

ELEMENTS OF GEOPHYSICAL FLUID MECHANICS

MATHEMATICS, KINEMATICS, DYNAMICS, SHALLOW WATER FLOW,
VORTICITY, BALANCED FLOW, VERTICAL COORDINATES, SCALAR MECHANICS

Stephen M. Griffies

Program in Atmospheric and Oceanic Sciences

Princeton University

SMG@princeton.edu

Draft from October 17, 2021



COPYRIGHT ©2021 BY STEPHEN M. GRIFFIES

ALL RIGHTS RESERVED

GRIFFIES, STEPHEN M., 1962-

ELEMENTS OF GEOPHYSICAL FLUID MECHANICS

THIS BOOK WAS TYPESET USING LATEX.

Contents

PREFACE	iii
A GUIDE TO THIS BOOK	ix
PRINCETON UNIVERSITY AOS 571	xv
LIST OF SYMBOLS	xxiii
Part I. Mathematical methods and concepts	1
1 Cartesian tensor algebra	5
2 Cartesian tensor calculus	21
3 Partial differential equations	45
4 Green's function methods	67
5 Curves and surfaces	109
6 General tensors in brief	119
7 General tensor analysis	129
8 Orthogonal coordinates	151
9 Differential forms	165
Part II. Geophysical particle mechanics	173
10 Particle mechanics	175
11 Symmetry and conservation	207
Part III. Kinematics of fluid flow	229
12 Continuum approximation	231
13 Fundamentals of fluid kinematics	241
14 Material fluid objects	269
15 Mass conservation	289
16 Tracer conservation	311

17 Non-divergent flows	331
 Part IV. Dynamics of geophysical fluid flows 349	
18 Momentum dynamics	353
19 Stress in fluids	369
20 Pressure form stress	405
21 Equilibrium thermodynamics	423
22 Energy dynamics	457
23 Hydrostatic and tangent plane approximations	491
24 Oceanic Boussinesq fluid	511
25 Buoyancy	553
26 Rotating flows	577
27 Horizontal flow balances	605
28 Ekman mechanics	621
29 Space and time dependent gravity	643
30 Surface gravity waves and Stokes drift	655
 Part V. Shallow water mechanics 671	
31 Shallow water models	673
32 Shallow water dynamics	695
 Part VI. Vorticity and potential vorticity 743	
33 Vorticity and circulation	745
34 Non-divergent barotropic flow	761
35 Shallow water vorticity and potential vorticity	787
36 Vorticity and circulation mechanics	829
37 Angular momentum, vorticity, and strain	871
38 Potential vorticity mechanics	879

39 Potential vorticity budgets	897
Part VII. Nearly geostrophic balanced flows 913	
40 Shallow water planetary and quasi-geostrophy	915
41 Continuously stratified planetary geostrophy	935
42 Continuously stratified quasi-geostrophy	959
Part VIII. Elements of generalized vertical coordinates 981	
43 Mathematical foundations	985
44 Kinematics	1003
45 Dynamical equations	1027
46 Isopycnal primitive equations	1037
47 Shallow water thickness weighted averaging	1049
Part IX. Mechanics of ocean scalars 1065	
48 Tracer advection and diffusion	1067
49 Eddy and mean tracer kinematics	1115
50 Subgrid scale tracer transport	1143
51 Ocean buoyancy and sea level	1165
52 Water mass analysis	1181

Contents

Preface

Synopsis

Geophysical fluid mechanics (GFM) is the study of classical fluid flow on a rotating and gravitating planet. In this book, our primary inspiration comes from terrestrial fluids in the atmosphere and ocean. Geophysical flows are in near solid-body motion with the rotating planet, thus prompting a description from the rotating (non-inertial) planetary reference frame. Body forces from both rotation (Coriolis) and gravitation (buoyancy) are fundamental features of geophysical fluid flows. We encounter a variety of conceptual models within these pages. Some are formulated within the context of a *perfect fluid* comprising a single material constituent and where fundamental processes are limited to the reversible and mechanical. Others are posed within *real fluids* comprised of multiple matter constituents exposed to irreversible processes such as friction, mixing and heating.

We develop geophysical fluid mechanics from a mathematical physics perspective at a level accessible to the graduate student in atmospheric sciences, oceanography, physics, engineering, applied mathematics, and allied fields. Mathematical derivations are offered in detail to expose the physics and to develop techniques required for the practitioner. Subjects are approached by first establishing general concepts and afterwards considering specializations, case studies, and approximations that exemplify the fundamentals. Though guided by natural phenomena, the treatment is centered on the mechanics of geophysical fluid mechanics. Hence, we study many topics to a degree beyond that needed if primarily focused on the transfer of phenomenological knowledge. For this reason, we find occasions to examine physics through multiple lenses that render a variety of perspectives and insights.

The presentation is based on the premise that skills as a physicist are best taught by nurturing physical reasoning, and with physical reasoning supported by mathematical precision coupled to the elucidation of concepts and principles. Hence, the presentation style is both deductive and descriptive. The deductive approach supports understanding through elementary physical notions that are expressed mathematically and then used as a foundation for conceptual theories. The descriptive serves to build skills in physical reasoning and the ability to articulate physical ideas using words as a complement to maths. In support of this style, we cultivate both the conceptual and technical by embracing the physics and maths at the heart of geophysical fluid mechanics. There is a synergy between the physical content of geophysical fluid mechanics and its mathematical expression, whereby physics informs the maths and maths reveals the physics. This synergy is fundamental to mathematical physics and this book is a humble attempt at its realization.

From lecture notes to book

This book grew from lecture notes developed for a first semester graduate course in geophysical fluid mechanics at Princeton University. The course develops fluid mechanics as framed within the arena of geophysical fluid flows and it drew from standard texts such as [Gill \(1982\)](#), [Pedlosky \(1987\)](#), [Apel \(1987\)](#), [Salmon \(1998\)](#), [McWilliams \(2006\)](#), [Marshall and Plumb \(2008\)](#), [Cushman-Roisin and Beckers \(2011\)](#), [Olbers et al. \(2012\)](#), and [Vallis \(2017, 2019\)](#). Additional insights were gleaned from the fluid mechanics texts by [Aris \(1962\)](#), [Batchelor \(1967\)](#), [Meyer \(1971\)](#), [Landau and Lifshitz \(1987\)](#), [Acheson \(1990\)](#), and [Kundu et al. \(2016\)](#).

The study of rotating and stratified fluids has evolved much over its history. During recent decades the study has seen particular evolution through deepening physical foundations, refining mathematical formulations, increasing the intellectual and predictive value of numerical simulations, extending applications across terrestrial and planetary systems, and expanding observational and laboratory measurements and techniques. What has emerged is a recognition that a fruitful study of rotating and stratified fluid flows makes use of ideas that go beyond the traditional notions of *geophysical fluid dynamics* (GFD). A contemporary practitioner develops insights by weaving together concepts and tools from mathematics, kinematics, dynamics, thermodynamics, scalar mechanics, numerical simulations, laboratory experiments, and field measurements. Acknowledging this broadening of the practice motivates the term “mechanics” in this book’s title, rather than the more focused “dynamics”. It is a seemingly trivial change in verbiage and yet it represents a nontrivial change in perspective.

Two pillars of geophysical fluid mechanics

We conceive of two pillars to theoretical geophysical fluid mechanics. The *elements pillar*, which is the focus of this book, comprises the physical and mathematical formulation of conceptual models used to garner insight into rotating and stratified fluid motion. This pillar is concerned with understanding how physical concepts are mathematically expressed in a mechanical description of geophysical flows. The particular study of geophysical fluid phenomena and corresponding solutions to equations, such as waves, instabilities, turbulence and general circulation, forms the *emergent phenomena pillar*. We introduce emergent phenomena, in particular waves and turbulence, where it serves as motivation for and illustration of elementary models. Even so, emergent phenomena are not our primary focus. Rather, this book focuses on physics forming the core of the equations and less on physics emerging from solutions to the equations.

The choice to focus on elements rather than emergent phenomena is partly a product of the course from which this book grew, which is a course that lays the foundations for the subject by deductively formulating the equations and exploring their physical content. Clearly the elements pillar and the emergent phenomena pillar are synergistic, thus offering lessons and guidance to one another. In many ways, the elements pillar is less widely emphasized by existing texts in geophysical fluid mechanics. This book attempts to partly remedy that situation while, hopefully, remaining accessible and relevant to students.

Causality and association

Throughout this book we develop equations describing the evolution of fluid properties, with such equations derived from foundational physical principles such as Newton’s laws of motion, thermodynamic principles, mass conservation, and vorticity mechanics. As part of this development we

often seek information about what *causes* a particular fluid motion. The question of causality is posed when studying Newton's equation of motion, which says that acceleration (motion) arises from a force (the cause of the motion). However, on the surface, this fundamental statement of classical mechanics offers little more than the definition of a force. We make physical progress only after specifying the nature of the force; e.g., gravitational, electromagnetic, etc.¹

In geophysical fluid mechanics, we refer to equations that determine time changes as *prognostic equations*, whereby knowledge of forces enable one to predict the flow evolution. The question arises how to determine the forces, particularly when the forces are commonly dependent on the flow. Such is the complexity and beauty inherent in nonlinear field theories, such as fluid mechanics, where cause and effect are intrinsically coupled. We can sometimes make progress by turning the problem around, whereby kinematic knowledge of the motion offers inferential knowledge of the dynamical forces. This situation is exemplified by pressure forces acting within a non-divergent flow whereby pressure acts, instantaneously and globally, to satisfy the constraint that the velocity remains non-divergent.

Besides seeking causal relations pointing us toward the future, many basic questions of fluid mechanics arise either instantaneously, as in the constraints maintaining non-divergent flows, or when the flow is steady, in which case the flow at each point in space has no time dependence. In steady flows, the net acceleration, and hence the net force, vanish at each point within the fluid, although the fluid itself can still be moving. Hence, for steady flows we are unable to make statements about causality since the steady equations are *diagnostic* rather than prognostic. Diagnostic relations thus provide statements about *associations* between physical processes. The *geostrophic balance* is the canonical diagnostic relation appearing in geophysical fluid mechanics, where the Coriolis force is associated with (i.e., balances) the horizontal pressure force. Further associative statements are particularly common when studying steady vorticity balances. So although a predictive theory requires prognostic equations that manifest causal relations, an understanding of how fluid motion appears, and in particular how it is constrained, can be fruitfully studied through diagnostic relations.

Geophysical Fluid Mechanics and Climate Science 2.0

Fluid mechanics has a history of applications that span science and engineering, from blood flow to the stability of galaxies. A key 21st century application of geophysical fluid mechanics concerns the questions of climate science associated with the uncontrolled greenhouse gas experiment pursued by industrialized civilization's carbon centered energy use. Leading order questions about climate warming have been sufficiently addressed to recognize that the planet has reached a crisis point threatening the viability of the biosphere. Even so, mechanistic answers to a number of questions remain at the cutting edge of climate science research. What will happen to the atmospheric jet stream and storm tracks in a world without summer Arctic sea ice? Will tropical storms be more powerful in a warmer world? What are the patterns for coastal sea level rise and their connections to large-scale ocean circulation? What are the key processes acting to bring relatively warm ocean waters to the base of high latitude ice shelves? How stable is the ocean's large-scale overturning circulation and its associated poleward heat transport? Are there feasible and sustainable geo-engineering options that equitably reduce the negative impacts of climate warming without introducing new problems? These questions, and countless others, constitute the scientific challenges of *Climate Science 2.0*.

¹For more on this perspective of Newton's laws, see Chapter 1 of *Symon* (1971).

Numerical circulation models, a core tool for Climate Science 2.0, have reached a level of maturity that they vividly reveal details of the complex and multi-scaled nature of planetary fluid flow. Geophysical fluid mechanics is essential for the development of robust numerical methods and subgrid scale parameterizations for such models, along with mechanistic analyses and interpretations of simulated data. In turn, geophysical fluid mechanics furthers predictive capability for weather and climate forecast systems and enhances confidence in projections for future climate. Moving from the numerical world to the natural world, we see that the notions of geophysical fluid mechanics play a key role in the design of observational field campaigns and the analysis of their measurements. Such field measurements probe process-oriented questions through the use of technologies that facilitate increasingly ambitious investigations within natural fluid motions. So in brief, the concepts and tools from geophysical fluid mechanics provide the means to conceptually digest the underlying physics probed by both natural and simulated experiments, and to communicate the findings to the broader scientific community. Hence, geophysical fluid mechanics, as part of the scientific investigation of the earth's climate, serves an essential role in humanity's quest for sustainability, equity, and justice.

Gratitudes

This book benefitted from interactions with students taking part in Princeton University's AOS 571, a course that I taught during the autumn semester from 2014-present. Further inspiration was offered by students, postdocs, and fellow research scientists who I encountered as part of my research career. I wish to particularly highlight a variety of journal clubs held as part of my research group, where a great deal of head scratching led to insights sprinkled through these pages.

I remain humbled and grateful for being part of the unique research environment cultivated at NOAA's Geophysical Fluid Dynamics Laboratory as well as Princeton University's Atmospheric and Oceanic Sciences program. As part of my research and mentoring in this environment, I have encountered thinkers whose style, questions, and insights have taken root in my work. This environment has also afforded me the opportunity to travel the world to interact with various teachers and colleagues whose wisdom and love of science and life are infectious and inspiring. Throughout these interactions, I have entered into trusting and non-judgmental spaces where deep learning and understanding spontaneously arise. Partaking in these spaces, where heart and mind meld, has been among the most satisfying experiences of my life. Safe spaces for honest, diverse, and inclusive learning are precious, and I am grateful to those who nurture that space.

A book of this nature is a not a simple endeavor. It starts modestly, grows over time, and eventually becomes a passion if not an obsession. I was particularly drawn to writing during the many months of COVID-19 pandemic that kept me inside far more than during non-pandemic times. I found little objective guidance to know when to let this book live outside of my hands. That decision, when it occurs, will no doubt be a combination of intuition and exhaustion. Indeed, when nearing the end of my PhD I was told that a thesis is never finished but instead it must be abandoned. That pithy remark also works well for this book.

Writing this book has been an exercise in rational thought that was fed by spiritual food from meditation, yoga, family, and community. In particular, each step of this project was supported by my wife, Adi, and our son, Francisco. I am deeply grateful for their patience and trust as I satisfied the goal of writing this book through countless nights, weekends, and holidays in our home attic space. I treasure being part of this family and I dedicate this work to you two amazing human beings.

Disclaimer

Facets of this book originate from research papers published while I was a US Government federal employee. Reference to these papers is provided at the appropriate place in the text. However, no sentence in this book was taken verbatim from published papers. Rather, the material has been extensively reworked and refined to enhance pedagogical value for the reader, thus going well beyond that appropriate for research papers or reports. Additionally, this book's writing occupied personal time and was not part of any official assigned government duty, and I made use of personal equipment for typesetting and producing figures.

About the cover

I took the cover photo of an iceberg, ocean, clouds, and sea bird (can you find the bird?) in the Orkney Passage region of the Southern Ocean during a cruise from March-May 2017 aboard the British research ship James Clark Ross. I encourage any person, particularly theorists such as me, to partake in field research. It will enhance your science and deepen your soul.

Caveats and limitations

Although reaching some level of maturity, this book remains a work in progress not yet ready for publication. Here are just a few of the items required before it is ready for its next stage.

- A number of further topics will fill these pages.
- More effort is needed to unify notation, complete and organize the index, develop further exercises, enhance figures, improve discussions, and correct errors.
- Many sections and chapters have yet to be scrutinized by readers. I hope to garner help from those interested in providing candid feedback to improve presentation and correct mistakes. Feedback is solicited particularly where the reader identifies poor or confused writing, incorrect concepts, or math errors. Nothing is too trivial for comment.



A guide to this book

No book is an island, with this book generously making use of other books, review articles, research papers, and online tutorials. Many readers find value in studying a subject from a variety of perspectives, thus justifying the proliferation of books with overlapping subject matter. In an attempt to support this form of study, we commonly provide pointers to written and/or video presentations. Many more resources are available through a quick internet search.

There is no pretense that any reader will penetrate all topics nor read this book cover-to-cover. This recognition is particularly keen in a world where research and educational agendas often spread rather than focus our attention. Hence, an attempt has been made to facilitate picking up the book at a variety of starting points. For this purpose, each chapter and/or part is written in a reasonably self-contained manner even at the cost of modest redundancy. When redundancy becomes onerous, cross-referencing identifies allied material treated elsewhere in the book. In the remainder of this front matter, we summarize the book's organization and then offer some general pointers on how to pursue its study.

Organization

This book is organized into parts according to their particular focus, with each chapter starting with a brief guide to the material and pointing to dependencies to other chapters. Some chapters focus on topics required for a basic understanding of the subject and offer exercises to test that understanding. Other chapters offer monograph style topics that further the foundations and exemplify applications largely taken from a selection of the author's research interests. Although this book is sizable, there are many topics omitted. These omissions reflect on the author's limited energy as well as the desire to keep the book from growing even larger.

MATHEMATICAL METHODS AND CONCEPTS

Part I provides a suite of mathematics chapters of use in studying geophysical fluid mechanics. Some chapters concern topics readily found in applied mathematics or mathematical physics texts, whereas others provide a grounding in mathematics rarely seen outside of geophysical fluid mechanics, in particular the maths of generalized vertical coordinates. The casual reader can skim these chapters without sacrificing too much from later chapters, assuming a working knowledge of vector differential and integral calculus. Where more difficult math topics arise in later chapters, the

reader is encouraged to return here to develop skills and to firm understanding. We propose that a variety of avenues for physical insights should be embraced by the practitioner, with mathematical acuity offering great rewards to the physicist.

GEOPHYSICAL PARTICLE MECHANICS

Part II provides a summary of Newtonian particle mechanics with particular emphasis on particles moving around a rotating sphere. This study exposes the physics of motion when viewed from the non-inertial reference frame of a terrestrial observer. By doing so, we encounter such topics as the Coriolis and centrifugal accelerations as well as spherical coordinates, all within the relatively simple context of point particle mechanics.

KINEMATICS OF FLUID FLOW

Mechanics is comprised of kinematics (the study of intrinsic properties of motion) and dynamics (the study of forces causing motion). In Part III we initiate a study of fluid mechanics by focusing on the kinematics of fluid flow and matter transport within that flow. Our treatment exposes both the Eulerian and Lagrangian viewpoints and emphasizes the variety of kinematic notions and tools key to describing fluid motion. We also encounter facets of material transport as described by the tracer equation, thus laying the foundation for *tracer mechanics*. It is in these chapters that we nurture the brain-muscle to think about fluid motion. Fluid flow, and the transport of matter within that flow, have many features fundamentally distinct from point particle and rigid body motion. We thus cultivate distinct tools and concepts that can take much practice to intellectually digest.

Quite often a course in fluid mechanics skims over fluid kinematics and tracer mechanics, preferring instead to focus mostly on dynamics. Indeed, some kinematic topics can seem a bit esoteric, particularly the study of Lagrangian kinematics. However, an incomplete understanding of fluid kinematics and tracer mechanics can lead to many difficulties with fluid dynamics. The reader is thus encouraged to study the chapters here in detail, and to revisit the material as the need arises during later chapters.

DYNAMICS OF GEOPHYSICAL FLUID FLOW

Part IV presents the core chapters of geophysical fluid dynamics as studied in this book. Within these chapters we study how Newton's laws of mechanics and the postulates of thermodynamics are used to describe fluid motion on a rotating and gravitating sphere. These concepts and methods, along with those from kinematics and tracer mechanics, leads to elemental features of atmospheric and oceanic flows. We study the subject by focusing on how forces that act on fluid elements and fluid regions lead to accelerations and thus to motion. These forces both act throughout the volume of a fluid element (*body forces* from gravity, Coriolis, and centrifugal) as well as on the boundary of a fluid element (*contact forces* from pressure and friction).

SHALLOW WATER MECHANICS

In Part V we study the mechanics of a shallow water fluid, with a shallow water fluid comprised of hydrostatically balanced homogeneous fluid layers. The layers are also typically assumed to be immiscible, so that interactions between layers occur only via mechanical forces from pressure acting at the layer interfaces. The shallow water fluid allows us to focus on rotation and stratification without the complexities of vertically continuous stratification and thermodynamics. Many physical

insights garnered by studying shallow water fluids extend to more realistic fluids, thus making the shallow water model very popular among theorists.

VORTICITY AND POTENTIAL VORTICITY

Part VI dives into the subject of vorticity and potential vorticity. This subject is rich and elegant, though it requires perhaps the most work for the student to penetrate given the heavy reliance on vector and integral calculus. Vorticity plays a role in the motion of all geophysical fluids since motion on a rotating planet provides a nonzero *planetary vorticity* even to fluids at rest on the planet. This feature of geophysical fluids contrasts to many other areas of fluid mechanics, where irrotational flows are commonly encountered. Potential vorticity is a strategically chosen component of the vorticity vector that marries mechanics (vorticity) to thermodynamics (stratification). Material conservation properties of potential vorticity are striking and render important constraints on fluid motion. Indeed, perhaps the most practical reason to study vorticity concerns the various constraints imposed on the flow moving on a rotating and gravitating sphere. These constraints provide conceptual insights and predictive power.

NEARLY GEOSTROPHIC BALANCED FLOWS

Part VII introduces the topic of balanced models, with our attention limited to the non-divergent barotropic model, quasi-geostrophy, and planetary geostrophy. Balanced models generally remove the divergent motions associated with gravity waves, thus allowing us to focus on the large-scale vortical motions. Balanced models have a rich history among theoretical geophysical fluid studies, providing insights into both laminar flows through planetary geostrophy, and turbulent flows through quasi-geostrophy. Our treatment is relatively terse.

ELEMENTS OF GENERALIZED VERTICAL COORDINATES

Part VIII brings together the various elements of a generalized vertical coordinate description of geophysical fluid mechanics. The chapters offer details of the maths, kinematics, dynamics, and applications. This material is central to many current research activities, including subgrid scale parameterizations and the design of numerical atmosphere and ocean models.

MECHANICS OF OCEAN SCALARS

Part IX dives into the mechanics of scalar fields as realized in the ocean, with our focus on active tracers (temperature and salinity), passive tracers, and buoyancy. Much of this study forms the basis of *tracer mechanics*, which has proven very important for the ocean given that it is generally very difficult to measure dynamical fields such as velocity and vorticity, whereas tracer distributions are much more readily measured.

Written and spoken communication

Reading lecture notes, research articles, and books exposes one to sound and clear writing, as well as to unsound and obscure writing. Attending lectures and seminars in person or online also exposes one to a range of speakers. Some speakers offer the benefits of a sleeping aide without the cost and side-effects of a pharmaceutical. Others rush through more material than even they can digest, whereas others pedagogically engage in a lecture even without any prior knowledge of the

material. To succeed in research and teaching, you need to master elements of both written and spoken communication. Hence, it is useful to nurture these skills throughout your career.

CLEAR THINKING LEADS TO CLEAR COMMUNICATION

Clear communication is the sign of clear thinking. Some people communicate better in writing, where one has the opportunity to carefully compose, organize thoughts, and edit. Others are better at speaking, where spontaneous and interactive reflections and experience can bolster the clarity of a presentation. Both modes of communication are important in science and engineering.

As inspiration, for both clear and obscure, pick up one of your textbooks or class lecture notes and analyze the presentation for clarity. Where are you confused? Where is the material crystal clear? Then pick up a journal article and perform the same analysis. What do you like? What do you dislike? Then go to YouTube and find a science or engineering lecture, old or new. What makes the speaker engaging and clear, or boring and obscure?

EMPATHY IS KEY

Empathy is a basic facet of effective communication and teaching. Place your mind inside that of an interested and smart reader or listener. Identify with their quest to understand what you wish to communicate. What assumptions are you making? Are the assumptions justified based on the audience? How compelling is your scientific story? As a start along the path towards clear written communication, I have accumulated some pointers in an online document: [Elements of Style for Journal Papers \(2013\)](#).

CLARITY HELPS, BUT SOME MATERIAL IS JUST TOUGH

Although poor communication hinders our ability to digest new ideas and concepts, it is also important to appreciate that some material is tough no matter how well it is communicated. We should aim to make a subject matter as simple as possible, but not simpler (paraphrasing Einstein). Furthermore, it sometimes takes one or two generations of teaching before some scientific material can be sufficiently digested to allow for the core conceptual nuggets to be revealed. As an example, try reading Newton or Maxwell's original works as compared to a modern presentation of Classical Mechanics or Electromagnetism. So we do strive for clear communication, but we cannot presume that clarity is sufficient to remove the struggles we all experience when learning.

Pointers on problem solving

Most people are not born with *a priori* physics problem solving skills. Rather, it takes extensive practice to develop the necessary brain muscle. Here are some general pointers to keep in mind when diving into a physics problem, whether it is for a class or the basis of a broader research question.

CHECK FOR DIMENSIONAL CONSISTENCY

The symbols we use in mathematical physics correspond to geometrical objects (e.g., points, vectors, tensors) describing a physical concept (e.g., position in space, velocity, temperature, angular momentum, stress). Hence, the symbols generally carry physical dimensions. The three physical

dimensions we are concerned with in this course are length (L), time (T), and mass (M). Physical dimensions of the equations must be self-consistent. For example, if one writes an equation

$$A = B, \quad (1)$$

where A and B have different physical dimensions, then the equation makes no sense physically. Something is wrong. Although not always sufficient to uncover errors, dimensional analysis is an incredibly powerful means to debug the maths.

CHECK FOR TENSORIAL CONSISTENCY

In the same way that mathematical equations in physics need to maintain dimensional consistency, they must also respect tensor rules. For example, the equation

$$A = B, \quad (2)$$

makes mathematical sense if A and B are both scalars. Likewise,

$$\mathbf{A} = \mathbf{B} \quad (3)$$

makes sense if \mathbf{A} and \mathbf{B} are both vectors. However, if both \mathbf{A} and \mathbf{B} are vectors, then the equation

$$\mathbf{A} = \nabla \cdot \mathbf{B} \quad (4)$$

does not make sense because the left hand side is a vector and the right hand side is a scalar. A bit more subtle example is when \mathbf{A} is a vector yet \mathbf{B} is an axial vector. In this case \mathbf{A} remains invariant under a change from right hand to left hand coordinates whereas \mathbf{B} flips sign. Maintaining basic tensorial rules can be considered the next level of sophistication beyond dimensional analysis.

USE WORDS AND PICTURES

It is important to explain the problem and solution using words and pictures. Hence, it is good practice to liberally include words and sentences in between the key equations, with the purpose to explain what the maths means using clear English. Here are some practical payoffs for this style of presentation.

- In the process of trying to explain the maths using words and pictures, you generally must dive deeper into the logic of the problem. In doing so, you often identify weak points, incomplete assumptions, and errors. This process is a very important learning stage in preparing to stand in front of people to present results and to answer questions. It is a key facet of research and teaching.
- Physics teachers are often more forgiving of math errors if you convince the teacher that you have a sensible physical understanding of the problem. Plain English and pictures are very useful means for this purpose.

THERE IS OFTEN MORE THAN ONE PATH TO A SOLUTION

In physics, there is often more than one path to a solution. Pursuing distinct paths offers added physical and mathematical insight, exposes assumptions, and allows one to double-check a solution. Some of the most profound findings in physics came from pursuing distinct formulations. One example concerns the distinct formulation of mechanics offered by Newton (1642-1746), and then

later by Lagrange (1736-1813) and then Hamilton (1805-1865). Had Lagrange or Hamilton rested on the merits of Newton's formulation, we may well have had a very different intellectual evolution of 19th and 20th century physics.



Princeton University AOS 571

Welcome to AOS 571 for the 2021 autumn semester. This course focuses on geophysical fluid mechanics, which is a physics discipline concerned with motion of rotating and stratified fluids such as those in the earth's atmosphere (mostly a gas) and the ocean (mostly a liquid). We develop certain of the conceptual and mathematical foundations of the subject by studying elemental features of planetary fluid motion, and by providing a mathematical physics basis for understanding these patterns.

Here are details of your teacher and the course.

Dr. Stephen M. Griffies

SMG@princeton.edu

<https://stephengriffies.github.io/>

Class Monday and Wednesday 2:30pm-3:50pm

In-person classes held at 154 Guyot Hall

Zoom for Q & A from 2pm-3pm each Friday

Worked homeworks uploaded by students to Canvas

Class materials and communication via Canvas

Zoom recorded class lectures from 2020 are available via Canvas

Class structure and expectations

Our goal for this course is learn how to formulate and to solve problems in geophysical fluid mechanics, and to develop an understanding of the basic physical theory. To help reach this goal, I have prepared extensive notes (now a book in progress) and we will work though selected pieces in class. You will also read far more than covered in class.

Garnering an understanding of geophysical fluid mechanics can require great effort and practice. To help that process, you are expected to read the assigned material, work through the derivations, and hand-in homework exercises. You will have many opportunities to develop the necessary brain muscle assuming you maintain the discipline to keep up with the material. Please ask questions, preferably in class, when you are unsure of anything. Also, please strive to solve all of the assigned exercises, either alone or collaboratively with other students.

CULTIVATING A SAFE SPACE FOR HEALTHY LEARNING

A fundamental feature of any class concerns the atmosphere developed for learning. My aim as class teacher is to foster an inclusive, friendly, generous, patient, and non-judgmental space for learning by students and the teacher. Key principles that support this space include equity, diversity and inclusion, each of which are intrinsically valuable and an essential feature of ethical research and education. We also acknowledge and honor past generations whose efforts, some of which were garnered through force and oppression, have led to the rewarding research environment that we now work. It is our sincere hope that practicing the above principles supports present and future generations in a way that helps to heal past injustices. Cultivating this safe and grateful space supports deep learning while genuinely appreciating contributions from individuals without regard to race, ethnicity, culture, religion, sexual orientation, gender identity and expression, physical ability, age, socioeconomic status or nationality. As a participant in this space, we each celebrate diversity and nurture an inclusive and friendly community that is optimized for shared learning and mutual understanding.

For many, this class will require a tremendous amount of effort. It will require much focus and energy to master the material even for those with a firm background. Regardless your background, talents, or interests, I am here to help in anyway you need. So please reach out if you are struggling. I also encourage you to develop working relations with your classmates. Homework exercises can be done with groups, and working with others offers great opportunities for learning.

Even if you prefer to work alone, I ask that you develop some form of a relation with one or more of your classmates. Part of this recommendation is based on the need to informally monitor the health of each other, particularly give that the early years of graduate school can be very tough. Even so, it can be a time for building deep friendships and community as we share in the process of learning how to thrive as budding researchers and teachers.

ONLINE LECTURE NOTES FOR THE CLASS

The class lectures closely follow selected material from [Griffies \(2021\)](#) available online. These notes started in 2014 as annotations to the textbooks from [Vallis \(2017\)](#) and [Vallis \(2019\)](#), with both books highly recommended supplements to the course. In recent years, the class notes have grown into an autonomous document from which the course material directly derives.

You are asked to read through the class notes and to view the assigned videos **prior to the class**. During class we will discuss salient points from the assigned notes. To allow sufficient time for interactive discussion and questions, not all of the assigned reading material will be directly covered in class. Hence, you will be expected to read and to understand more material than is covered on the chalkboard. Correspondingly, you are welcome to bring any questions related to the notes to the class, even if the material was not covered in class.

CLASS GRADE = HOMEWORK (75%) + FINAL EXAM (25%)

There are two areas where grades are earned: homework exercises and take-home final exam. Additionally, I ask that everyone aim to attend all classes and participate by asking questions.

Worked homework exercises are due at the start of class (uploaded to Canvas) one week after the exercises are assigned (unless otherwise noted). Students can make use of any resources for solving homework exercise, including other people. Clarification of questions can also be obtained via email to me. If you find a solution from a source other than your own head, then be sure that you fully understand both the essence and the detail of the solution. Although you are encouraged

to discuss the problems with other students, you are cheating yourself if you merely parrot another person's answer without fully grokking it yourself.

There is no pretense that the exercises offered in class are clearly formulated. Indeed, a certain degree of ambiguity reflects the *status quo* in research, where formulating a novel and insightful question is generally the most difficult part of the research process. Additionally, the solutions may not be 100% correct or ideal from a pedagogical perspective. Rather, they represent a work in progress. If mistakes or ambiguities are found, then please share your questions and concerns.

There is a take-home final exam during exam period. You are asked to do the work as a solo student, with no help or consultation from another human. However, you can make use of books, notes, online resources, etc. The questions are generally taken from published papers with relevant references provided to the student.

PLEASE DISCONNECT ELECTRONICS FROM INTERNET DURING CLASS

The class time is comprised of lectures based on the class notes along with discussion. The pace will be gauged on questions during the lectures and my sense for how well the class is grokking the material. You are encouraged to follow lectures by having a copy of the class notes on-hand, either electronically or printed.

To support your learning and teaching experience, and those of your classmates, please ensure that you turn all electronic devices into "airplane mode" so that you are not tempted to divert attention to non-class issues. To get the most from the lectures and class time requires focused attention and participation.

Assignments and final exam

Here we consider some context for the class assignments and final exam.

AIM FOR A BALANCE BETWEEN BEING THOROUGH AND BRIEF

There is often a conflict between showing full mastery of a problem and keeping the solution write-up brief. In general, there is no need to re-derive equations already presented in the class lecture notes. Proper referencing of the equation is all that you need; i.e., tell me something like "starting from equation (X.YY) from the class notes." Additionally, when presenting a derivation, you may choose to show just the key steps rather than all intermediate steps. Determining what is a "key" step is largely up to you, but it should be something you learn to do in time. Nonetheless, as per the previous pointer, you are encouraged to show more than one approach to a solution.

QUESTIONS FOR POINTS OF CLARIFICATION

Questions for points of clarification will be entertained if you feel the problem is ill-posed or if you are totally lost. Email is the most efficient means to communicate to me. Responses will generally be sent through Canvas so that all students can see the response, thus keeping everyone with the same information. Correspondingly, questions within 24 hours of the deadline are generally not entertained so to ensure that all students have time to see the response.

MISTAKES OFFER SIGNIFICANT OPPORTUNITIES FOR LEARNING

Everyone makes mistakes, some more than others. The toughest part is the self-imposed shame or embarrassment. Please try to keep a positive mind about your mistakes. As you will discover,

mistakes offer significant opportunities for learning. I am a poster-child for this process!

So do not be overly anxious if you find many marks on your homeworks and exams. Rather, aim to use mistakes as learning opportunities. That is how life in academics (life in general!) works. Furthermore, be completely honest with yourself to candidly identify weaknesses. I will do my best to work personally with you if something remains uncertain or you feel there is a weakness in your skills that needs some extra help. Please seek help should you wish it. And finally, please do question my marks should you feel they are unfair or incorrect. I am prone to mistakes in my grading.

SOME SPECIFIC POINTS ABOUT GRADING YOUR WORK

Assignments are generally marked using the following rules, with grading less forgiving as the class progresses through the semester. The following mistakes are marked down by an increasing amount moving through the list.

1. **SIGN ERRORS:** Sign errors are a nuisance. We all must spend time to uncover them. One means of detecting errors is to try explaining the maths to yourself or someone else. Does the result make sense? If not, then perhaps there is a sign error. I am generally not too upset with sign errors if they have minimal physical relevance. But when they indicate a physical misconception then I will mark it more harshly.
2. **MATH ERRORS:** Math errors, such as those associated with basic calculus mistakes, are generally marked.
3. **DIMENSIONAL AND TENSORIAL ERRORS:** I am relatively unforgiving of dimensional mistakes and tensorial inconsistencies.
4. **PHYSICALLY MISSING THE POINT:** Evidence of physically missing the point will generally invoke the most negative marks depending on the depth of the misconception. The best way to convince me you grasp the basic physics is to use words and pictures. If the maths is missing or totally wrong, but you present some sensible words and pictures, then that will help earn some credit.

Here are some further considerations for your homework assignments.

- **PRESENTATION OF THE SOLUTION:** Please write clearly and legibly. I greatly appreciate solutions submitted in L^AT_EX, particularly if your handwriting is poor. If the equations and words are sloppy, you will find teachers less forgiving of errors. You must convince the teacher that you understand the solution and present the maths in a legible manner. Good communication skills are key to being a good scientist or engineer.
- **DEADLINES:** Please do your best to be on time with handing in homeworks and exams, with disasters, personal tragedy, and significant accidents the only excuses for late assignments. Fairness to those meeting the deadlines is the fundamental reason to insist on this rule.
 - You may hand-in homework late but only with prior arrangement. Otherwise, the homework set will receive zero credit.

Course material and strategies for thriving

The course material consists of class notes plus a selection of online videos. Ideally, the notes are to be read and the videos viewed **prior to each class**. Salient points related to the material will be presented during class on the chalkboard, along with further discussion and questions. We will *not* cover all the material in class that is expected to be learned as part of the reading assignments and problem sets.

Here is a strategy for thriving in this course.

- PRIOR TO EACH CLASS
 - View assigned videos (roughly 5 to 45 minutes).
 - Read assigned class notes (can take one to three hours).
- AFTER EACH CLASS
 - Re-read lecture notes and review assigned videos.
 - As needed, view recorded class lecture from 2020 to fill in gaps.
 - Work on assigned exercises alone and then gather with classmates (in-person or Zoom) to discuss.
- QUESTIONS: Address questions by reaching out to the teacher or classmates during class, after class, during office hours, during group gatherings, or via Ed Discussion. There are no silly questions, so please ask any and all!

Course outline

1. Course introduction and mathematics refresher
 - [9-minute video from Prof. N. Hall](#) introducing some basic images from terrestrial fluid flows.
 - [16-minute video from 3Blue1Brown](#) on the divergence and curl operations with examples taken from fluid mechanics.
 - Course introduction
 - Chapter 1: cartesian tensors
 - Chapter 2: vector calculus
 - Chapter 12: the continuum hypothesis used for describing fluids as a continuous media
2. Geophysical particle mechanics
 - [Watch the first 5-minutes from this UCLA Spin Lab video](#) on Coriolis effect.
 - Chapter 10: kinematics of a particle moving around a rotating sphere, including position, velocity, acceleration, Cartesian and spherical coordinates, rotating reference frame, Coriolis acceleration, planetary centrifugal acceleration
 - Chapter 10: dynamics of a particle moving around a rotating sphere using Newton's equation of motion, rotating reference frames, gravitational geopotential
3. Symmetries, conservation laws, and constrained motion

- Chapter 11: mechanical energy, potential momentum, inertial oscillations, axial angular momentum

4. Fluid kinematics

- [4-minute video on Eulerian and Lagrangian descriptions](#) from Prof. Hogg
- [27-minute video on Eulerian and Lagrangian descriptions](#) from Prof. Lumley.
- Chapter 13: Fluid kinematics, Eulerian and Lagrangian descriptions, Galilean invariance, material time derivative, flow lines

5. Mass conservation

- [5-minute video on mass conservation](#) from Prof. Hogg.
- Chapter 15: continuity equation, mass budget for fluid elements and finite regions, kinematic boundary conditions

6. Tracer conservation

- Chapter 16: barycentric velocity, tracer equation, budgets for infinitesimal fluid elements, budgets for finite fluid regions, Leibniz-Reynolds transport theorem, boundary conditions

7. Kinematics of incompressible flow

- [4-minute video on streamlines](#) from Prof. Hogg.
- Chapter 17: scalar streamfunction, vector streamfunction, area and volume conservation, meridional-depth overturning circulation

8. Momentum dynamics

- [6-minute video on momentum](#) from Prof. Hogg.
- Chapter 18: momentum dynamics, accelerations, contact forces, body forces, special forms of the momentum equation, axial angular momentum

9. Stress in fluids

- [2.5-minute video on stress and strain](#) from Prof. Hogg.
- [8-minute video on stress](#) from Prof. Hogg.
- Chapter 19: stresses and the stress tensor, linear momentum budget, relating stress to strain, form stress, boundary conditions
- Chapter 20: pressure form stress

10. Energy dynamics and filtered equations

- Chapter 22: thermodynamics of a moving fluid, mechanical energy, internal energy, total energy
- [8-minute video on hydrostatic pressure](#) from Prof. Hogg.
- Chapter 23: primitive equations, hydrostatic approximation, tangent plane approximation, basics of time evolving a fluid state

11. Boussinesq ocean equations and geostrophic mechanics I

- Chapter 24: oceanic Boussinesq approximation
- [30-minute video from Prof. Fultz](#) for an overview of rotating fluids.
- [4-minute video from the UCLA SpinLab](#) for examples of Taylor columns.
- Chapter 26: Rossby number, geostrophy, planetary geostrophy, Taylor-Proudman, thermal wind, isopycnal form stress

12. Buoyancy

- [49-minute video from Prof. Lewin](#) for an overview of buoyancy as well as other features of fluid mechanics.
- Chapter 25: Archimedes' principle, buoyancy, stratification, gravitational stability, mass density for perfect and realistic fluids

13. Geostrophic mechanics II and balanced inviscid horizontal flows

- Chapter 26: Rossby number, geostrophy, planetary geostrophy, Taylor-Proudman, thermal wind, isopycnal form stress
- Chapter 27: natural coordinates; centripetal, centrifugal, and Coriolis accelerations; exact geostrophic flow; inertial motion of fluid particles; cyclostrophic balance; gradient wind balance

14. Ekman layer mechanics

- Start around the 23-minute mark [of this video from Prof. Fultz](#) for his discussion of Ekman layers.
- Chapter 28: natural coordinates, spiral motion across isobars, non-dimensionalization and the Ekman number, net mass transport

15. Formulation of shallow water models

- [30-minute video](#) on shallow water model from Prof. Hall.
- Chapter 31: thickness equation, momentum equation, reduced gravity model, stacked shallow water layers, shallow water layer in a rotating tank

16. Shallow water dynamics

- Chapter 32: geostrophy, form stress, mechanical energy including available potential energy, angular momentum in a tank

17. Vorticity, circulation, and potential vorticity

- [23-minute video on vorticity](#) from Prof. Shapiro
- Chapter 33: vorticity and circulation
- Chapter 35: shallow water vorticity and potential vorticity

18. Vorticity mechanics

- [21-minute video on vorticity](#) from Prof. Shapiro.

- [5-minute video on vortex rings and Helmholtz's theorems](#) from the Physics Girl.
- Chapter 36: vortex lines and tubes, Kelvin's circulation theorem, mechanics of baroclinicity, β -effect

19. Balanced models I (single shallow water layer)

- Chapter 34: barotropic vorticity equation
- Chapter 40: Buckingham's II theorem, asymptotic expansion in terms of small Rossby number, shallow water planetary geostrophy and quasi-geostrophy

20. Balanced models II (continuous stratification)

- Chapter 41: asymptotic derivation of continuously stratified planetary geostrophy and properties of these equations.
- Chapter 42: asymptotic derivation of continuously stratified quas-geostrophy and properties of these equations.

We generally have 24 classes per semester. So that above 20 lectures will be supplemented with four further lectures depending on the class needs and wants. The following topics have been covered in the past and will be presented if there is time.

1. Kinematics of material objects

- Chapter 14: Eulerian and Lagrangian description of material lines, surfaces, and volumes within a perfect fluid.

2. Advection and diffusion

- Chapter 48: advection maths and physics; diffusion maths and physics

3. Thermodynamics

- Section 2.8: refresher on exact and inexact differentials
- Chapter 21: First law of thermodynamics, thermodynamic potentials, ideal gas atmosphere

4. Potential vorticity mechanics

- Chapter 38: PV material invariance for perfect fluid; PV evolution with friction and heating; impermeability theorem; isopycnal layer integrated PV

5. Shallow water gravity waves and geostrophic adjustment

- [14-minute video on gravity waves](#) from Prof. N. Hall.
- Section 32.8: gravity waves in single shallow water layer
- Section 35.7: geostrophic adjustment



List of Symbols

Many symbols encountered in this book are defined local to their usage and are not found far outside of that location. Many other symbols appear in a variety of places and are included in the following list that offers a quick reminder of their meaning. Symbols are here categorized according to whether they are English (Latin-based), Greek, or mathematical operators.

NON-DIMENSIONAL NUMBERS

SYMBOL	MEANING
Bu	Burger number: $Bu = \text{deformation radius}/\text{horizontal length scale of flow} = (L_d/L)^2$
Ek	Ekman number: $Ek = \text{frictional acceleration from vertical shears}/\text{coriolis acceleration}$
Fr	Froude number: $Fr = \text{fluid particle speed}/\text{wave speed}$
Ge	Geostrophic number: $Ge = \text{Coriolis acceleration}/\text{pressure gradient acceleration}$
Kn	Knudsen number: $Kn = \text{molecular mean free path}/\text{macroscopic length scale}$
Re	Reynolds number: $Re = \text{inertial acceleration}/\text{frictional acceleration} = U L/\nu$
Ri	Richardson number: $Ri = \text{buoyancy stratification}/\text{squared vertical shear} = N^2/ \partial_z \mathbf{u} ^2$
Ro	Rossby number: $Ro = \text{inertial acceleration}/\text{Coriolis acceleration} = U/(f L)$



ENGLISH-BASED SYMBOLS: PART I

SYMBOL	MEANING
A_v	Avogadro's number: $A_v = 6.022 \times 10^{23} \text{ mole}^{-1}$
b	buoyancy of a fluid element, with $b > 0$ for fluid relatively light: $b = -g(\rho - \rho_0)/\rho_0$
c_s	sound speed: $c_s^{-2} = \partial\rho/\partial p$
c_p	heat capacity at constant pressure: $c_p = [\partial\mathcal{H}/\partial T]_{p,C}$
C	tracer concentration = mass of tracer per mass of fluid = tracer mass fraction
C_D	dimensionless bottom drag coefficient: $C_D > 0$
\mathcal{C}	circulation of velocity around the boundary of a surface $\mathcal{C} \equiv \oint_{\partial\delta} \mathbf{v} \cdot d\mathbf{r}$
\vec{e}_a	basis vectors for a chosen coordinate system
f, f_0	Coriolis parameter and a reference value, also the planetary vorticity: $f = 2\Omega \sin\phi$
\mathbf{F}	frictional acceleration
F_i^m	components to the transformation matrix between material & position coordinates; also the deformation tensor
$G = G^{\text{grav}}$	Newton's gravitational constant: $G = 6.674 \times 10^{-11} \text{ N m}^2 \text{ kg}^{-2} = 6.674 \times 10^{-11} \text{ m}^3 \text{ kg}^{-1} \text{ s}^{-2}$
$G(\mathbf{x} \mathbf{x}_0)$	Green's function with \mathbf{x} the observation point (or field point) and \mathbf{x}_0 the source point
$G(\mathbf{x} \mathbf{x}_0)$	Free space Green's function; i.e., the Green's function without boundaries
g	effective gravitational acceleration at earth's surface from central gravity + planetary centrifugal: $g \approx 9.8 \text{ m s}^{-2}$
g^r	reduced gravity defined between two fluid layers; e.g., $g_1^r = g(\rho_2 - \rho_1)/\rho_1 \ll g$
h	layer thickness for a shallow water fluid
h	layer thickness for a continuously stratified fluid: $h = \bar{h} \delta\sigma$
h	specific thickness for a generalized vertical coordinate: $h = \partial z/\partial\sigma$
$\mathcal{H}(x)$	Heaviside step function: $\mathcal{H}(x) = 0$ for $x < 0$ whereas $\mathcal{H}(x) = 1$ for $x > 0$, and $\text{sgn}(x) = 2\mathcal{H}(x) - 1$
H	vertical depth scale
H	sometimes used as depth of the ocean bottom: $z = -H(x, y) = \eta_b(x, y)$
i	$i = \sqrt{-1}$ used for imaginary numbers
\mathbf{J}	tracer flux; for material tracers the dimensions are mass per time per area
K	kinetic energy for a point particle of mass m : $K = m \mathbf{V} \cdot \mathbf{V}/2$
\mathcal{K}	kinetic energy per mass of a fluid element: $\mathcal{K} = \mathbf{v} \cdot \mathbf{v}/2$
\mathbb{K}	diffusivity tensor; a positive semi-definite symmetric second order tensor
k	integer index to label a layer in a shallow water model with $k = 1, N$ layers ($k = 1$ is top layer)
k_B	Boltzmann constant: $k_B = 1.3806 \times 10^{-23} \text{ m}^2 \text{ kg s}^{-2} \text{ K}^{-1}$
L	horizontal length scale
L_d	deformation radius: (a) shallow water $L_d = \sqrt{gH}/f$; (b) continuously stratified internal $L_d = H N/f$
M_{air}	grams per mole of air: $M_{\text{air}} = 28.8 \times 10^{-3} \text{ kg mole}^{-1}$
p	pressure at a point in the fluid
p_a	pressure applied to the ocean surface or shallow water surface from, for example, the atmosphere
p_b	pressure at the bottom of a fluid column
p_{slp}	sea level pressure with an area average, $\langle p_{\text{slp}} \rangle = 101.325 \times 10^3 \text{ N m}^{-2}$
$p_{k-1/2}$	hydrostatic pressure at the layer interface with vertical position $z = \eta_{k-1/2}$ in a shallow water fluid
P_k	pressure integrated over a shallow water layer: $P_k \equiv \int_{\eta_{k+1/2}}^{\eta_{k-1/2}} p_k(z) dz = h_k(g \rho_k h_k/2 + p_{k-1/2})$
\mathcal{P}_k	potential energy for a shallow water fluid column: $\mathcal{P}_k = g \rho_k \int_{\eta_{k+1/2}}^{\eta_{k-1/2}} z dz = (g \rho_k/2)(\eta_{k-1/2}^2 - \eta_{k+1/2}^2)$
Q	potential vorticity either for a continuous fluid (Ertel PV) or shallow water fluid (Rossby PV)
q	quasi-geostrophic potential vorticity either for a continuous fluid or shallow water fluid
Q_m	mass flux per horizontal area crossing the ocean upper surface, with $Q_m > 0$ for mass entering ocean



ENGLISH-BASED SYMBOLS: PART II

SYMBOL	MEANING
r	radial position relative to an origin typically taken as center of a spherical earth
\mathbb{R}	real number line, also written \mathbb{R}^1
\mathbb{R}^2	two-dimensional Euclidean space of real numbers
\mathbb{R}^3	three-dimensional Euclidean space of real numbers
R	radius of a sphere
R_e	radius of sphere whose volume approximates that of the earth: $R_e = 6.371 \times 10^6 \text{ m}$
R_g	universal gas constant: $R_g = 8.314 \text{ J mole}^{-1} \text{ K}^{-1} = 8.314 \text{ kg m}^2 \text{ s}^{-2} \text{ mole}^{-1} \text{ K}^{-1}$
R^{air}	specific gas constant for air: $R^{\text{air}} = R_g/M_{\text{air}} = 2.938 \times 10^2 \text{ m}^2 \text{ s}^{-2} \text{ K}^{-1}$
\mathbb{S}	rate of strain tensor: $2 \mathbb{S}_{mn} = \partial_n v_m + \partial_m v_n$
\mathbb{S}^{dev}	deviatoric rate of strain tensor: $\mathbb{S}_{mn}^{\text{dev}} = \mathbb{S}_{mn} - \delta_{mn} \mathbb{S}_{qq}/3$
S	salt concentration = mass of salt in a fluid element per mass of seawater
S	salinity: $S = 1000 \mathbb{S}$
s	expression for a generic surface: $s = s(x, y, z, t)$.
s	arc-length along a curve $\mathbf{x}(s)$ with infinitesimal increment $ds = \sqrt{d\mathbf{x} \cdot d\mathbf{x}}$
\hat{s}	unit tangent vector to a curve, also written as $\hat{s} = \hat{\mathbf{t}}$ (see below)
T	absolute thermodynamic <i>in situ</i> temperature (in Kelvin when part of a thermodynamic equation)
\mathbb{T}	stress tensor
$\mathbb{T}^{\text{kinetic}}$	kinetic stress tensor: $\mathbb{T}^{\text{kinetic}} = -\rho \mathbf{v} \otimes \mathbf{v}$
$\mathbb{T}^{\text{sw kinetic}}$	kinetic stress tensor for shallow water fluid: $\mathbb{T}^{\text{sw kinetic}} = -\rho \mathbf{u} \otimes \mathbf{u}$
t	time (universal Newtonian time)
\hat{t}	unit tangent vector to a curve: $\hat{t} = d\mathbf{x}/ds$, where s is the arc-length so that $ds = \sqrt{d\mathbf{x} \cdot d\mathbf{x}}$
\mathbf{u}	horizontal velocity of a fluid element: $\mathbf{u} = \hat{\mathbf{x}} u + \hat{\mathbf{y}} v$
U	horizontal velocity scale
\mathbf{U}	depth integrated horizontal velocity: $\mathbf{U} = \int_{\eta_b}^{\eta} \mathbf{u} dz$
\mathbf{v}	velocity of a fluid element: $\mathbf{v} = D\mathbf{x}/Dt = \hat{\mathbf{x}} u + \hat{\mathbf{y}} v + \hat{\mathbf{z}} w$
W	vertical velocity scale
w	vertical component to the velocity: $w = Dz/Dt$
$w^{(\sigma)}$	dia-surface velocity = fluid volume per <i>horizontal area</i> per time crossing a σ surface: $w^{(\sigma)} = (\partial z/\partial \sigma) D\sigma/Dt$
w^{dia}	dia-surface flux = fluid volume per <i>surface area</i> per time crossing a σ -surface: $w^{\text{dia}} = (1/ \nabla \sigma) D\sigma/Dt$
(x, y, z)	triplet of Cartesian coordinates
\mathbf{x}	spatial position with Cartesian representation: $\mathbf{x} = \hat{\mathbf{x}} x + \hat{\mathbf{y}} y + \hat{\mathbf{z}} z$
$(\hat{\mathbf{x}}, \hat{\mathbf{y}}, \hat{\mathbf{z}})$	triplet of Cartesian unit vectors oriented in a righthand sense
z_σ	specific thickness for a generalized vertical coordinate: $z_\sigma = \partial z/\partial \sigma = h$



GREEK SYMBOLS

SYMBOL	MEANING
α	thermal expansion coefficient in terms of either potential or Conservative Temperature: $\alpha = -\rho^{-1} \partial \rho / \partial \theta$
β	saline contraction coefficient: $\beta = \rho^{-1} \partial \rho / \partial S$
β	meridional derivative of planetary vorticity: $\beta = \partial_y f$
δ_{ab}	components to the Kronecker tensor in Cartesian coordinates
δ^a_b	components to the Kronecker tensor in general coordinates
ϵ	kinetic energy dissipation per mass due to viscous effects
ϵ_{ab}	components to the permutation symbol in two dimensions
ϵ_{abc}	components to the permutation symbol in three dimensions
$\eta = \eta_0$	vertical position of the free upper surface of a fluid domain: $z = \eta(x, y, t)$
$\eta_{k-1/2}$	depth of the interface at the top of the shallow water layer k
$\eta_{k+1/2}$	depth of the interface at the bottom of the shallow water layer k
$\eta_b = -H$	vertical position of rigid and static lower solid-earth boundary of a fluid domain: $z = \eta_b(x, y) = -H(x, y)$
(λ, ϕ)	longitude and latitude position on the sphere: $0 \leq \lambda \leq 2\pi$ and $-\pi/2 \leq \phi \leq \phi/2$
κ	molecular kinematic diffusivity
κ^{eddy}	eddy kinematic diffusivity: $\kappa^{\text{eddy}} \gg \kappa$
μ_n	chemical potential for constituent n
μ	relative chemical potential for a binary fluid
μ_{vsc}	dynamic viscosity = $\rho \nu$
ν_s	specific volume: $\nu_s = \rho^{-1}$
ν	molecular kinematic viscosity
ν_{air}	molecular kinematic viscosity of air: $\nu_{\text{air}} \approx 1.3 \times 10^{-5} \text{ m}^2 \text{ s}^{-1}$
ν_{water}	molecular kinematic viscosity of fresh water: $\nu_{\text{water}} \approx 10^{-6} \text{ m}^2 \text{ s}^{-1}$
ν^{eddy}	eddy viscosity: $\nu^{\text{eddy}} \gg \nu$
ω	relative vorticity vector: $\omega = \nabla \wedge \mathbf{v}$
ω	radial frequency for a wave so that the wave period is $2\pi/\omega$
Ω	earth's angular velocity oriented through north pole with magnitude $\Omega = 7.2921 \times 10^{-5} \text{ s}^{-1}$
Φ	geopotential, which in its simplest form is $\Phi = g z$
ψ	streamfunction for two-dimensional incompressible flow: $\nabla \cdot \mathbf{u} = 0 \implies \mathbf{u} = \hat{z} \wedge \nabla \psi$
Ψ	vector streamfunction for three-dimensional incompressible flow: $\nabla \cdot \mathbf{v} = 0 \implies \mathbf{v} = \nabla \wedge \Psi$
ρ	<i>in situ</i> density (mass per volume) of a fluid element: $\rho = \rho(S, \theta, p)$
ρ_0	constant reference density used for the Boussinesq approximation
σ	an arbitrary surface where $\sigma(x, y, z, t) = \text{constant}$, commonly used for generalized vertical coordinates
τ	stress vector
θ	potential temperature
Θ	Conservative Temperature
ζ	vertical component to the relative vorticity, with Cartesian expression $\zeta = \partial_x v - \partial_y u$
ζ_a	vertical component to the absolute vorticity, with Cartesian expression $\zeta_a = f + \partial_x v - \partial_y u$



MATHEMATICAL OPERATIONS AND SYMBOLS

SYMBOL	MEANING
$[\equiv]$	read “has dimensions” and used when referring to the physical dimensions of an object
\wedge	vector cross product of vector calculus or wedge product of exterior calculus
∇	gradient operator that acts on a scalar
∇_z	horizontal gradient operator: $\nabla_z = \hat{\mathbf{x}} (\partial/\partial x) + \hat{\mathbf{y}} (\partial/\partial y) = \hat{\mathbf{x}} \partial_x + \hat{\mathbf{y}} \partial_y$
$\nabla \cdot$	divergence operator that acts on a vector to produce a scalar
$\nabla \wedge$	curl operator that acts on a vector to produce a vector
∇_σ	horizontal gradient operator acting on a constant σ -surface: $\nabla_\sigma = \hat{\mathbf{x}} (\partial/\partial x)_\sigma + \hat{\mathbf{y}} (\partial/\partial y)_\sigma$
∂_σ	vertical derivative operator using the generalized vertical coordinate: $\partial_\sigma = \partial/\partial\sigma = (\partial z/\partial\sigma) \partial/\partial z$
$\partial/\partial t$	Eulerian time derivative acting at a fixed spatial position, \mathbf{x}
$[\partial/\partial t]_\sigma$	time derivative computed on constant σ -surface
D/Dt	material, Lagrangian, or substantial time derivative following a fluid element
$d\Psi$	inexact differential
δ	differential increment often written for a field following the fluid motion
$\delta(x)$	one-dimensional Dirac delta distribution with dimensions inverse length
$\delta^{(2)}(\mathbf{x})$	two-dimensional Dirac delta distribution with dimensions inverse area
$\delta(\mathbf{x})$	three-dimensional Dirac delta distribution with dimensions inverse volume
$\delta(t)$	temporal Dirac delta distribution with dimensions inverse time
Δ	finite difference increment in space: $\Delta_x, \Delta_y, \Delta_z, \Delta_\sigma$
dA	infinitesimal horizontal area element: $dA = dx dy$
$d\mathcal{S}$	infinitesimal area element on a surface
dV	infinitesimal volume element
δV	infinitesimal volume for a region moving with the fluid (Lagrangian region)
$\int_{\mathcal{R}} dV$	integral over an arbitrary volume region, \mathcal{R}
$\int_{\mathcal{R}(\mathbf{v})} dV$	integral over a region that follows the fluid motion (Lagrangian integral)
$\int_{\mathcal{S}} d\mathcal{S}$	integral over a surface \mathcal{S}
$\oint_{\partial\mathcal{R}} d\mathcal{S}$	integral over a closed surface $\partial\mathcal{R}$ that bounds the volume \mathcal{R}
$\oint d\ell$	closed line integral over a periodic domain
$\oint_{\partial\mathcal{S}} d\ell$	counter-clockwise closed line integral over the boundary of a surface, $\partial\mathcal{S}$
\sim	similar to or scales as
\approx	approximately equal to
\dot{A}	time derivative following a particle trajectory: $\dot{A} = DA/Dt$.



Part I

Mathematical methods and concepts

Fluid mechanics is a field theory based on Newton's laws of mechanics and quasi-equilibrium thermodynamics as applied to a continuous fluid media. Geophysical fluid mechanics is concerned with gravitationally stratified fluids with multiple matter constituents moving on a rotating sphere. The speed for all motion is small relative to the speed of light so that relativistic effects are negligible, thus enabling the use of universal Newtonian time for all observers. Our goal for this part of the book is to review mathematical topics for use in our study of geophysical fluid mechanics. Depending on the reader's background and interests, this material can be readily skipped on first reading and/or accessed later to fill gaps.

PHYSICS PROVIDES RELATIONS BETWEEN GEOMETRIC OBJECTS

Mathematical objects of use for the study of fluid mechanics include scalar fields (e.g., temperature, mass density, specific entropy), vector fields (e.g., velocity, vorticity), and second order tensor fields (e.g., diffusion tensor, stress tensor, moment of inertia tensor). These and other fields have an existence independent of the arbitrary coordinate choices used for their description. Thinking abstractly, these physical fields are geometric objects such as points, vectors, surfaces, volumes, etc. In the study of geophysical fluid mechanics, we use physical principles to develop differential equations relating geometric objects. Mathematical tools are used to compute numbers as required to compare with experiments and field measurements, and to formulate discrete equations for numerical simulations.

Our perspective can be summarized by “physics as geometry”, which forms a foundation to theoretical physics such as that detailed in [Thorne and Blandford \(2017\)](#). It has both conceptual and practical use for our study. It furthermore provides the framework for this part of the book, in which we develop mathematical tools that are later used to formulate a variety of theoretical geophysical fluid models. Our overall aim is to develop mathematical tools to help unpack the physics encapsulated by equations. This aim extends to those cases where analytical solutions are unavailable, which is the norm for nonlinear field theories such as fluid mechanics or even in many cases when the equations are linearized. Such qualitative and conceptual tools are of great value for the analysis of numerical simulations and field measurements.

TENSOR ANALYSIS AND GEOPHYSICAL FLUID MECHANICS

There are many occasions where a geophysical fluid system is more physically transparent when using a particular coordinate description or reference frame. However, there is no *a priori* choice that fits all systems. Thus, being adept at transforming from one description to another eases the study. Tensor analysis is the proven means for systematically performing such transformations thus motivating its use throughout fluid mechanics. In its more abstract realization via differential forms, tensor analysis provides the means to mathematically express physical ideas without any display of coordinate artefacts, thus exposing the underlying physical and mathematical essence.

The following offers an incomplete list of geophysical fluid systems where various coordinate descriptions or reference frames are encountered, and thus where tensor analysis can be put to use. Granted, each system listed here can be studied without the formalism of tensor analysis. However, by doing so one often encounters clumsy manipulations that can obfuscate the underlying physical concepts. Indeed, imagine the tedium required to write continuous field equations in multiple dimensions prior to vector analysis! That situation is akin to the tedium and awkward nature required to work across multiple coordinate systems and reference frames absent the formalism of tensor analysis. Hence, an adept use of vector analysis, its generalization to tensor analysis, and its further abstraction to differential forms, reveals how maths can inform the physics and how physics can be transparently embodied by the maths.

- RELATING EULERIAN AND LAGRANGIAN KINEMATICS: There is a duality in fluid kinematics between Eulerian and Lagrangian descriptions of fluid motion. To develop an understanding of this duality we make use of tensor analysis to facilitate the transformation between the two kinematic descriptions.
- SPHERICAL PLANET: Geophysical fluids move on an approximately spherical planet, making spherical coordinates the preferred choice for studying planetary flows. We make use of tensor methods to transform between planetary Cartesian coordinates (origin at the center of the planet) and spherical coordinates.
- CYLINDRICAL ROTATING TANK: Rotating laboratory fluids move in a circular tank, with cylindrical polar coordinates of use to respect symmetry of the domain. We make use of tensor methods to transform between Cartesian and cylindrical polar coordinates when considering rotating tank systems.
- ROTATING REFERENCE FRAME: Geophysical fluids move around a rotating earth and the fluid maintains a close to solid-body motion. Terrestrial observers also move in near solid-body motion. We are thus motivated to study geophysical fluids from a rotating reference frame. We use rudimentary tensor methods to transform between a fixed inertial frame and the non-inertial rotating reference frame, with this transformation revealing non-inertial accelerations (Coriolis and centrifugal) that impact on the observed fluid flow.
- STRATIFIED FLUIDS AND GENERALIZED VERTICAL COORDINATES: Geophysical fluids move in a gravitational field that acts to stratify the fluid according to its local buoyancy. For many purposes it can be useful to describe the vertical position of a fluid element according to its buoyancy rather than its height. This “isopycnal” vertical coordinate choice leads to a non-orthogonal coordinate description of the fluid motion. There are other vertical coordinates that can be of use for other situations, such as the terrain following coordinates commonly used for atmospheric simulations and coastal ocean models. Transforming between Cartesian coordinates and such *generalized vertical coordinates* requires the mathematical precision of general tensors.
- WATER MASS ANALYSIS: We may use properties such as temperature, salinity, carbon concentration to define coordinates for a point within a water mass configuration space. This space has arbitrary dimensionality and has no natural metric structure, thus making the use of *differential forms* of use to expose the underlying mathematical structure of this space.

SUMMARY OF THE MATH CHAPTERS

Some of the chapters in this part of the book are essential for nearly all subsequent chapters, whereas others target aficionados and thus serve somewhat limited purposes. Readers are encouraged to take a close look at each chapter if only to know where to find topics that might be of use later in the book or later in one’s career.

- CARTESIAN TENSOR ALGEBRA: Chapter 1 is a synopsis of Cartesian tensor analysis. This topic provides a systematization of ideas from Cartesian geometry and linear algebra. Material in this chapter is essential for nearly every topic in this book.
- CARTESIAN TENSOR CALCULUS: Chapter 2 extends the algebraic ideas from Chapter 1 to differential and integral calculus. This chapter provides a resume of multivariate or vector calculus of use for fluid mechanics. Material in this chapter is essential for nearly every topic in this book.

- PARTIAL DIFFERENTIAL EQUATIONS: Chapter 3 provides a summary of linear partial differential equations (PDEs) commonly encountered in mathematical physics. Even though the equations of fluid mechanics are nonlinear PDEs, their linear counterparts offer much insight into fluid behavior.
- GREEN'S FUNCTIONS: Chapter 4 provides a survey of Green's function methods used to solve linear partial differential equations of mathematical physics. Although this book is less concerned with specific solutions to these equations, we study the Green's function method since it facilitates rather general insights into both the mathematical and physical content of fluid equations. Furthermore, Green's function methods are commonly used across various areas of the geophysical sciences, thus making it quite useful to be familiar to the basic notions.
- GEOMETRY OF CURVES AND SURFACES: Chapter 5 introduces rudimentary differential geometry used to characterize curves (such as fluid particle trajectories) and surfaces (such as isopycnals) embedded in a background Euclidean space. Here we encounter such notions as normal and tangent directions and curvature.
- GENERAL TENSOR INTRODUCTION: Chapter 6 provides an introduction to general tensor analysis and its applications to geophysical fluids. The discussion is accessible to anyone who has read Chapter 1, and is recommended for all readers of this book, even for those who do not wish to study the details of general tensors in Chapter 7.
- GENERAL TENSOR ANALYSIS: Chapter 7 extends the Cartesian tensor algebra and calculus to allow for the use of arbitrary, or general, coordinates. This chapter is essential to understand the mathematics underlying non-Cartesian coordinates, such as spherical coordinates and isopycnal coordinates.
- ORTHOGONAL COORDINATES: Chapter 8 offers a reference for the three locally orthogonal coordinate systems (Cartesian, spherical, cylindrical) used in this book and how various mathematical objects appear when written in these coordinates.
- DIFFERENTIAL FORMS: Chapter 9 introduces us to *differential forms*, which are totally anti-symmetrized tensor objects that arise in the study of calculus on topological manifolds that are not endowed with a metric. In physics, differential forms appear in areas such as Hamiltonian mechanics, thermodynamics, fluid mechanics, and field theory. We make use of differential forms in Section 52.12 to provide a mathematical foundation for water mass analysis.

Cartesian tensor algebra

In this chapter we introduce the formalism of Cartesian tensor analysis with a focus on the basic algebraic relations. The use of Cartesian tensors restricts attention to Cartesian coordinates on a flat Euclidean space, with all Cartesian coordinates related by orthogonal transformations representing rotations.

READER'S GUIDE TO THIS CHAPTER

We follow standard treatments of Cartesian tensors such as that in Chapter 2 of [Aris \(1962\)](#). The discussion should be accessible to those having studied undergraduate calculus and linear algebra. For geophysical fluid mechanics, mastery of Cartesian tensors is nearly sufficient for mastery of general tensors. This chapter is basic to all of the maths in this book.

1.1	Introduction to tensors and tensor fields	6
1.2	Points and vectors	6
1.3	Distance and the scalar product	8
1.3.1	Distance between points	8
1.3.2	Magnitude of a vector and the scalar product	8
1.4	Vector product	9
1.4.1	Basis vector orientation and the Levi-Civita tensor	9
1.4.2	Axial (or pseudo) vector	10
1.4.3	Orthogonality relations between cross products	10
1.4.4	Vector product of arbitrary vectors	11
1.4.5	Geometric interpretation of the vector product	11
1.4.6	Generalization to arbitrary vectors	12
1.5	Measuring volume	13
1.5.1	Volume defined by three vectors	13
1.5.2	Cartesian volume element for integration	14
1.5.3	n -space volumes and the Levi-Civita tensor	14
1.6	Example vector identities using the Levi-Civita tensor	14
1.6.1	Double vector product	15
1.6.2	Scalar product of two vector products	15
1.7	Transforming the coordinate representations	15
1.7.1	Inverse transformation	16
1.7.2	Orthogonal transformation	16
1.7.3	Geometric interpretation of orthogonal transformations	17
1.7.4	Transforming the coordinate representation of a vector	17
1.7.5	Invariance of the scalar product	17
1.7.6	Transforming the coordinate representation of a second order tensor	18

1.7.7	Importance of distinguishing between tensors and matrices	18
1.8	Homogeneity and isotropy	18
1.8.1	Homogeneous tensor	19
1.8.2	Isotropic and anisotropic tensors	19
1.9	Exercises	19

1.1 Introduction to tensors and tensor fields

Fluid mechanics involves fields of scalars, vectors, and higher order tensors. We generically refer to all of these geometric objects as *tensors*, with a scalar a zero order tensor and a vector a first order tensor. A scalar field provides a single number at each point in space-time. Example scalar fields include temperature, mass density, entropy, salinity, humidity, and mechanical energy. On a flat manifold, such as the Euclidean space considered in this chapter, a vector connects two points on the manifold and is specified by a direction and a magnitude. Correspondingly, a vector field provides a vector at each point in space. Example vector fields include the fluid velocity and forces acting on fluid elements. A second order tensor can be represented by a matrix, with a tensor field providing a matrix at each point in space. The stress tensor and the diffusion tensor are examples encountered in this book.

Geophysical fluids are embedded in the Newtonian universal time and flat Euclidean space. This space and time structure induces the familiar Euclidean norm when measuring the spatial distance between points, whether the points are on a plane, a sphere, or an arbitrary surface within the fluid such as a surface of constant specific entropy. We can thus make use of Cartesian coordinates as the starting point for a mathematical formulation of geophysical fluid mechanics. Transformations to alternative coordinates are made when they lend insight to the symmetry of the flow and/or the geometry of the space on which the flow occurs. It is for this reason that we devote this chapter to developing the formalism of Cartesian tensor analysis. Furthermore, it is for this reason that Cartesian tensors give us nearly all of the formalism necessary to study general tensors in geophysical fluid mechanics.

1.2 Points and vectors

Consider a point, \mathcal{P} , in three dimensional Euclidean space \mathbb{R}^3 . We can represent its spatial position by providing its Cartesian coordinates relative to an arbitrary origin. As such, the position is a vector whose tail is at the origin and head at the point as shown in Figure 1.1. We write this coordinate representation as

$$\mathcal{P} \mapsto \vec{P} = \hat{x} P_1 + \hat{y} P_2 + \hat{z} P_3. \quad (1.1)$$

Vectors are denoted by an arrow. The right hand side of equation (1.1) provides the representation of the position vector in terms of a triplet of Cartesian coordinates, (P_1, P_2, P_3) , that measure distance along their corresponding Cartesian unit vectors, $(\hat{x}, \hat{y}, \hat{z})$. The Cartesian unit vectors form a basis for three dimensional Euclidean space.¹ Hence, the position vector for any point in space can be represented in terms of these three basis vectors.

At times we find it useful to make use of alternative notations in which the position vector is

¹The unit vectors are sometimes denoted $(\hat{i}, \hat{j}, \hat{k})$ in the literature. We generally avoid that notation in this book.

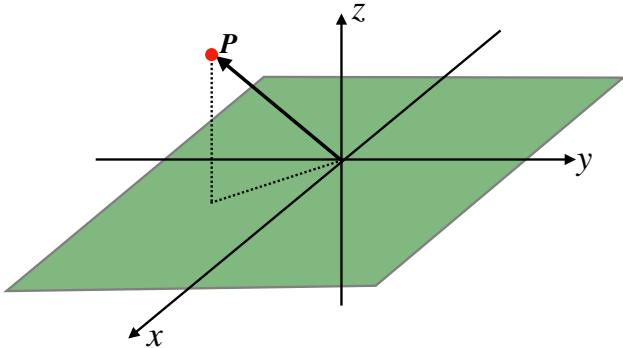


FIGURE 1.1: An arbitrary point in space, \mathcal{P} , has an objective existence independent of our subjective choice of coordinate system used to describe its position. We here represent its position with respect to the origin of a right-handed Cartesian coordinate system. The Cartesian representation of the position vector is $\mathcal{P} \mapsto \vec{P} = \mathbf{P} = \hat{\mathbf{x}} P_1 + \hat{\mathbf{y}} P_2 + \hat{\mathbf{z}} P_3$, with the Cartesian basis vectors the normalized triplet of unit vectors $(\hat{\mathbf{x}}, \hat{\mathbf{y}}, \hat{\mathbf{z}})$ and the triplet of Cartesian coordinates given by (P_1, P_2, P_3) . There are an infinite number of possible Cartesian coordinate systems that are rotated and/or translated with respect to the one shown here.

written

$$\vec{P} = \mathbf{P} \quad (1.2a)$$

$$= \hat{\mathbf{x}} P_1 + \hat{\mathbf{y}} P_2 + \hat{\mathbf{z}} P_3 \quad (1.2b)$$

$$= \sum_{a=1}^3 \vec{e}_a P_a \quad (1.2c)$$

$$= \vec{e}_a P_a. \quad (1.2d)$$

The first equality introduced the boldface notation, which we commonly use for the representation of vectors in Cartesian coordinates. Indeed, although less convenient when moving to the general tensors of Chapter 6, we still maintain the boldface in this book given its common usage in the literature and since much of this book makes use of Cartesian tensors. Equation (1.2d) introduced the Einstein summation convention in which repeated indices are summed over their range, thus allowing us to drop the summation symbol. Equation (1.2c) introduced a generic notation for the basis vectors

$$\vec{e}_1 = \hat{\mathbf{x}} \quad \vec{e}_2 = \hat{\mathbf{y}} \quad \vec{e}_3 = \hat{\mathbf{z}}. \quad (1.3)$$

The Cartesian basis vectors each have unit magnitude (normalized), meaning that their scalar products yield²

$$\hat{\mathbf{x}} \cdot \hat{\mathbf{x}} = \hat{\mathbf{y}} \cdot \hat{\mathbf{y}} = \hat{\mathbf{z}} \cdot \hat{\mathbf{z}} = 1. \quad (1.4)$$

Use of a hat or carot symbol over a vector signifies that the vector is normalized to unity. For Cartesian coordinates we generally work with the normalized basis vectors (1.3). Furthermore, a normalized vector can change only through rotation since by definition it remains of unit norm (see Section 2.1.4). For the general tensors of Chapters 6, 8 and 43, we sometimes find it more convenient to work with unnormalized basis vectors. Hence, the basis vectors, \vec{e}_a , retain the arrow symbol rather than the carot since they are not always normalized.

We emphasize that the tensor labels denote components of a specific coordinate representation, P_a , as well as members from the set of basis vectors, \vec{e}_a . These labels are not to be confused with

²The Cartesian scalar (or dot or inner) product is discussed more formally in Section 1.3.

partial derivative operations.³ Finally, we sometimes write the vector components and basis vectors in the form of a list or triplet

$$P_a = (P_1, P_2, P_3) \quad \text{and} \quad \vec{e}_a = (\vec{e}_1, \vec{e}_2, \vec{e}_3) = (\hat{x}, \hat{y}, \hat{z}). \quad (1.5)$$

1.3 Distance and the scalar product

In defining a vector to have unit magnitude, we are assuming we know how to measure the magnitude of a vector. We here make this notion precise.

1.3.1 Distance between points

Consider two points in Euclidean space, \mathcal{P} and $\mathcal{P} + d\mathcal{P}$, separated by a small distance and specified by their respective position vectors

$$\mathcal{P} \mapsto \vec{P} = \vec{e}_a P_a \quad (1.6a)$$

$$\mathcal{P} + d\mathcal{P} \mapsto \vec{P} + d\vec{x} = \vec{e}_a (P_a + dx_a). \quad (1.6b)$$

Euclidean space is afforded a metric whereby the squared distance between two points is measured via Pythagoras' Theorem

$$[\text{distance}(\mathcal{P}, \mathcal{P} + d\mathcal{P})]^2 = (\vec{P} + d\vec{x} - \vec{P}) \cdot (\vec{P} + d\vec{x} - \vec{P}) \quad (1.7a)$$

$$= (\mathbf{P} + d\mathbf{x} - \mathbf{P}) \cdot (\mathbf{P} + d\mathbf{x} - \mathbf{P}) \quad (1.7b)$$

$$= dx_a dx_b (\vec{e}_a \cdot \vec{e}_b) \quad (1.7c)$$

$$= dx_a dx_b \delta_{ab} \quad (1.7d)$$

$$= dx_a dx_a \quad (1.7e)$$

$$= (dx_1)^2 + (dx_2)^2 + (dx_3)^2. \quad (1.7f)$$

To reach this result we introduced the components to the Kronecker delta tensor, which can be represented by the 3×3 identity matrix

$$\vec{e}_a \cdot \vec{e}_b = \delta_{ab} = \begin{bmatrix} \delta_{11} & \delta_{12} & \delta_{13} \\ \delta_{21} & \delta_{22} & \delta_{23} \\ \delta_{31} & \delta_{32} & \delta_{33} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}. \quad (1.8)$$

The Kronecker tensor provides the Cartesian coordinate representation of the *metric* for Euclidean space. The metric provides the means to measure the distance between points on a manifold, and how to measure the length of a vector. It thus allows us to *normalize* a vector to have unit magnitude, motivating the often used term *norm* rather than metric. In Section 7.1 we introduce alternative representations for the metric based on the use of non-Cartesian coordinates and non-Euclidean manifolds.

1.3.2 Magnitude of a vector and the scalar product

By defining the distance between two points, we in turn have a prescription for defining the squared magnitude of a vector

$$|\mathbf{P}|^2 = \mathbf{P} \cdot \mathbf{P} = P_a P_b (\vec{e}_a \cdot \vec{e}_b) = P_a P_a = (P_1)^2 + (P_2)^2 + (P_3)^2. \quad (1.9)$$

³In this book we eschew the notation where partial derivatives are denoted by a subscript.

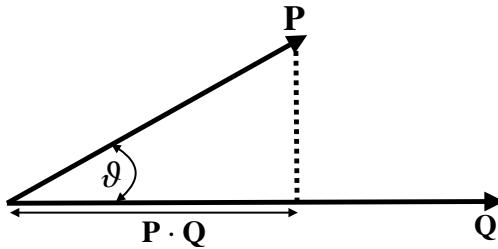


FIGURE 1.2: Illustrating the geometry associated with forming the scalar product between two vectors, $\mathbf{P} \cdot \mathbf{Q} = |\mathbf{P}| |\mathbf{Q}| \cos \vartheta$. Since the scalar or inner product requires a metric, we see that angles also require a metric.

Correspondingly, we have the *scalar product* between two arbitrary vectors

$$\mathbf{P} \cdot \mathbf{Q} = P_a Q_b (\vec{e}_a \cdot \vec{e}_b) = P_a Q_a, \quad (1.10)$$

which also goes by the name *dot product* or *inner product*. Given our expression for the scalar product and the magnitude of vectors, we can introduce a geometrical interpretation by defining the angle between the vectors according to

$$\cos \vartheta \equiv \frac{\mathbf{P} \cdot \mathbf{Q}}{|\mathbf{P}| |\mathbf{Q}|} = \frac{P_a Q_a}{\sqrt{P_b P_b} \sqrt{Q_c Q_c}}. \quad (1.11)$$

We illustrate this equation in Figure 1.2. It is useful to convince oneself that this definition is consistent with $-1 \leq \cos \vartheta \leq 1$. Furthermore, note that this definition of an angle between two vectors makes use of the metric tensor, so that angles on a manifold, just as distance, require a metric.

1.4 Vector product

The scalar product provides a means to measure the magnitude of a vector and the distance between two points. We here introduce the vector (or cross) product, which provides a means to measure area associated with two vectors and to specify the orientation of that area.

1.4.1 Basis vector orientation and the Levi-Civita tensor

Consider a flat plane defined by any two of the Cartesian basis vectors, \vec{e}_a and \vec{e}_b . We seek a means to specify what side of the plane is up and what side is down. Doing so allows us to orient objects within space.⁴ Notably, there is no objective means for this specification, since “up” and “down” are subject to our chosen orientation. Therefore, we must choose a convention. For that purpose, we follow the *right hand rule*, in which the out-stretched thumb, index, and middle fingers of the right hand orient the three Cartesian basis vectors.

We algebraically specify the right hand rule for the basis vectors through the relation⁵

$$\vec{e}_a \wedge \vec{e}_b = \epsilon_{abc} \vec{e}_c. \quad (1.12)$$

⁴There are surfaces, such as the Möbius strip, that are not orientable. We only consider orientable surfaces in this book.

⁵Many authors choose the symbol \times for the vector product rather than the wedge symbol, \wedge . The wedge is used here as it lends itself to less confusion with the coordinate x , particularly when written on a chalkboard. We also prefer the wedge symbol since it provides a connection to the exterior product of exterior algebra to be discussed in Chapter 9.

The left hand side introduces the vector (or cross) product of two basis vectors. The right hand side algebraically defines the vector product as the *contraction* of the Levi-Civita tensor with another basis vector. Here, the term contraction can be considered a generalization of the inner product. The Cartesian components of the Levi-Civita tensor are given by the totally anti-symmetric permutation symbol or ϵ -tensor

$$\epsilon_{123} = 1 \quad (1.13a)$$

$$\epsilon_{abc} = \begin{cases} 1, & \text{even permutation of } abc \text{ (123, 312, 231)} \\ -1, & \text{odd permutation of } abc \text{ (321, 132, 213)} \\ 0, & \text{all other } abc. \end{cases} \quad (1.13b)$$

Exchanging indices (an odd permutation) flips the sign of the permutation symbol

$$\epsilon_{abc} = -\epsilon_{bac} = -\epsilon_{acb}, \quad (1.14)$$

whereas cycling indices (an even permutation) preserves the sign

$$\epsilon_{abc} = \epsilon_{cab} = \epsilon_{bca}. \quad (1.15)$$

Correspondingly, we have

$$\vec{e}_a \wedge \vec{e}_b = \epsilon_{abc} \vec{e}_c = -\vec{e}_b \wedge \vec{e}_a. \quad (1.16)$$

1.4.2 Axial (or pseudo) vector

The vector cross product defines a vector whose properties are dependent on the orientation of the space. That is, the vector cross product changes its sign upon changing from a right handed to left handed orientation of the coordinate system. Such vectors are known as *axial vectors* or *pseudo vectors*. Example axial vectors that we encounter in this book include the angular velocity for the planet as well as the vorticity of a fluid.

1.4.3 Orthogonality relations between cross products

As defined, the permutation symbol ensures that $\vec{e}_a \wedge \vec{e}_b$ is orthogonal to both \vec{e}_a and \vec{e}_b

$$\vec{e}_a \cdot (\vec{e}_a \wedge \vec{e}_b) = \vec{e}_a \cdot \epsilon_{abc} \vec{e}_c \quad \text{definition} \quad (1.17a)$$

$$= \epsilon_{abc} \vec{e}_a \cdot \vec{e}_c \quad \text{rearrangement} \quad (1.17b)$$

$$= \epsilon_{cba} \vec{e}_c \cdot \vec{e}_a \quad \text{relabel } a \text{ to } c \text{ and } c \text{ to } a \quad (1.17c)$$

$$= -\epsilon_{abc} \vec{e}_c \cdot \vec{e}_a \quad cba \text{ is an odd permutation of } abc \quad (1.17d)$$

$$\Rightarrow \vec{e}_a \cdot (\vec{e}_a \wedge \vec{e}_b) = 0. \quad (1.17e)$$

To fully digest step (1.17c) it can be useful to reintroduce the summation symbol so that

$$\epsilon_{abc} \vec{e}_a \cdot \vec{e}_c = \sum_{a=1}^3 \sum_{c=1}^3 \epsilon_{abc} \vec{e}_a \cdot \vec{e}_c \quad \text{summation symbols reintroduced} \quad (1.18a)$$

$$= \sum_{c=1}^3 \sum_{a=1}^3 \epsilon_{cba} \vec{e}_c \cdot \vec{e}_a \quad \text{swap } a \text{ and } c \quad (1.18b)$$

$$= \epsilon_{cba} \vec{e}_c \cdot \vec{e}_a \quad \text{reintroduce summation convention} \quad (1.18c)$$

Additionally, to digest step (1.17d) we step through the permutations

$$\epsilon_{cba} = -\epsilon_{bca} \quad \text{swap } c \text{ with } b \text{ to pick up a minus sign} \quad (1.19a)$$

$$= \epsilon_{bac} \quad \text{swap } c \text{ with } a \text{ to pick up a minus sign} \quad (1.19b)$$

$$= -\epsilon_{abc} \quad \text{swap } b \text{ with } a \text{ to pick up a minus sign.} \quad (1.19c)$$

The same procedure shows that $\vec{e}_b \cdot (\vec{e}_a \wedge \vec{e}_b) = 0$. Hence, the vector product is orthogonal to the plane specified by any two of the basis vectors. That is, the vector product points orthogonal to that plane and in a direction determined by the right hand rule. We note that this proof reveals a general property. Namely, the contraction of a symmetric tensor (e.g., the scalar product $\vec{e}_c \cdot \vec{e}_a$) with an anti-symmetric tensor results in zero. We further illustrate this property in Exercise 1.2.

1.4.4 Vector product of arbitrary vectors

The expression (1.12) for the vector product of two basis vectors renders the vector product of arbitrary vectors.

$$\mathbf{P} \wedge \mathbf{Q} = P_a \vec{e}_a \wedge Q_b \vec{e}_b \quad (1.20a)$$

$$= P_a Q_b \vec{e}_a \wedge \vec{e}_b \quad (1.20b)$$

$$= P_a Q_b \epsilon_{abc} \vec{e}_c \quad (1.20c)$$

$$= (P_2 Q_3 - P_3 Q_2) \vec{e}_1 + (P_3 Q_1 - P_1 Q_3) \vec{e}_2 + (P_1 Q_2 - P_2 Q_1) \vec{e}_3. \quad (1.20d)$$

We can write the vector product as a determinant

$$\mathbf{P} \wedge \mathbf{Q} = \det \begin{bmatrix} \vec{e}_1 & \vec{e}_2 & \vec{e}_3 \\ p_1 & p_2 & p_3 \\ q_1 & q_2 & q_3 \end{bmatrix}. \quad (1.21)$$

As with the basis vectors, the vector product is orthogonal to both of the individual vectors

$$\mathbf{P} \cdot (\mathbf{P} \wedge \mathbf{Q}) = (P_d \vec{e}_d) \cdot (P_a Q_b \epsilon_{abc} \vec{e}_c) \quad (1.22a)$$

$$= P_c P_a Q_b \epsilon_{abc} \quad (1.22b)$$

$$= 0, \quad (1.22c)$$

where the final equality follows since the product $P_c P_a$ is symmetric on the labels ac , whereas ϵ_{abc} is anti-symmetric.

1.4.5 Geometric interpretation of the vector product

The expression (1.20d) leads to the identity

$$|\mathbf{P} \wedge \mathbf{Q}|^2 = |\mathbf{P}|^2 |\mathbf{Q}|^2 - (\mathbf{P} \cdot \mathbf{Q})^2 \quad (1.23a)$$

$$= |\mathbf{P}|^2 |\mathbf{Q}|^2 (1 - \cos^2 \vartheta), \quad (1.23b)$$

$$= |\mathbf{P}|^2 |\mathbf{Q}|^2 \sin^2 \vartheta, \quad (1.23c)$$

where we used the scalar product expression (1.11) to introduce the angle subtended by the two vectors. Trigonometry indicates that the area of the parallelogram defined by \mathbf{P} and \mathbf{Q} is given by $|\mathbf{P}| |\mathbf{Q}| \sin \vartheta$. Hence, the vector product has a magnitude given by this area

$$\text{area}(\mathbf{P}, \mathbf{Q}) = |\mathbf{P}| |\mathbf{Q}| \sin \vartheta = |\mathbf{P} \wedge \mathbf{Q}|. \quad (1.24)$$

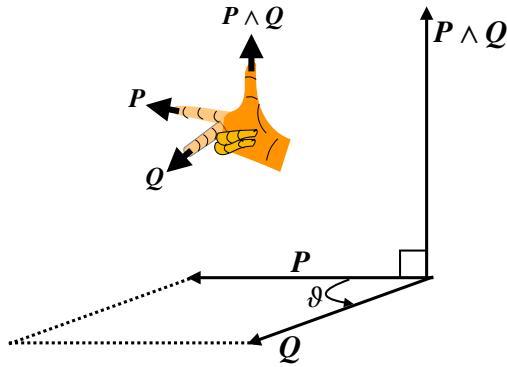


FIGURE 1.3: The magnitude of the vector product between two vectors is given by the product of their magnitudes and the sine of the angle between them, $|P \wedge Q| = |P||Q|\sin\vartheta$. This magnitude equals to the area of the parallelogram subtended by the two vectors. The vector product is directed perpendicular to the plane determined by the two vectors and oriented according to the right hand rule. We depict the right hand rule with the insert, whereby the index finger orients the first vector, the middle finger the second vector, and the thumb orients their vector product.

Since $P \wedge Q$ is orthogonal to the plane defined by P and Q , we can write the vector product in the purely geometric manner

$$P \wedge Q = \hat{n} \text{area}(P, Q) = \hat{n} |P||Q|\sin\vartheta, \quad (1.25)$$

where \hat{n} is a unit vector pointing normal to the area and in a direction given by the right hand rule. This formula is illustrated in Figure 1.3.

To further emphasize the geometric interpretation, let P equal to the vertical direction, $P = \hat{z}$. The vector cross product then defines a vector,

$$Q^{\hat{z}\perp} = \hat{z} \wedge Q \quad (1.26a)$$

$$= \hat{z} \wedge [Q - (\hat{z} \cdot Q)\hat{z}] \quad (1.26b)$$

$$(\hat{z} \wedge \hat{x})(\hat{x} \cdot Q) + (\hat{z} \wedge \hat{y})(\hat{y} \cdot Q) \quad (1.26c)$$

$$= \hat{y}(\hat{x} \cdot Q) - \hat{x}(\hat{y} \cdot Q). \quad (1.26d)$$

By construction, $Q^{\hat{z}\perp}$ is in the horizontal plane and it is perpendicular to the horizontal projection of Q

$$Q^{\hat{z}\perp} \cdot \hat{z} = 0 \quad \text{and} \quad Q^{\hat{z}\perp} \cdot Q = 0 \implies Q^{\hat{z}\perp} \cdot [Q - (\hat{z} \cdot Q)\hat{z}] = 0. \quad (1.27)$$

Hence, the vector $Q^{\hat{z}\perp}$ is geometrically computed by rotating the horizontal component of Q by $\pi/2$ radians counter-clockwise about the \hat{z} axis. That interpretation holds for all coordinate directions so that

$$\hat{x} \wedge Q = \hat{z}(\hat{y} \cdot Q) - \hat{y}(\hat{z} \cdot Q) = \text{project } Q \text{ to y-z plane} + \text{rotate CCW by } \pi/2 \text{ around } \hat{x}. \quad (1.28a)$$

$$\hat{y} \wedge Q = \hat{x}(\hat{z} \cdot Q) - \hat{z}(\hat{x} \cdot Q) = \text{project } Q \text{ to z-x plane} + \text{rotate CCW by } \pi/2 \text{ around } \hat{y}. \quad (1.28b)$$

$$\hat{z} \wedge Q = \hat{y}(\hat{x} \cdot Q) - \hat{x}(\hat{y} \cdot Q) = \text{project } Q \text{ to x-y plane} + \text{rotate CCW by } \pi/2 \text{ around } \hat{z}. \quad (1.28c)$$

1.4.6 Generalization to arbitrary vectors

Thus far the discussion has considered vectors to represent the position of a point in space. As such, the vectors have the physical dimensions of length and $\text{area}(P, Q)$ has dimensions of area. However, the vector analysis is general, so that the above notions extend to vectors of arbitrary physical dimensions, such as velocity. In these more general cases the physical dimensions must be adjusted accordingly.

1.5 Measuring volume

The vector product offers a means to measure area defined by two vectors. We now extend that result to measure the volume determined by three non-parallel vectors. This result has particular relevance to the volume element used for integration over space.

1.5.1 Volume defined by three vectors

Consider the scalar product of an arbitrary vector with the vector product, $(\mathbf{P} \wedge \mathbf{Q}) \cdot \mathbf{R}$. This scalar product projects that portion of the vector \mathbf{R} onto the direction parallel to the normal to the plane defined by $\mathbf{P} \wedge \mathbf{Q}$. Given that $|(\mathbf{P} \wedge \mathbf{Q})|$ is the area of the parallelogram defined by \mathbf{P} and \mathbf{Q} , we conclude that $(\mathbf{P} \wedge \mathbf{Q}) \cdot \mathbf{R}$ is the volume of the parallelepiped defined by the three vectors. However, note that this volume is not positive definite since the sign depends on the relative orientation of $\mathbf{P} \wedge \mathbf{Q}$ and \mathbf{R} . So more precisely, we need to apply an absolute value around the triple product to get the volume.

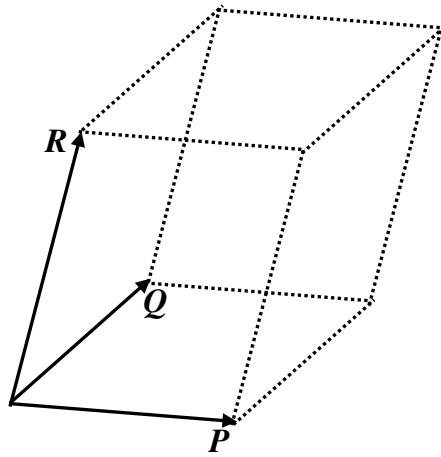


FIGURE 1.4: Three linearly independent position vectors determine a volume given by $|(\mathbf{P} \wedge \mathbf{Q}) \cdot \mathbf{R}| = |(\mathbf{R} \wedge \mathbf{P}) \cdot \mathbf{Q}| = |(\mathbf{Q} \wedge \mathbf{R}) \cdot \mathbf{P}|$

We can prove cyclic symmetry of $(\mathbf{P} \wedge \mathbf{Q}) \cdot \mathbf{R}$ through the following manipulations

$$(\mathbf{P} \wedge \mathbf{Q}) \cdot \mathbf{R} = (P_a \vec{e}_a \wedge Q_b \vec{e}_b) \cdot R_d \vec{e}_d \quad (1.29a)$$

$$= P_a Q_b (\vec{e}_a \wedge \vec{e}_b) \cdot R_d \vec{e}_d \quad (1.29b)$$

$$= P_a Q_b (\epsilon_{abc} \vec{e}_c) \cdot \vec{e}_d R_d \quad (1.29c)$$

$$= P_a Q_b \epsilon_{abc} (\vec{e}_c \cdot \vec{e}_d) R_d \quad (1.29d)$$

$$= P_a Q_b \epsilon_{abc} \delta_{cd} R_d \quad (1.29e)$$

$$= P_a Q_b \epsilon_{abc} R_c \quad (1.29f)$$

$$= R_c P_a Q_b \epsilon_{abc} \quad (1.29g)$$

$$= R_a P_b Q_c \epsilon_{bca} \quad (1.29h)$$

$$= R_a P_b Q_c \epsilon_{abc} \quad (1.29i)$$

$$= (\mathbf{R} \wedge \mathbf{P}) \cdot \mathbf{Q}. \quad (1.29j)$$

This identity yields the geometric result illustrated in Figure 1.4

$$\text{volume}(\mathbf{P}, \mathbf{Q}, \mathbf{R}) = |(\mathbf{P} \wedge \mathbf{Q}) \cdot \mathbf{R}| = |(\mathbf{R} \wedge \mathbf{P}) \cdot \mathbf{Q}| = |(\mathbf{Q} \wedge \mathbf{R}) \cdot \mathbf{P}|. \quad (1.30)$$

1.5.2 Cartesian volume element for integration

We need the volume of an infinitesimal region when performing an integration over space. When making use of Cartesian coordinates we need the volume of a rectangular prism defined by infinitesimal distances along each of the axes. We thus set

$$\mathbf{P} = \hat{\mathbf{x}} dx \quad \text{and} \quad \mathbf{Q} = \hat{\mathbf{y}} dy \quad \text{and} \quad \mathbf{R} = \hat{\mathbf{z}} dz, \quad (1.31)$$

in which case the volume element is

$$dV = (\mathbf{P} \wedge \mathbf{Q}) \cdot \mathbf{R} = dx dy dz (\hat{\mathbf{x}} \wedge \hat{\mathbf{y}}) \cdot \hat{\mathbf{z}} = dx dy dz. \quad (1.32)$$

This expression for the volume element could have been written down without the formalism of a vector triple product. However, in Chapter 6 we find the general relation $(\mathbf{P} \wedge \mathbf{Q}) \cdot \mathbf{R}$ provides a useful starting point for deriving the volume element with arbitrary coordinates.

1.5.3 n -space volumes and the Levi-Civita tensor

We combine the geometric specification of the vector product as a means to measure area, (1.25), with the algebraic specification (1.20d) by writing

$$\text{2-volume} = \epsilon(\mathbf{P}, \mathbf{Q}) = \epsilon_{ab} P_a Q_b = \det \begin{bmatrix} P_1 & Q_1 \\ P_2 & Q_2 \end{bmatrix}. \quad (1.33)$$

In this equation, ϵ_{ab} is the totally anti-symmetric 2×2 tensor. It has Cartesian components that can be organized as a matrix according to

$$\begin{bmatrix} \epsilon_{11} & \epsilon_{12} \\ \epsilon_{21} & \epsilon_{22} \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}. \quad (1.34)$$

In words, the first equality in equation (1.33) states that the ϵ -tensor in two dimensions takes two vectors as its argument and produces a 2-volume (i.e., an area). The three dimensional generalization yields

$$\text{3-volume} = \epsilon(\mathbf{P}, \mathbf{Q}, \mathbf{R}) = \epsilon_{abc} P_a Q_b R_c = \det \begin{bmatrix} P_1 & Q_1 & R_1 \\ P_2 & Q_2 & R_2 \\ P_3 & Q_3 & R_3 \end{bmatrix}. \quad (1.35)$$

Suppressing the first vector argument in the 3-volume produces a vectorial surface area defined by the other two vectors

$$\text{surface area} = \epsilon(, \mathbf{Q}, \mathbf{R}). \quad (1.36)$$

By construction, the vectorial surface area is orthogonal to both \mathbf{Q} and \mathbf{R} .

1.6 Example vector identities using the Levi-Civita tensor

The Levi-Civita tensor is a versatile tool for deriving vector identities. We illustrated some of these features in the previous discussion and here illustrate two more. These examples, and others in this chapter, generally expose many of the *index gymnastics* involved with tensor manipulations. A bit of practice and confidence readily allows one to suppress many of these steps.

1.6.1 Double vector product

Consider the double vector product

$$\mathbf{P} \wedge (\mathbf{Q} \wedge \mathbf{R}) = P_a Q_b R_c \vec{e}_a \wedge (\vec{e}_b \wedge \vec{e}_c) \quad (1.37a)$$

$$= P_a Q_b R_c \vec{e}_a \wedge (\epsilon_{bcd} \vec{e}_d) \quad (1.37b)$$

$$= P_a Q_b R_c \epsilon_{bcd} \epsilon_{ade} \vec{e}_e \quad (1.37c)$$

$$= -P_a Q_b R_c \epsilon_{bcd} \epsilon_{aed} \vec{e}_e. \quad (1.37d)$$

Explicit substitution verifies that the product $\epsilon_{bcd} \epsilon_{aed}$ equals to

$$\epsilon_{bcd} \epsilon_{aed} = \delta_{ba} \delta_{ce} - \delta_{be} \delta_{ca}. \quad (1.38)$$

This identity then leads to

$$\epsilon_{bcd} \epsilon_{aed} \vec{e}_e = \delta_{ba} \vec{e}_c - \delta_{ca} \vec{e}_b \quad (1.39)$$

so that

$$\mathbf{P} \wedge (\mathbf{Q} \wedge \mathbf{R}) = -P_a Q_b R_c (\delta_{ba} \vec{e}_c - \delta_{ca} \vec{e}_b) \quad (1.40a)$$

$$= -(\mathbf{P} \cdot \mathbf{Q}) \mathbf{R} + (\mathbf{P} \cdot \mathbf{R}) \mathbf{Q}. \quad (1.40b)$$

1.6.2 Scalar product of two vector products

We make further use of the Levi-Civita identity (1.38) to write

$$(\mathbf{P} \wedge \mathbf{Q}) \cdot (\mathbf{R} \wedge \mathbf{S}) = (\epsilon_{abc} p_a q_b) (\epsilon_{dec} R_d s_e) \quad (1.41a)$$

$$= p_a q_b R_d s_e \epsilon_{abc} \epsilon_{dec} \quad (1.41b)$$

$$= p_a q_b R_d s_e (\delta_{ad} \delta_{be} - \delta_{ae} \delta_{bd}) \quad (1.41c)$$

$$= (\mathbf{P} \cdot \mathbf{R}) (\mathbf{Q} \cdot \mathbf{S}) - (\mathbf{P} \cdot \mathbf{S}) (\mathbf{Q} \cdot \mathbf{R}). \quad (1.41d)$$

1.7 Transforming the coordinate representations

The Cartesian basis vectors are mutually orthogonal and fixed in space. However, the orientation of the basis vectors is arbitrary. We thus consider an alternative specification to the basis vectors by performing a linear transformation

$$\vec{e}_{\bar{a}} = \mathcal{R}_{\bar{a}b} \vec{e}_b. \quad (1.42)$$

This expression introduced components to the transformation matrix moving between the unbarred and the barred Cartesian coordinates

$$\mathcal{R}_{\bar{a}b} = \begin{bmatrix} \mathcal{R}_{\bar{1}1} & \mathcal{R}_{\bar{1}2} & \mathcal{R}_{\bar{1}3} \\ \mathcal{R}_{\bar{2}1} & \mathcal{R}_{\bar{2}2} & \mathcal{R}_{\bar{2}3} \\ \mathcal{R}_{\bar{3}1} & \mathcal{R}_{\bar{3}2} & \mathcal{R}_{\bar{3}3} \end{bmatrix}. \quad (1.43)$$

In Cartesian tensor analysis studied in this chapter, the transformation matrix is assumed to be independent of space.⁶ Although the transformation matrix carries two indices, it is not a tensor. Instead, it is a matrix operator used to transform from one set of basis vectors to another. We now deduce some constraints on this transformation matrix.

⁶The transformation matrix is a function of space and time for the general tensors considered in Chapter 7.

1.7.1 Inverse transformation

Assuming the transformation is invertible leads to the inverse transformation

$$\vec{e}_a = (\mathcal{R}^{-1})_{a\bar{b}} \vec{e}_{\bar{b}}. \quad (1.44)$$

As a self-consistency check we combine this relation with equation (1.42) thus rendering

$$\vec{e}_a = (\mathcal{R}^{-1})_{a\bar{b}} \vec{e}_{\bar{b}} = (\mathcal{R})_{a\bar{b}}^{-1} (\mathcal{R}_{\bar{b}c} \vec{e}_c). \quad (1.45)$$

This relation holds since

$$(\mathcal{R}^{-1})_{a\bar{b}} \mathcal{R}_{\bar{b}c} = \delta_{ac}, \quad (1.46)$$

or as a matrix identity

$$\mathcal{R}^{-1} \mathcal{R} = I. \quad (1.47)$$

1.7.2 Orthogonal transformation

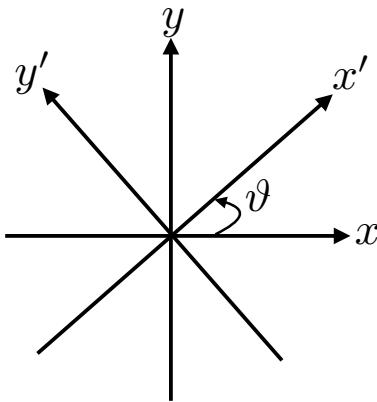


FIGURE 1.5: Counter-clockwise rotation of horizontal Cartesian axes through an angle ϑ .

We now assume that the two sets of Cartesian basis vectors are orthonormal. That assumption leads to the following constraint on the transformation matrix

$$\delta_{\bar{a}\bar{b}} = \vec{e}_{\bar{a}} \cdot \vec{e}_{\bar{b}} \quad (1.48a)$$

$$= \mathcal{R}_{\bar{a}a} \vec{e}_a \cdot \mathcal{R}_{\bar{b}b} \vec{e}_b \quad (1.48b)$$

$$= \mathcal{R}_{\bar{a}a} \mathcal{R}_{\bar{b}b} \vec{e}_a \cdot \vec{e}_b \quad (1.48c)$$

$$= \mathcal{R}_{\bar{a}a} \mathcal{R}_{\bar{b}b} \delta_{ab} \quad (1.48d)$$

$$= \mathcal{R}_{\bar{a}a} \mathcal{R}_{\bar{b}a} \quad (1.48e)$$

$$= \mathcal{R}_{\bar{a}a} (\mathcal{R}^T)_{a\bar{b}}, \quad (1.48f)$$

where \mathcal{R}^T is the matrix transpose with components

$$(\mathcal{R}^T)_{a\bar{b}} = \mathcal{R}_{\bar{b}a}. \quad (1.49)$$

Written as a matrix equation we see that

$$\mathcal{R} \mathcal{R}^T = I. \quad (1.50)$$

This relation defines an *orthogonal transformation*, whereby the inverse matrix equals to the matrix transpose

$$\mathcal{R}^{-1} = \mathcal{R}^T. \quad (1.51)$$

1.7.3 Geometric interpretation of orthogonal transformations

Orthogonal transformations convert one set of Cartesian coordinates to another. Geometrically, an orthogonal transformation corresponds to a rotation, with Figure 1.5 illustrating this axis rotation in two dimensions. For this two dimensional example, the rotation matrix can be written in terms of the cosine of the angles between the unit vectors; i.e., the *direction cosines*

$$\mathcal{R}_{\bar{a}a} = \begin{bmatrix} \cos \vartheta & \sin \vartheta \\ -\sin \vartheta & \cos \vartheta \end{bmatrix} = \begin{bmatrix} \cos \vartheta & \cos(\pi/2 - \vartheta) \\ \cos(\pi/2 + \vartheta) & \cos \vartheta \end{bmatrix} = \begin{bmatrix} \vec{e}_1 \cdot \vec{e}_1 & \vec{e}_1 \cdot \vec{e}_2 \\ \vec{e}_2 \cdot \vec{e}_1 & \vec{e}_2 \cdot \vec{e}_2 \end{bmatrix}. \quad (1.52)$$

The final form of the rotation matrix reveals that it is built from the projection of the rotated basis vectors onto the original basis vectors. This result holds for rotations in three dimensions as well, thus leading to

$$\mathcal{R}_{\bar{a}a} = \begin{bmatrix} \vec{e}_1 \cdot \vec{e}_1 & \vec{e}_1 \cdot \vec{e}_2 & \vec{e}_1 \cdot \vec{e}_3 \\ \vec{e}_2 \cdot \vec{e}_1 & \vec{e}_2 \cdot \vec{e}_2 & \vec{e}_2 \cdot \vec{e}_3 \\ \vec{e}_3 \cdot \vec{e}_1 & \vec{e}_3 \cdot \vec{e}_2 & \vec{e}_3 \cdot \vec{e}_3 \end{bmatrix}. \quad (1.53)$$

In summary, Cartesian tensor analysis considers arbitrary Cartesian coordinates as related through a rotation matrix built from the *direction cosines*.

1.7.4 Transforming the coordinate representation of a vector

We introduced the transformation (1.44) according to how it acts on the basis vectors. Now consider how it acts on the coordinate representation of an arbitrary vector by moving brackets

$$\mathbf{P} = P_a \vec{e}_a \quad (1.54a)$$

$$= P_a (\mathcal{R}^T)_{a\bar{a}} \vec{e}_{\bar{a}} \quad (1.54b)$$

$$\equiv P_{\bar{a}} \vec{e}_{\bar{a}}, \quad (1.54c)$$

where we defined the transformation of the vector components

$$P_{\bar{a}} = P_a (\mathcal{R}^T)_{a\bar{a}} = \mathcal{R}_{\bar{a}a} P_a. \quad (1.55)$$

1.7.5 Invariance of the scalar product

The above properties of an orthogonal transformation ensure that the scalar product

$$\mathbf{P} \cdot \mathbf{Q} = P_a Q_a \quad (1.56)$$

is the same regardless the choice of Cartesian coordinates

$$\mathbf{P} \cdot \mathbf{Q} = P_a Q_a \quad (1.57a)$$

$$= P_a \vec{e}_a \cdot Q_b \vec{e}_b \quad (1.57b)$$

$$= P_a (\mathcal{R}^T)_{a\bar{a}} Q_b (\mathcal{R}^T)_{b\bar{b}} (\vec{e}_{\bar{a}} \cdot \vec{e}_{\bar{b}}) \quad (1.57c)$$

$$= P_a (\mathcal{R}^T)_{a\bar{a}} Q_b (\mathcal{R}^T)_{b\bar{b}} \delta_{\bar{a}\bar{b}} \quad (1.57d)$$

$$= P_a (\mathcal{R}^T)_{a\bar{a}} Q_b (\mathcal{R}^T)_{b\bar{a}} \quad (1.57e)$$

$$= P_{\bar{a}} Q_{\bar{a}}. \quad (1.57f)$$

1.7.6 Transforming the coordinate representation of a second order tensor

The stress tensor in Chapter 18; the moment of inertia tensor in Section 37.4.3; and the diffusion tensor in Chapter 48 are examples of second order tensors. Second order tensors have a coordinate representation given by

$$\mathcal{T} = T_{ab} \vec{e}_a \otimes \vec{e}_b, \quad (1.58)$$

with T_{ab} the Cartesian coordinates that can be organized as a matrix. Notably, there is not a scalar product between the basis vectors in equation (1.58). Instead, the \otimes symbol represents the *outer product* of the two basis vectors. For our purposes the outer product simply provides a formal means for two vectors to be positioned next to one another. For many purposes we drop the \otimes symbol for brevity.

We determine how the Cartesian coordinate components T_{ab} transform by following the now familiar procedure for transforming the basis vectors. The key new facet is that now we have two basis vectors to carry around

$$\mathcal{T} = T_{ab} \vec{e}_a \vec{e}_b \quad (1.59a)$$

$$= T_{ab} (\mathcal{R}^T)_{a\bar{a}} \vec{e}_{\bar{a}} (\mathcal{R}^T)_{b\bar{b}} \vec{e}_{\bar{b}} \quad (1.59b)$$

$$= T_{ab} (\mathcal{R}^T)_{a\bar{a}} (\mathcal{R}^T)_{b\bar{b}} \vec{e}_{\bar{a}} \vec{e}_{\bar{b}} \quad (1.59c)$$

$$\equiv T_{\bar{a}\bar{b}} \vec{e}_{\bar{a}} \vec{e}_{\bar{b}}. \quad (1.59d)$$

The final equality introduced the transformed tensor components

$$T_{\bar{a}\bar{b}} = T_{ab} (\mathcal{R}^T)_{a\bar{a}} (\mathcal{R}^T)_{b\bar{b}} = T_{ab} \mathcal{R}_{\bar{a}a} \mathcal{R}_{\bar{b}b}. \quad (1.60)$$

The transformation of the components to higher order tensors follows analogously.

1.7.7 Importance of distinguishing between tensors and matrices

A matrix is an ordered array of objects. Hence, matrices are useful for organizing the coordinate components to a tensor. For example, the coordinate components to a first-order tensor (a vector) can be organized into a row or column matrix. Likewise, the coordinate components to a second-order tensor can be organized into a matrix. Consequently, the algebra of Cartesian tensors shares much with that of matrices familiar from linear algebra. However, tensors are not equivalent to matrices. The key distinction concerns how the components to tensors transform under changes to coordinates.

Namely, tensor components transform in a precise manner when modifying coordinates. In contrast, elements of an arbitrary matrix may or may not transform, with details depending on what the matrix elements represent. One means to help maintain focus on the distinction is to recall that a tensor is a geometric object (Section 1.1) that can be represented using arbitrary coordinates. Since the tensor has an existence independent of coordinates, its coordinate components are constrained to transform in a precise manner under changes to the coordinates. In particular, the Cartesian tensors of this chapter transform only via rotations.

1.8 Homogeneity and isotropy

We have many occasions to consider basic symmetry properties of tensor fields, with homogeneity and isotropy two rather basic symmetry properties.

1.8.1 Homogeneous tensor

A tensor field is *homogeneous* if it possesses the same value at each point in space. For example, a uniform temperature field is homogeneous, as is a uniform velocity field. As defined, a homogeneous tensor field has no spatial dependence and thus it does not provide any means to distinguish points in space. Likewise, a time independent tensor is said to be homogeneous in time.

1.8.2 Isotropic and anisotropic tensors

A tensor field is *isotropic* if its representation remains independent of coordinate basis. A scalar tensor is, by definition, isotropic since it has no information about spatial directions. A nonzero vector field cannot be isotropic since it points in a particular direction and so its representation is dependent on the orientation of the basis vectors.

A second order isotropic tensor, I , satisfies

$$I_{\bar{a}\bar{b}} = I_{ab} \mathcal{R}_{\bar{a}a} \mathcal{R}_{\bar{b}b} \equiv I_{ab} \quad (1.61)$$

where we used the identity (1.60) for the component transformation. For nonzero tensors, this equation is satisfied by

$$I_{ab} = \lambda \delta_{ab}, \quad (1.62)$$

with λ an arbitrary scalar and δ_{ab} the Kronecker tensor (1.8), which is the identity tensor. Hence, the most general second order isotropic tensor is proportional to the identity tensor. We often find it useful to decompose an arbitrary second order tensor into its anisotropic and isotropic components according to

$$T_{ab} = \underbrace{T_{ab} - \frac{T_{pp}}{\delta_{pp}} \delta_{ab}}_{\text{anisotropic}} + \underbrace{\frac{T_{pp}}{\delta_{pp}} \delta_{ab}}_{\text{isotropic}}, \quad (1.63)$$

where

$$T_{pp} = \sum_{p=1}^N T_{pp} = \text{trace of } T_{ab} \quad \text{and} \quad \delta_{pp} = \sum_{p=1}^N \delta_{pp} = N \quad (1.64)$$

with N the number of space dimensions.



1.9 Exercises

EXERCISE 1.1: PRODUCT OF SYMMETRIC MATRICES

Let $A = A^T$ and $B = B^T$ be two symmetric matrices. Under what condition is their product also symmetric: $AB = (AB)^T$?

EXERCISE 1.2: PRODUCT OF SYMMETRIC AND ANTI-SYMMETRIC TENSORS

Let $A = -A^T$ be an anti-symmetric matrix, and $S = S^T$ be a symmetric matrix. Show that the trace of their product vanishes: $\text{Tr}(AS) = 0$. Alternatively, in terms of tensors, show that the full contraction of an anti-symmetric tensor with a symmetric tensor vanishes: $A_{mn} S_{mn} = 0$.

EXERCISE 1.3: PROJECTION OPERATOR

Consider an arbitrary direction in space specified by the unit vector \hat{n} with components \hat{n}_a . Define the projection operator

$$P_{ab} = \delta_{ab} - \hat{n}_a \hat{n}_b, \quad (1.65)$$

and show that

$$P_{ab} T_{bc} \hat{n}_a = 0. \quad (1.66)$$

Hence, $S_{ac} = P_{ab} T_{bc}$ is the projection of T onto the plane perpendicular to the direction \hat{n} .



Cartesian tensor calculus

This chapter presents differential and integral calculus for our study of fluid mechanics. We build from the Cartesian tensor algebra of Chapter 1 to develop elements of Cartesian tensor calculus, which is the basis for vector calculus. The material is used throughout this book so the reader is encouraged to master the material.

READER'S GUIDE TO THIS CHAPTER

For further study consult any book on calculus with analytic geometry. Particularly effective treatments, with applications to physics, are given in the following resources.

FEYNMAN LECTURES: Chapters 2 and 3 in Volume II of the [Feynman Lectures](#) offers insightful discussions of vector differential calculus. Although written for students of electrodynamics, many of the examples are drawn from fluid mechanics.

DIV, GRAD, CURL AND ALL THAT ([Schey, 2004](#)): This text pedagogically presents the methods and theorems of vector calculus in a manner that greatly assists the development of intuition.

THEORY AND PROBLEMS OF VECTOR ANALYSIS ([Spiegel, 1974](#)): This “Schaum’s Outline Series” book has nearly 500 worked exercises. Hence, it is a particularly useful resource for those wishing a refresher on problem solving in vector calculus. Some of the exercises here in Section 2.10 are drawn from this book.

[This video from 3Blue1Brown](#) provides some compelling graphics to help develop intuition for the divergence and curl operators as encountered in fluid mechanics.

2.1	The gradient of a scalar field	22
2.1.1	Direction of steepest ascent	23
2.1.2	Tangent to an isosurface	23
2.1.3	Normal to an isosurface	24
2.1.4	Unit vectors change only by rotation	24
2.1.5	Showing that $\delta\hat{\mathbf{n}} \cdot \hat{\mathbf{n}} = 0$	25
2.2	The divergence of a vector field	25
2.2.1	Divergence of a scalar field times a vector field	26
2.2.2	Laplacian of a scalar field	26
2.3	The curl of a vector field	26
2.3.1	Computing the curl	26
2.3.2	Curl-free vector fields	27
2.3.3	Curl-free and divergence-free fields	28
2.3.4	Identities involving the curl	28

2.4	Path integral of a scalar function	30
2.5	Path integral of a vector function	32
2.5.1	Circulation	32
2.5.2	Circulation example	32
2.5.3	Fundamental theorem of calculus	32
2.6	Stokes' theorem	33
2.6.1	Statement of Stokes' theorem	33
2.6.2	Stokes' theorem for a rectangular region	33
2.7	Gauss's divergence theorem	34
2.7.1	An example rectangular volume	34
2.7.2	Divergence theorem for scalar fields	35
2.7.3	Surface integral of the outward normal	35
2.7.4	First and second form of Green's identities	36
2.7.5	Integral of a curl over a closed surface	36
2.7.6	The domain integral of a non-divergent field	37
2.8	Exact and inexact differentials	37
2.8.1	Exact differentials	37
2.8.2	Inexact differentials	38
2.8.3	Integrating factors	38
2.8.4	An example using the velocity field	39
2.8.5	Heuristic physics of exact and inexact differential operations	39
2.9	Helmholtz decomposition	39
2.10	Exercises	40

2.1 The gradient of a scalar field

Consider a scalar field $\psi(\mathbf{x})$ defined on Euclidean space with position

$$\mathbf{x} = \sum_{a=1}^3 x_a \vec{e}_a = x_a \vec{e}_a, \quad (2.1)$$

where the second equation made use of the Einstein summation convention used throughout this book. For example, the scalar field, ψ , can be the temperature, the mass density, or the specific entropy. We may estimate its value at an adjacent spacial point $\mathbf{x} + d\mathbf{x}$ through a Taylor series

$$\psi(\mathbf{x} + d\mathbf{x}) = \psi(\mathbf{x}) + \frac{\partial \psi}{\partial x_1} dx_1 + \frac{\partial \psi}{\partial x_2} dx_2 + \frac{\partial \psi}{\partial x_3} dx_3 + \mathcal{O}(d\mathbf{x} \cdot d\mathbf{x}) \quad (2.2a)$$

$$\approx [1 + dx_a \partial_a] \psi(\mathbf{x}), \quad (2.2b)$$

where we dropped higher order terms to reach the final approximate expression, and introduced the shorthand notation for the partial derivative operator

$$\partial_a = \frac{\partial}{\partial x_a}. \quad (2.3)$$

We can introduce the gradient operator according to

$$\nabla = \vec{e}_a \partial_a = \hat{\mathbf{x}} \partial_x + \hat{\mathbf{y}} \partial_y + \hat{\mathbf{z}} \partial_z \quad (2.4)$$

in which case

$$\psi(\mathbf{x} + d\mathbf{x}) \approx (1 + d\mathbf{x} \cdot \nabla) \psi(\mathbf{x}). \quad (2.5)$$

Note that in some treatments, ∇ is referred to as *Hamilton's operator*.

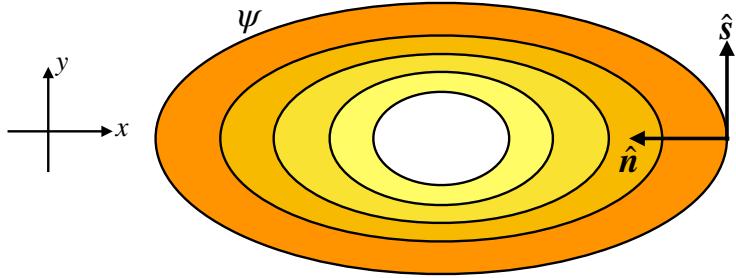


FIGURE 2.1: Contours of a scalar field $\psi(x, y)$, with values increasing toward the center. At any point in space, $\nabla\psi$ points in the direction of steepest increase (ascent) and orients the normal vector $\hat{n} = |\nabla\psi|^{-1} \nabla\psi$. The unit tangent vector, \hat{s} , points in a direction tangent to a ψ isosurface so that it follows the surface of constant ψ and it is orthogonal to the direction of steepest ascent: $\hat{n} \cdot \hat{s} = 0$.

2.1.1 Direction of steepest ascent

Using the approximate relation (2.5), and the geometric expression (1.11) for the scalar product, renders

$$\psi(\mathbf{x} + d\mathbf{x}) - \psi(\mathbf{x}) \approx |d\mathbf{x}| |\nabla\psi| \cos \vartheta, \quad (2.6)$$

where ϑ is the angle between the differential increment $d\mathbf{x}$ and the gradient $\nabla\psi$. Orienting the increment $d\mathbf{x}$ so that $\vartheta = 0$ ensures that $\psi(\mathbf{x} + d\mathbf{x}) - \psi(\mathbf{x})$ is maximal. Consequently, $\nabla\psi$ points in the direction of *steepest ascent* across constant ψ isosurfaces (Figure 2.1). The opposite direction is that of *steepest descent*, where $\vartheta = \pi$.

2.1.2 Tangent to an isosurface

Consider a family of isosurfaces defined by points satisfying

$$\psi(\mathbf{x}) = \text{constant} \quad (2.7)$$

Figure 2.1 shows a two dimensional example where the isosurfaces are lines where ψ is a constant. As another example, consider $\psi(\mathbf{x}) = \psi(r)$, where $r^2 = \mathbf{x} \cdot \mathbf{x}$ is the squared radius of a sphere. Isosurfaces for this spherically symmetric function are spherical shells of radius r .

In general, moving along an isosurface keeps the scalar field unchanged. Let \hat{s} be a unit vector that points in the direction tangent to the isosurface at any point \mathbf{x} . By construction

$$\psi(\mathbf{x} + \hat{s} ds) - \psi(\mathbf{x}) = 0, \quad (2.8)$$

where ds is an infinitesimal increment. In words, this identity says that if we move an infinitesimal distance in the direction tangent to the isosurface, then the function ψ does not change its value. Now expanding this identity in a Taylor series leads to the vanishing of the tangential partial derivative

$$\hat{s} \cdot \nabla\psi = \frac{\partial\psi}{\partial s} = 0. \quad (2.9)$$

That is, isosurfaces of a function ψ are defined by directions along which the partial derivative of the function vanishes. For the spherically symmetric function, $\psi(\mathbf{x}) = \psi(r)$, the tangent vector points in either of the two angular directions along the spherical surface.

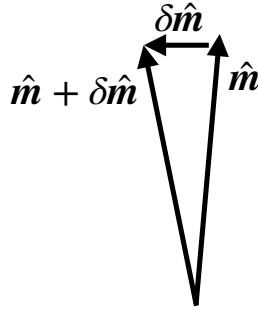


FIGURE 2.2: The infinitesimal change to a unit vector is itself perpendicular to the unit vector: $\delta\hat{\mathbf{m}} \cdot \hat{\mathbf{m}} = 0$. The reason is that the unit vector is constrained to remain unit length, so that the only way that it can change is to change its direction.

2.1.3 Normal to an isosurface

We may normalize the direction of maximal ascent, in which case we define the normal direction

$$\hat{\mathbf{n}} = |\nabla\psi|^{-1} \nabla\psi. \quad (2.10)$$

By construction, the gradient computed in the $\hat{\mathbf{n}}$ direction yields the maximum change for the function so that the *normal derivative* is given by

$$\hat{\mathbf{n}} \cdot \nabla\psi = \frac{\partial\psi}{\partial n} = |\nabla\psi|. \quad (2.11)$$

For the spherically symmetric example,

$$\hat{\mathbf{n}} = \frac{\mathbf{x}}{|\mathbf{x}|} = \hat{\mathbf{r}}, \quad (2.12)$$

where $\hat{\mathbf{r}}$ is the unit vector pointing radially outward from the origin. In this case the normal derivative is equal to the radial derivative

$$\hat{\mathbf{n}} \cdot \nabla\psi = \frac{\partial\psi}{\partial r} \quad \text{spherically symmetric } \psi. \quad (2.13)$$

2.1.4 Unit vectors change only by rotation

Consider an arbitrary unit vector, $\hat{\mathbf{m}}$. The defining feature of a unit vector is that it has unit magnitude

$$\hat{\mathbf{m}} \cdot \hat{\mathbf{m}} = 1. \quad (2.14)$$

Unit vectors can only be modified through changes in their direction since their magnitude is fixed at unity. Hence, they are only modified by rotations. An important consequence of this constraint is that the infinitesimal change to a unit vector is perpendicular to unit vector itself (see Figure 2.2). We see this property through considering an arbitrary infinitesimal change, symbolized by δ , in which

$$0 = \delta(1) = \delta(\hat{\mathbf{m}} \cdot \hat{\mathbf{m}}) = 2\hat{\mathbf{m}} \cdot \delta\hat{\mathbf{m}}. \quad (2.15)$$

In Section 10.4, we formally show that the constraint

$$\delta\hat{\mathbf{m}} \cdot \hat{\mathbf{m}} = 0 \quad (2.16)$$

means that unit vector changes can only arise from rotations. Even so, the above assertion should make intuitive sense.

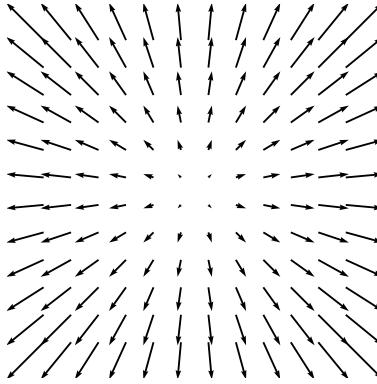


FIGURE 2.3: A vector field with a non-zero horizontal divergence. With $\mathbf{F} = x\hat{\mathbf{x}} + y\hat{\mathbf{y}}$ the field diverges at each point with a spatially constant divergence $\nabla \cdot \mathbf{F} = 2$. Note that this vector field has zero curl, $\nabla \wedge \mathbf{F} = 0$.

2.1.5 Showing that $\delta\hat{\mathbf{n}} \cdot \hat{\mathbf{n}} = 0$

As an illustration of the constraint (2.16), let us verify that it holds for the special case of a unit normal vector (2.10) defined according to surfaces of constant scalar field

$$\hat{\mathbf{n}} = |\nabla\psi|^{-1} \nabla\psi. \quad (2.17)$$

The proof follows first by writing

$$\delta\hat{\mathbf{n}} = |\nabla\psi|^{-1} [\delta(\nabla\psi) - \hat{\mathbf{n}} \delta|\nabla\psi|], \quad (2.18)$$

so that

$$|\nabla\psi| \hat{\mathbf{n}} \cdot \delta\hat{\mathbf{n}} = \hat{\mathbf{n}} \cdot \delta(\nabla\psi) - \delta|\nabla\psi| = \frac{\nabla\psi \cdot \delta(\nabla\psi)}{|\nabla\psi|} - \delta|\nabla\psi| = 0. \quad (2.19)$$

The last equality made use of the identity

$$\delta(|\nabla\psi|) = \delta(\sqrt{\nabla\psi \cdot \nabla\psi}) = \frac{1}{2\sqrt{\nabla\psi \cdot \nabla\psi}} \delta(\nabla\psi \cdot \nabla\psi) = \frac{\nabla\psi \cdot \delta(\nabla\psi)}{|\nabla\psi|}. \quad (2.20)$$

2.2 The divergence of a vector field

The divergence of a vector field, \mathbf{F} , is the scalar product of the divergence operator with the vector

$$\operatorname{div}(\mathbf{F}) = \nabla \cdot \mathbf{F} = \partial_a F_a \begin{cases} > 0 \Rightarrow \text{diverging vector field,} \\ < 0 \Rightarrow \text{converging vector field} \\ = 0 \Rightarrow \text{divergence-free vector field.} \end{cases} \quad (2.21)$$

If the vector field in the surrounding neighborhood of a point is directed away from that point, then the vector field is diverging as if there is a source at the point (Figure 2.3). In this case the divergence of the vector field is positive. The converse occurs for a vector field converging to a point as if there is a sink. If the vector field under consideration is the velocity field of a fluid, then these considerations are directly related to the conservation of matter (see Chapter 15).

The above example of a point source or sink provide a good introduction to divergence. More generally, a vector field possess a positive divergence at any point where there is more “stuff” leaving than entering. For example, if the vector field is the velocity for a fluid, then there is a net divergence if more fluid leaves a point than enters, and the converse holds if the fluid is converging. We further discuss these ideas when studying Gauss’s divergence theorem in Section 2.7.

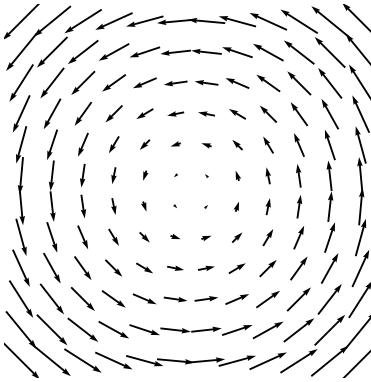


FIGURE 2.4: A horizontal vector field with a constant curl and zero divergence: $\mathbf{F} = -y \hat{x} + x \hat{y}$, $\Rightarrow \nabla \wedge \mathbf{F} = 2 \hat{z}$ and $\nabla \cdot \mathbf{F} = 0$.

2.2.1 Divergence of a scalar field times a vector field

We have many opportunities to make use of properties of the divergence operator following from application of the chain rule. For example, use of the chain rule indicates that the divergence of a scalar field times a vector field is given by

$$\nabla \cdot (\phi \mathbf{F}) = \partial_a(\phi F_a) \quad (2.22a)$$

$$= \partial_a(\phi) F_a + \phi \partial_a F_a \quad (2.22b)$$

$$= \mathbf{F} \cdot \nabla \phi + \phi \nabla \cdot \mathbf{F}. \quad (2.22c)$$

2.2.2 Laplacian of a scalar field

The Laplacian of a scalar field is the divergence of the gradient

$$\nabla^2 \psi = \nabla \cdot \nabla \psi. \quad (2.23)$$

Scalar fields that have a vanishing Laplacian are said to be *harmonic*

$$\nabla^2 \psi = 0 \quad \text{harmonic function.} \quad (2.24)$$

Familiar examples of harmonic functions are the sines and cosines used for Fourier analysis in Euclidean space, and the spherical harmonics used for Fourier analysis on the sphere. The name *harmonic* originates from the relation of harmonic functions to characteristic vibrational modes of a taut string such as those found on musical instruments (when played with skill). Furthermore, harmonic functions play a central role in complex analysis.

2.3 The curl of a vector field

The curl characterizes how a vector field spins around each point in space. In fluid mechanics we make great use of the vorticity field, which is the curl of the velocity (Chapter 36).

2.3.1 Computing the curl

We measure the curl of a vector by computing the cross product of the divergence operator and the vector field. Hence, just like the cross product from Section 1.4, the curl is specified by both a

magnitude and a direction

$$\text{curl}(\mathbf{F}) = \nabla \wedge \mathbf{F} \quad (2.25a)$$

$$= \vec{e}_a \partial_a \wedge \vec{e}_b F_b \quad \text{coordinate representation and basis} \quad (2.25b)$$

$$= \vec{e}_a \wedge \partial_a (\vec{e}_b F_b) \quad \text{move derivative operator onto } \mathbf{F} = \vec{e}_b F_b \quad (2.25c)$$

$$= (\vec{e}_a \wedge \vec{e}_b) \partial_a F_b + F_b (\vec{e}_a \wedge \partial_a \vec{e}_b) \quad \text{perform product rule} \quad (2.25d)$$

$$= \epsilon_{abc} \vec{e}_c \partial_a F_b \quad \partial_a \vec{e}_b = 0 \text{ for Cartesian coordinates.} \quad (2.25e)$$

To reach this result we set $\partial_a \vec{e}_b = 0$ since the Cartesian basis vectors are fixed in space.¹ We also made use of the relation (1.12) for the cross product of basis vectors. Expanding the final expression leads to the familiar expression

$$\text{curl}(\mathbf{F}) = \nabla \wedge \mathbf{F} = \left(\frac{\partial F_3}{\partial x_2} - \frac{\partial F_2}{\partial x_3} \right) \hat{x} + \left(\frac{\partial F_1}{\partial x_3} - \frac{\partial F_3}{\partial x_1} \right) \hat{y} + \left(\frac{\partial F_2}{\partial x_1} - \frac{\partial F_1}{\partial x_2} \right) \hat{z}, \quad (2.26)$$

which can also be written as a determinant

$$\nabla \wedge \mathbf{F} = \det \begin{bmatrix} \vec{e}_1 & \vec{e}_2 & \vec{e}_3 \\ \partial_1 & \partial_2 & \partial_3 \\ F_1 & F_2 & F_3 \end{bmatrix}. \quad (2.27)$$

The horizontal vector field $\mathbf{F} = x \hat{x} + y \hat{y}$ shown in Figure 2.3 has zero curl yet non-zero divergence. Figure 2.4 shows another vector field, $\mathbf{F} = -y \hat{x} + x \hat{y}$, with zero divergence yet nonzero curl $\nabla \wedge \mathbf{F} = 2 \hat{z}$. As seen in Section 33.5, this vector field corresponds to the velocity due to solid-body motion on a rotating planet, with its curl determining the planetary vorticity.

2.3.2 Curl-free vector fields

There are some cases of physically relevant vector fields that have a vanishing curl

$$\nabla \wedge \mathbf{F} = 0. \quad (2.28)$$

We sometimes refer to such curl-free vectors as *irrotational*. In fluid mechanics a curl-free velocity field has zero vorticity, which is a property maintained by linear gravity waves in the absence of rotation (Section 32.8.3). We illustrate a curl-free vector field in Figure 2.5.

The curl of a gradient vanishes

$$\nabla \wedge \nabla \psi = \vec{e}_a \partial_a \wedge \vec{e}_b \partial_b \psi \quad (2.29a)$$

$$= (\vec{e}_a \wedge \vec{e}_b) \partial_a \partial_b \psi \quad (2.29b)$$

$$= 0, \quad (2.29c)$$

where the final equality follows since $\vec{e}_a \wedge \vec{e}_b$ is anti-symmetric on the labels ab whereas $\partial_a \partial_b$ is symmetric. This property allows us to introduce a scalar field ψ for curl-free vector fields so that

$$\mathbf{F} = -\nabla \psi \quad \text{scalar potential.} \quad (2.30)$$

The scalar ψ is known as the *scalar potential*. In the specific case of \mathbf{F} representing the gravitational force, then ψ is called the gravitational potential (see Section 10.12 and Chapter 29).

¹Basis vectors corresponding to non-Cartesian coordinates are spatially dependent (see Chapters 6 and 7), thus making this step invalid for general tensors. We will find a “fix” for this step in Section 7.17 by defining the *covariant curl operator*.

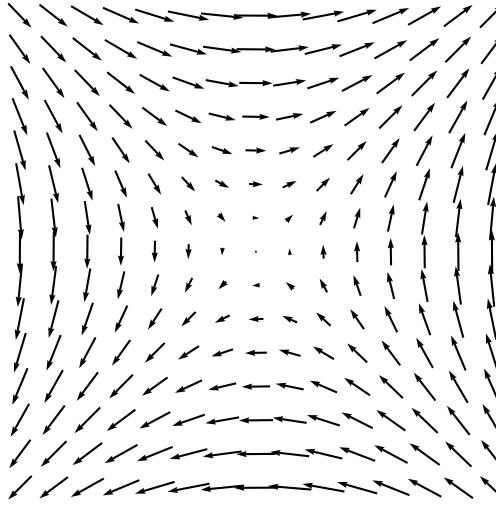


FIGURE 2.5: A horizontal vector field with a zero curl, where $\mathbf{F} = -\nabla\psi$ with the scalar potential given by $\psi = \sin(x/10) \sin(y/10)$.

2.3.3 Curl-free and divergence-free fields

Consider a vector field that has zero curl *and* zero divergence. The curl-free property means that

$$\nabla \wedge \mathbf{F} = 0 \Rightarrow \mathbf{F} = -\nabla\psi. \quad (2.31)$$

The divergence-free property means that ψ is a harmonic function (Section 2.2.2)

$$\nabla \cdot \nabla\psi = \nabla^2\psi = 0. \quad (2.32)$$

The velocity field arising from a linear non-rotating gravity wave (Section 32.8.3) in a Boussinesq fluid (Section 24.1) maintains zero vorticity and zero divergence. Furthermore, curl-free and divergence-free velocity fields are commonly encountered in engineering applications such as aerodynamics (e.g., see [Acheson \(1990\)](#) for many elementary examples).

2.3.4 Identities involving the curl

We close this section by deriving a suite of identities involving the curl operator. These identities are especially useful when developing dynamical equations for vorticity. Furthermore, by making use of the rules for general tensor analysis developed in Chapters 6 and 7, these formula take on the same form regardless the coordinate choice.

Divergence of the curl vanishes

The divergence of the curl vanishes, as seen through the following

$$\nabla \cdot (\nabla \wedge \mathbf{F}) = \partial_a(\epsilon_{abc} \partial_b F_c) \quad (2.33a)$$

$$= \epsilon_{abc} \partial_a \partial_b F_c \quad (2.33b)$$

$$= 0. \quad (2.33c)$$

The final equality holds since $\partial_a \partial_b$ is symmetric on ab whereas ϵ_{abc} is anti-symmetric.

Divergence of a cross product

We now derive an expression for the divergence of a cross product

$$\nabla \cdot (\mathbf{F} \wedge \mathbf{E}) = \mathbf{E} \cdot (\nabla \wedge \mathbf{F}) - \mathbf{F} \cdot (\nabla \wedge \mathbf{E}) \quad (2.34)$$

through the following manipulations

$$\nabla \cdot (\mathbf{F} \wedge \mathbf{E}) = \vec{e}_a \cdot \partial_a (F_b \vec{e}_b \wedge E_c \vec{e}_c) \quad (2.35a)$$

$$= \vec{e}_a \cdot (\vec{e}_b \wedge \vec{e}_c) \partial_a (F_b E_c) \quad (2.35b)$$

$$= \epsilon_{abc} \partial_a (F_b E_c) \quad (2.35c)$$

$$= F_b \epsilon_{abc} \partial_a E_c + E_c \epsilon_{abc} \partial_a F_b \quad (2.35d)$$

$$= -\mathbf{F} \cdot (\nabla \wedge \mathbf{E}) + \mathbf{E} \cdot (\nabla \wedge \mathbf{F}). \quad (2.35e)$$

Curl of a scalar times a vector

We can compute the curl of a scalar field $\psi \mathbf{F}$ through the following steps

$$\nabla \wedge (\psi \mathbf{F}) = \vec{e}_a \partial_a \wedge \psi \vec{e}_b F_b \quad (2.36a)$$

$$= (\vec{e}_a \wedge \vec{e}_b) \partial_a (\psi F_b) \quad (2.36b)$$

$$= \epsilon_{abc} \vec{e}_c (\psi \partial_a F_b + F_b \partial_a \psi) \quad (2.36c)$$

$$= \psi \nabla \wedge \mathbf{F} + \nabla \psi \wedge \mathbf{F}. \quad (2.36d)$$

Curl of a cross product

The curl of a cross product of two vectors is given by

$$\nabla \wedge (\mathbf{F} \wedge \mathbf{E}) = \vec{e}_a \partial_a \wedge (\vec{e}_b F_b \wedge \vec{e}_c E_c) \quad (2.37a)$$

$$= \vec{e}_a \wedge (\vec{e}_b \wedge \vec{e}_c) \partial_a (F_b E_c) \quad (2.37b)$$

$$= \vec{e}_a \wedge (\epsilon_{bcd} \vec{e}_d) \partial_a (F_b E_c) \quad (2.37c)$$

$$= \epsilon_{ade} \epsilon_{bcd} \vec{e}_e \partial_a (F_b E_c) \quad (2.37d)$$

$$= -\epsilon_{aed} \epsilon_{bcd} \vec{e}_e \partial_a (F_b E_c) \quad (2.37e)$$

$$= -(\delta_{ab} \delta_{ec} - \delta_{ac} \delta_{eb}) \vec{e}_e \partial_a (F_b E_c) \quad (2.37f)$$

$$= (-\delta_{ba} \vec{e}_c + \delta_{ca} \vec{e}_b) \partial_a (F_b E_c) \quad (2.37g)$$

$$= \mathbf{F} (\nabla \cdot \mathbf{E}) + (\mathbf{E} \cdot \nabla) \mathbf{F} - \mathbf{E} (\nabla \cdot \mathbf{F}) - (\mathbf{F} \cdot \nabla) \mathbf{E}. \quad (2.37h)$$

Curl of a curl

A special case of the identity (2.37h) allows us to write the curl of a curl as

$$\nabla \wedge (\nabla \wedge \mathbf{F}) = \vec{e}_a \wedge (\vec{e}_b \wedge \vec{e}_c) \partial_a \partial_b F_c \quad (2.38a)$$

$$= (-\delta_{ba} \vec{e}_c + \delta_{ca} \vec{e}_b) \partial_a \partial_b F_c \quad (2.38b)$$

$$= \nabla (\nabla \cdot \mathbf{F}) - \nabla^2 \mathbf{F}. \quad (2.38c)$$

Relating advection, curl, and kinetic energy

We now apply some of the previous manipulations to derive a relation required to derive the vorticity equation (Section 36.4.1). Here, we aim to show that

$$(\mathbf{v} \cdot \nabla) \mathbf{v} = \boldsymbol{\omega} \wedge \mathbf{v} + \nabla(\mathbf{v} \cdot \mathbf{v})/2, \quad (2.39)$$

where

$$\boldsymbol{\omega} = \nabla \wedge \mathbf{v} \quad (2.40)$$

is the vorticity, $\mathbf{v} \cdot \mathbf{v}/2$ is the kinetic energy per mass, and \mathbf{v} is the fluid velocity field. We here show all the steps along with their justification

$$\begin{aligned}
 \boldsymbol{\omega} \wedge \mathbf{v} &= (\nabla \wedge \mathbf{v}) \wedge \mathbf{v} && \text{insert } \boldsymbol{\omega} = \nabla \wedge \mathbf{v} && (2.41a) \\
 &= (\vec{e}_a \partial_a \wedge \vec{e}_b v_b) \wedge \vec{e}_c v_c && \text{Cartesian representation of } \mathbf{v} \text{ and } \nabla && (2.41b) \\
 &= (\vec{e}_a \wedge \vec{e}_b) \wedge \vec{e}_c (\partial_a v_b) v_c && \text{rearrange} && (2.41c) \\
 &= \epsilon_{abd} (\vec{e}_d \wedge \vec{e}_c) (\partial_a v_b) v_c && \text{first cross product expanded} && (2.41d) \\
 &= \epsilon_{abd} \epsilon_{dce} \vec{e}_e (\partial_a v_b) v_c && \text{second cross product expanded} && (2.41e) \\
 &= \epsilon_{abd} \epsilon_{ced} \vec{e}_e (\partial_a v_b) v_c && \text{arrange indices to prepare for next step} && (2.41f) \\
 &= (\delta_{ac} \delta_{be} - \delta_{ae} \delta_{bc}) \vec{e}_e (\partial_a v_b) v_c && \text{use identity (1.38)} && (2.41g) \\
 &= \vec{e}_a v_c \partial_c v_a - \vec{e}_a v_c \partial_a v_c && \text{contract the Kronecker deltas} && (2.41h) \\
 &= \vec{e}_a [(\mathbf{v} \cdot \nabla) v_a - \partial_a \mathbf{v}^2/2] && \text{re-express as Cartesian tensor} && (2.41i) \\
 &= (\mathbf{v} \cdot \nabla) \mathbf{v} - \nabla[\mathbf{v} \cdot \mathbf{v}/2] && \text{rearrange.} && (2.41j)
 \end{aligned}$$

Note that Section 4.4.4 of [Griffies \(2004\)](#) exhibits these steps making use of general coordinates rather than Cartesian.

2.4 Path integral of a scalar function

Consider the integral of a scalar function, ψ , over an arbitrary one-dimensional path in space, \mathcal{C}

$$\mathcal{I} = \int_{\mathcal{C}} \psi(\varphi) d\varphi. \quad (2.42)$$

A path is a one-dimensional curve, so that a point along the path can be specified by a single parameter, denoted here by φ (see Sections 5.1 and 5.2 for more on paths). Now lay down a Cartesian coordinate system with an arbitrary origin. The corresponding Cartesian coordinate representation of a point along the path is written

$$\mathcal{C} \mapsto \mathbf{x}(\varphi) = \hat{\mathbf{x}} x(\varphi) + \hat{\mathbf{y}} y(\varphi) + \hat{\mathbf{z}} z(\varphi) \quad (2.43)$$

so that the path integral can be written

$$\mathcal{I} = \int_{\mathcal{C}} \psi(\varphi) d\varphi = \int_{\mathcal{C}} \psi[\mathbf{x}(\varphi)] d\varphi. \quad (2.44)$$

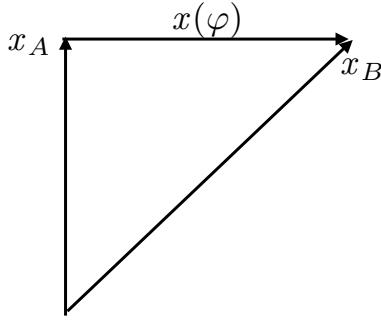


FIGURE 2.6: A linear path, $\mathbf{x}(\varphi)$ extending from \mathbf{x}_A to \mathbf{x}_B can be parameterized by a non-dimensional parameter $\varphi \in [0, 1]$ via $\mathbf{x}(\varphi) = \mathbf{x}_A + \varphi(\mathbf{x}_B - \mathbf{x}_A)$. Alternatively it can be parameterized by the arc-length along the path via $\mathbf{x}(s) = \mathbf{x}_A + \hat{\mathbf{t}} s$ with $s \in [0, L]$, $L = |\mathbf{x}_B - \mathbf{x}_A|$, and $\hat{\mathbf{t}}$ the unit tangent vector pointing from \mathbf{x}_A to \mathbf{x}_B .

Arc length to parameterize the path

A particularly common special case for path parameterization is where we choose φ to be the arc length along the path. For Euclidean space using Cartesian coordinates, the differential increment of arc length is given by

$$ds = \sqrt{d\mathbf{x} \cdot d\mathbf{x}}. \quad (2.45)$$

Inserting $\mathbf{x}(s)$ into this expression renders

$$ds = \sqrt{d\mathbf{x} \cdot d\mathbf{x}} = ds \sqrt{\frac{d\mathbf{x}}{ds} \cdot \frac{d\mathbf{x}}{ds}}. \quad (2.46)$$

This expression is self-consistent if

$$\sqrt{\frac{d\mathbf{x}}{ds} \cdot \frac{d\mathbf{x}}{ds}} = 1, \quad (2.47)$$

which is merely a rewrite of the defining expression (2.45). The derivative of the curve according to the arc-length, $d\mathbf{x}/ds$, defines a unit tangent vector to the curve

$$\hat{\mathbf{t}} = \frac{d\mathbf{x}}{ds} \implies \hat{\mathbf{t}} \cdot \hat{\mathbf{t}} = 1. \quad (2.48)$$

We offer a more focused discussion of curves and tangents in Section 5.2.1 (see in particular equation (5.6)).

Linear path example

As a specific example, consider a line between two points, \mathbf{x}_A and \mathbf{x}_B , as in Figure 2.6. We can parameterize the line using a dimensionless parameter φ according to

$$\mathbf{x}(\varphi) = \mathbf{x}_A + (\mathbf{x}_B - \mathbf{x}_A) \varphi \quad \varphi \in [0, 1]. \quad (2.49)$$

Alternatively, we can parameterize using the arc length

$$\mathbf{x}(s) = \mathbf{x}_A + \hat{\mathbf{t}} s \quad s \in [0, L], \quad (2.50)$$

where $L = \int_A^B ds = |\mathbf{x}_B - \mathbf{x}_A|$ is the total arc length of the line, and where $\hat{\mathbf{t}}$ is the unit tangent vector pointing along the path from \mathbf{x}_A to \mathbf{x}_B

$$\hat{\mathbf{t}} = \frac{\mathbf{x}'(s)}{|\mathbf{x}'(s)|} = \frac{\mathbf{x}_B - \mathbf{x}_A}{|\mathbf{x}_B - \mathbf{x}_A|}. \quad (2.51)$$

As defined we have $|\mathbf{x}'(s)| = |\hat{\mathbf{t}}| = 1$, so that the path integral is given by $\mathcal{I} = \int_0^L \psi[\mathbf{x}(s)] ds$.

2.5 Path integral of a vector function

Generalizing to a vector field, $\mathbf{F}(\mathbf{x})$, we could conceivably integrate each component of the vector along the curve independently using the approach for scalar fields in Section 2.4. In practice, however, that quantity rarely appears in physics. Instead, we more commonly wish to integrate that component of $\mathbf{F}(\mathbf{x})$ that projects onto a curve

$$\int_C \mathbf{F} \cdot d\mathbf{x} = \int_C \mathbf{F} \cdot \frac{d\mathbf{x}}{ds} ds \quad (2.52)$$

where $d\mathbf{x}/ds$ is tangent to the curve. The work performed by a force applied along a path provides a common example of this path integral (Section 10.1.5).

2.5.1 Circulation

For the case of a closed curve or a circuit (see Section 5.1), we refer to the above path integral as the *circulation* and use the convention of putting an arrowed circle on the integral sign

$$\text{circulation of vector field} = \oint_C \mathbf{F} \cdot d\mathbf{x}. \quad (2.53)$$

The arrow indicates that we conventionally traverse the closed path in a counter-clockwise (right hand) manner.

2.5.2 Circulation example

Consider the vector field $\mathbf{F} = -y \hat{\mathbf{x}} + x \hat{\mathbf{y}}$ shown in Figure 2.4. What is the circulation for this field computed around a circle of radius r whose center is the origin? To compute this circulation we make use of plane polar coordinates, in which $x = r \cos \varphi$ and $y = r \sin \varphi$, with $\varphi \in [0, 2\pi]$ the polar angle measured from the positive x -axis. The position of a point on the circle is thus written $\mathbf{x}(\varphi) = r(\hat{\mathbf{x}} \cos \varphi + \hat{\mathbf{y}} \sin \varphi)$, and the tangent to the circle is $d\mathbf{x}(\varphi)/d\varphi = r(-\hat{\mathbf{x}} \sin \varphi + \hat{\mathbf{y}} \cos \varphi)$. The integrand to the circulation (2.52) thus takes the form

$$\mathbf{F} \cdot \frac{d\mathbf{x}(\varphi)}{d\varphi} = r(y \sin \varphi + x \cos \varphi) = r^2. \quad (2.54)$$

Hence, the circulation around the constant radius circle is given by twice the area of the circle

$$\oint_C \mathbf{F} \cdot d\mathbf{x} = 2\pi r^2. \quad (2.55)$$

This result has application for geophysical fluids when computing the vorticity induced by the rotating planet (see Section 36.7.2).

2.5.3 Fundamental theorem of calculus

The special case of $\mathbf{F} = -\nabla\psi$ for a scalar field ψ recovers the fundamental theorem of calculus

$$\psi(\mathbf{x}_A) - \psi(\mathbf{x}_B) = \int_{\mathbf{x}_A}^{\mathbf{x}_B} d\psi = \int_{\mathbf{x}_A}^{\mathbf{x}_B} \nabla\psi \cdot d\mathbf{x}. \quad (2.56)$$

It follows that for any closed curve with $\mathbf{x}_A = \mathbf{x}_B$, the circulation of $\nabla\psi$ vanishes

$$\oint_C d\psi = \oint_C \nabla\psi \cdot d\mathbf{x} = 0. \quad (2.57)$$

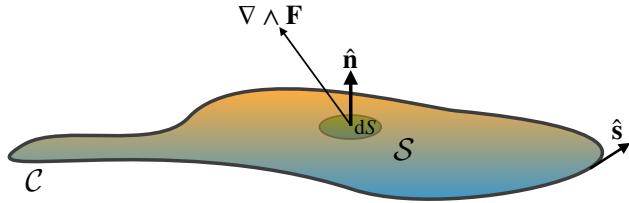


FIGURE 2.7: Illustrating the geometry of Stokes' Theorem. The unit normal $\hat{\mathbf{n}}$ points outward from the surface, \mathcal{S} , with $(\nabla \wedge \mathbf{F}) \cdot \hat{\mathbf{n}}$ the projection of the curl of a vector field onto the surface normal. The outer boundary of the area, $\partial\mathcal{S}$, is traversed counterclockwise following a tangent vector $\hat{\mathbf{s}}$ when computing the circulation.

2.6 Stokes' theorem

Stokes' theorem relates a vector field's behavior on the boundary of a surface to its behavior within the area of the surface. It is used extensively in our study of circulation and vorticity in Part VI of this book.

2.6.1 Statement of Stokes' theorem

Given an oriented surface \mathcal{S} with a boundary $\partial\mathcal{S}$, Stokes' Theorem says that the circulation around the boundary equals to the area integrated curl projected onto the surface outward normal

$$\oint_{\partial\mathcal{S}} \mathbf{F} \cdot d\mathbf{x} = \int_{\mathcal{S}} (\nabla \wedge \mathbf{F}) \cdot \hat{\mathbf{n}} d\mathcal{S}. \quad (2.58)$$

In this equation, $d\mathbf{x}$ is the vector line element along the path, $\partial\mathcal{S}$ is the closed path defining the boundary to an oriented two-dimensional surface \mathcal{S} , $\hat{\mathbf{n}}$ is the outward normal vector on the surface, and $d\mathcal{S}$ is the infinitesimal surface area element. The orientation of the outward normal is determined by the right hand rule according to the circulation direction.

2.6.2 Stokes' theorem for a rectangular region

To build experience with Stokes' theorem, consider the case of a rectangle in the x-y plane with dimensions $L_x \times L_y$. In this case $\hat{\mathbf{n}} = \hat{\mathbf{z}}$, so that

$$(\nabla \wedge \mathbf{F}) \cdot \hat{\mathbf{z}} = \frac{\partial F_2}{\partial x} - \frac{\partial F_1}{\partial y}, \quad (2.59)$$

in which case the right hand side of Stokes' theorem reduces to

$$\int_{\mathcal{S}} (\nabla \wedge \mathbf{F}) \cdot \hat{\mathbf{n}} dS = \int_0^{L_x} \int_0^{L_y} \left(\frac{\partial F_2}{\partial x} - \frac{\partial F_1}{\partial y} \right) dx dy. \quad (2.60)$$

Integration around the rectangle then leads to a direct verification of Stokes' Theorem

$$\int_0^{L_x} \int_0^{L_y} \left(\frac{\partial F_2}{\partial x} - \frac{\partial F_1}{\partial y} \right) dx dy = \int_0^{L_y} F_2(x, y) \Big|_{x=0}^{x=L_x} dy - \int_0^{L_x} F_1(x, y) \Big|_{y=0}^{y=L_y} dx \quad (2.61a)$$

$$= \int_0^{L_x} F_1(x, 0) dx + \int_0^{L_y} F_2(L_x, y) dy + \int_{L_x}^0 F_1(x, L_y) dx + \int_{L_y}^0 F_2(0, y) dy \quad (2.61b)$$

$$= \oint_{\partial\mathcal{S}} \mathbf{F} \cdot d\mathbf{x}. \quad (2.61c)$$

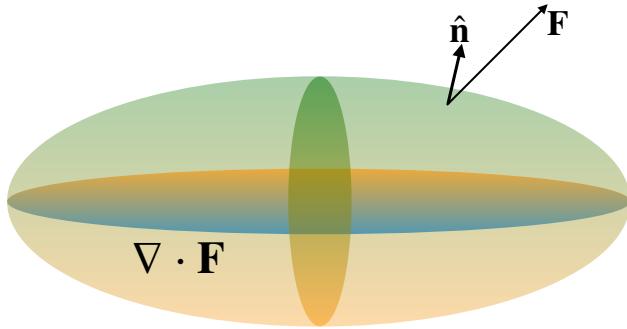


FIGURE 2.8: Illustrating the geometry of Gauss's divergence theorem for an ellipsoidal volume. The outward unit normal $\hat{\mathbf{n}}$ is projected onto the vector field \mathbf{F} via the scalar product, $\mathbf{F} \cdot \hat{\mathbf{n}}$. Integrating this scalar product over the closed surface \mathcal{S} yields the same result as computing the volume integral of the divergence, $\nabla \cdot \mathbf{F}$, over the region, \mathcal{R} , bounded by the closed surface.

We can generalize this result to verify Stokes' Theorem for an arbitrary surface. We do so by breaking the surface into a lattice of tiny rectangles. Integrating around the tiny rectangles and summing their contributions leads to a cancellation of the line integrals over all interior boundaries. The cancellation occurs since an internal edge of a rectangle is integrated once in each direction thus cancelling its contribution. The only nonzero contribution comes from integration over the external boundary.

2.7 Gauss's divergence theorem

Gauss's divergence theorem relates the divergence of a vector field integrated over a volume to the area integrated normal projection of the vector field through the closed surface bounding the volume. For a vector field \mathbf{F} , the divergence theorem states that

$$\int_{\mathcal{R}} \nabla \cdot \mathbf{F} dV = \oint_{\partial\mathcal{R}} \mathbf{F} \cdot \hat{\mathbf{n}} d\mathcal{S}, \quad (2.62)$$

where $\hat{\mathbf{n}}$ is the outward normal to the boundary surface and $d\mathcal{S}$ is the surface area element. We follow the convention that $\oint_{\partial\mathcal{R}}$ refers to a surface integral over a closed surface that bounds a volume. This notation is used to contrast with the surface integral $\int_{\mathcal{S}}$ that generally does not enclose a volume. Figure 2.8 illustrates the geometry of Gauss's divergence theorem. In physics jargon, we say that the divergence of a vector field, $\nabla \cdot \mathbf{F}$, integrated over a volume equals to the flux of that vector field, $\mathbf{F} \cdot \hat{\mathbf{n}}$, integrated over the area bounding the volume.

2.7.1 An example rectangular volume

To build intuition for Gauss's divergence theorem, consider a rectangular volume with dimensions $L_x \times L_y \times L_z$. The volume integral on the left hand side of equation (2.62) gives

$$\int_{\mathcal{R}} \left[\frac{\partial F_1}{\partial x} + \frac{\partial F_2}{\partial y} + \frac{\partial F_3}{\partial z} \right] dx dy dz. \quad (2.63)$$

Focusing on just the leftmost term, integration in x gives

$$\int_{\mathcal{R}} \frac{\partial F_1}{\partial x} dx dy dz = \int_{y=0}^{y=L_y} \int_{z=0}^{z=L_z} [F_1(L_x, y, z) - F_1(0, y, z)] dy dz \quad (2.64a)$$

$$= \int_{S_1+S_2} \mathbf{F} \cdot \hat{\mathbf{n}} d\mathcal{S}, \quad (2.64b)$$

where S_1 is the rectangle's face with outward normal $\hat{\mathbf{n}} = \hat{\mathbf{x}}$ and S_2 is the rectangle's face with normal $\hat{\mathbf{n}} = -\hat{\mathbf{x}}$. Repeating this procedure on the other terms in equation (2.63) gives the area integrated flux through the full boundary. To verify the theorem for a general volume V , we take the approach used to prove Stokes' theorem. First, divide the volume into many rectangular sub-volumes. Then apply the above result to each sub-volume and sum up the result. The fluxes through internal rectangular faces cancel exactly. Therefore, the sum of all the fluxes equals just the flux through the external boundary, yielding the divergence theorem.

2.7.2 Divergence theorem for scalar fields

We consider two corollaries of the divergence theorem, the first of which arises from the special case of a vector field $\mathbf{F} = \phi \mathbf{c}$ with \mathbf{c} an arbitrary *constant* vector. Substitution into the divergence theorem (2.62) yields

$$\oint_{\partial\mathcal{R}} \phi \mathbf{c} \cdot \hat{\mathbf{n}} d\mathcal{S} = \int_{\mathcal{R}} \nabla \cdot (\phi \mathbf{c}) dV. \quad (2.65)$$

Pulling the constant vector out of the integrals and rearrangement leads to

$$\mathbf{c} \cdot \left[\oint_{\partial\mathcal{R}} \phi \hat{\mathbf{n}} d\mathcal{S} - \int_{\mathcal{R}} \nabla \phi dV \right] = 0. \quad (2.66)$$

Since \mathbf{c} is an arbitrary vector, this equality is true in general only when

$$\oint_{\partial\mathcal{R}} \phi \hat{\mathbf{n}} d\mathcal{S} = \int_{\mathcal{R}} \nabla \phi dV. \quad (2.67)$$

In words, this result says that the integral of a scalar field over the boundary of a closed surface, when weighted by the outward normal to the surface, equals to the volume integral of the gradient of the scalar field integrated over the region bounded by the closed surface. We make use of this result in Section 18.1.3 when formulating the contribution of stresses to the motion of a fluid element, in particular when considering how pressure affects motion.

2.7.3 Surface integral of the outward normal

A corollary of equation (2.67) can be found by setting the scalar field, ϕ , to a constant so that $\nabla \phi = 0$. We thus find that the area integral of the outward normal vector vanishes when integrated over the surface of a closed volume

$$\oint_{\partial\mathcal{R}} \hat{\mathbf{n}} d\mathcal{S} = 0. \quad (2.68)$$

An example of this result can be seen by integrating over a closed rectangular volume, whereby the area integrals cancel component-wise. Another example is the sphere, where $\hat{\mathbf{n}} = \hat{\mathbf{r}}$ is the outward radial unit vector. Integration of the unit vector over the spherical surface identically vanishes. For some purposes one takes the result (2.68) as the definition of a simply closed volume.

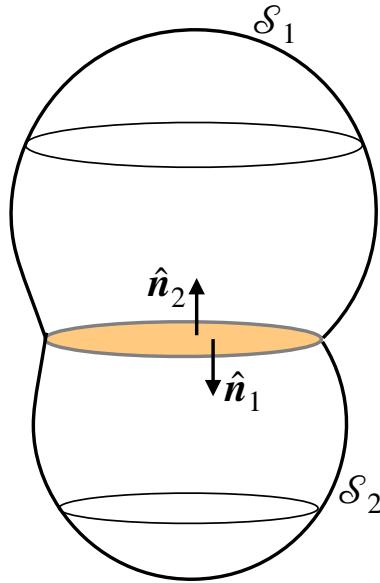


FIGURE 2.9: The integral of the curl of a function vanishes when integrated over an oriented closed surface. This result follows from both Gauss's divergence theorem as well as Stokes' theorem. Here we depict a closed volume that has been split into two surfaces \mathcal{S}_1 and \mathcal{S}_2 . Separately applying Stokes' theorem to the surfaces leads to the calculation of the circulation around the common boundary. Yet the orientation of the outward normal along the closed surface is opposite for the two regions, thus leading to a cancellation of the circulations.

2.7.4 First and second form of Green's identities

The second corollary to the divergence theorem arises from considering another special vector field

$$\mathbf{F} = \psi \nabla \phi, \quad (2.69)$$

with ψ and ϕ scalar fields. Substitution into the divergence theorem (2.62) leads to

$$\oint_{\partial\mathcal{R}} \psi \frac{\partial \phi}{\partial n} d\mathcal{S} = \int_{\mathcal{R}} [\nabla \psi \cdot \nabla \phi + \psi \nabla^2 \phi] dV \quad \text{Green's first integral identity.} \quad (2.70)$$

We can make this result more symmetric by swapping ψ and ϕ and then subtracting, thus yielding

$$\oint_{\partial\mathcal{R}} \left[\psi \frac{\partial \phi}{\partial n} - \phi \frac{\partial \psi}{\partial n} \right] d\mathcal{S} = \int_{\mathcal{R}} [\psi \nabla^2 \phi - \phi \nabla^2 \psi] dV \quad \text{Green's second integral identity.} \quad (2.71)$$

Finally, setting $\phi = 1$ renders

$$\oint_{\partial\mathcal{R}} \frac{\partial \psi}{\partial n} d\mathcal{S} = \int_{\mathcal{R}} \nabla^2 \psi dV \iff \oint_{\partial\mathcal{R}} \nabla \psi \cdot \hat{\mathbf{n}} d\mathcal{S} = \int_{\mathcal{R}} \nabla \cdot \nabla \psi dV. \quad (2.72)$$

These identities are fundamental to the Green's function method for solving linear partial differential equations (see Chapter 4).

2.7.5 Integral of a curl over a closed surface

Application of Gauss's divergence theorem leads us to conclude that the following integral vanishes

$$\oint_{\partial\mathcal{R}} (\nabla \wedge \Psi) \cdot \hat{\mathbf{n}} d\mathcal{S} = \int_{\mathcal{R}} \nabla \cdot (\nabla \wedge \Psi) dV = 0, \quad (2.73)$$

where the final equality follows since the divergence of a curl vanishes. Hence, the integral of the normal projection of the curl of a function over an oriented closed surface vanishes. We can understand this result geometrically by splitting the closed volume into two regions and then applying Stokes' theorem separately to the two regions (see Figure 2.9)

$$\oint_{\partial\mathcal{R}} (\nabla \wedge \Psi) \cdot \hat{\mathbf{n}} d\mathcal{S} = \int_{\mathcal{S}_1} (\nabla \wedge \Psi) \cdot \hat{\mathbf{n}} d\mathcal{S}_1 + \int_{\mathcal{S}_2} (\nabla \wedge \Psi) \cdot \hat{\mathbf{n}} d\mathcal{S}_2 \quad (2.74a)$$

$$= \oint_{\partial\mathcal{S}_1} \Psi \cdot d\mathbf{x} - \oint_{\partial\mathcal{S}_2} \Psi \cdot d\mathbf{x} \quad (2.74b)$$

$$= 0. \quad (2.74c)$$

The minus sign appearing in front of $\oint_{\partial\mathcal{S}_2}$ occurs since the orientation of the circulation integral is opposite that for $\oint_{\partial\mathcal{S}_1}$. We are thus left with a cancellation of the circulations. When applied to the vorticity of a fluid (Chapter 33), we see that

$$\oint_{\partial\mathcal{R}} \boldsymbol{\omega} \cdot \hat{\mathbf{n}} d\mathcal{S} = \int_{\mathcal{R}} \nabla \cdot \boldsymbol{\omega} dV = 0, \quad (2.75)$$

where $\boldsymbol{\omega} = \nabla \wedge \mathbf{v}$ is the vorticity and \mathbf{v} is the fluid velocity.

2.7.6 The domain integral of a non-divergent field

Consider a vector field that has zero divergence everywhere within a domain, \mathcal{R} . Consequently, $\int_{\mathcal{R}} \nabla \cdot \mathbf{F} dV = \oint_{\partial\mathcal{R}} \mathbf{F} \cdot \hat{\mathbf{n}} d\mathcal{S} = 0$. Now what can we say about $\int_{\mathcal{R}} \mathbf{F} dV$? One might be tempted to say that it vanishes. But that is generally an incorrect statement as we now show.

Since $\nabla \cdot \mathbf{F} = 0$ we know that

$$0 = \int_{\mathcal{R}} x_i \nabla \cdot \mathbf{F} dV = \int_{\mathcal{R}} [\nabla \cdot (x_i \mathbf{F}) - F_i] dV, \quad (2.76)$$

where we used $\partial_j x_i = \delta_{ij}$. Use of the divergence theorem leads to

$$\int_{\mathcal{R}} F_i dV = \int_{\mathcal{R}} \nabla \cdot (x_i \mathbf{F}) dV = \oint_{\partial\mathcal{R}} x_i (\hat{\mathbf{n}} \cdot \mathbf{F}) d\mathcal{S}. \quad (2.77)$$

The right hand side does not generally vanish since $\hat{\mathbf{n}} \cdot \mathbf{F}$ does not generally vanish at each point along $\partial\mathcal{R}$. Hence, we find that $\int_{\mathcal{R}} \mathbf{F} dV = 0$ only for those domains where $\hat{\mathbf{n}} \cdot \mathbf{F} = 0$ at each point along $\partial\mathcal{R}$.

2.8 Exact and inexact differentials

Thus far in this chapter all differentials have been exact. However, the thermodynamics discussed in Chapter 21 makes use of both exact and inexact differentials. We here introduce the mathematics of such differentials.

2.8.1 Exact differentials

As in Section 2.1, consider an arbitrary function of space, $\Phi(\mathbf{x})$. A differential increment for that function, computed between two close points \mathbf{x} and $\mathbf{x} + d\mathbf{x}$, is given by

$$d\Phi(\mathbf{x}) = \Phi(\mathbf{x} + d\mathbf{x}) - \Phi(\mathbf{x}) \quad (2.78a)$$

$$= d\mathbf{x} \cdot \nabla \Phi, \quad (2.78b)$$

where we dropped higher order terms due to the infinitesimal nature of the increments. It follows that we can determine the finite increment between two points through integration

$$\Phi(\mathbf{x}_B) - \Phi(\mathbf{x}_A) = \int_{\mathbf{x}_A}^{\mathbf{x}_B} d\Phi(\mathbf{x}) = \int_{\mathbf{x}_A}^{\mathbf{x}_B} d\mathbf{x} \cdot \nabla \Phi. \quad (2.79)$$

These results are familiar from elementary calculus, with the increment $d\Phi$ given by equation (2.78b) termed an *exact* differential. Importantly, the finite increment, $\Phi(\mathbf{x}_B) - \Phi(\mathbf{x}_A)$, depends only on the endpoint values of Φ . It does not depend on the path taken to go from \mathbf{x}_A to \mathbf{x}_B . Correspondingly, the integral of an exact differential around a closed loop vanishes

$$\oint d\Phi = 0. \quad (2.80)$$

2.8.2 Inexact differentials

Consider now a general differential expression written as

$$\mathbf{A} \cdot d\mathbf{x} = A dx + B dy + C dz, \quad (2.81)$$

where $\mathbf{A} = A \hat{\mathbf{x}} + B \hat{\mathbf{y}} + C \hat{\mathbf{z}}$ is an arbitrary vector function. If $\nabla \wedge \mathbf{A} = 0$, then \mathbf{A} can be written as the gradient of a scalar

$$\nabla \wedge \mathbf{A} = 0 \implies \mathbf{A} = \nabla \Psi, \quad (2.82)$$

in which case we have an exact differential expression

$$\mathbf{A} \cdot d\mathbf{x} = \nabla \Psi \cdot d\mathbf{x} = d\Psi. \quad (2.83)$$

That is, the differential $\mathbf{A} \cdot d\mathbf{x}$ is exact if

$$\nabla \wedge \mathbf{A} = 0 \implies \mathbf{A} \cdot d\mathbf{x} \text{ exact differential.} \quad (2.84)$$

In the more general case where $\nabla \wedge \mathbf{A} \neq 0$, then $\mathbf{A} \cdot d\mathbf{x}$ is termed an *inexact differential*. We make use of the following notation for inexact differentials

$$d\Psi = \mathbf{A} \cdot d\mathbf{x}. \quad (2.85)$$

Notably, the path integral of an inexact differential depends on the path taken between the endpoints. Correspondingly, the integral of an inexact differential around a closed loop does not generally vanish

$$\oint d\Psi \neq 0. \quad (2.86)$$

2.8.3 Integrating factors

Consider again the inexact differential $d\Psi = \mathbf{A} \cdot d\mathbf{x}$. Let us presume there exists a function τ so that the product $\tau^{-1} d\Psi$ is an exact differential. For τ to exist it must be such that

$$\nabla \wedge (\mathbf{A} \tau^{-1}) = 0. \quad (2.87)$$

Consequently, we can write

$$\mathbf{A} = \tau \nabla \Phi, \quad (2.88)$$

so that

$$d\Psi = \mathbf{A} \cdot d\mathbf{x} = \tau \nabla \Phi \cdot d\mathbf{x} = \tau d\Phi. \quad (2.89)$$

The function τ is known as an *integrating factor*. As seen in Section 21.2.3, pressure is the integrating factor for mechanical work, temperature is the integrating factor for heating, and the chemical potential is the integrating factor for chemical work.

2.8.4 An example using the velocity field

Consider the product $\mathbf{v} \cdot d\mathbf{x}$, where \mathbf{v} is the velocity field for a fluid and $d\mathbf{x}$ is a differential increment in space directed along a path. Furthermore, introduce the curl of the velocity, which defines the vorticity (Section 33.1) $\boldsymbol{\omega} = \nabla \wedge \mathbf{v}$. For cases where the vorticity vanishes, $\boldsymbol{\omega} = 0$, then $d\Psi = \mathbf{v} \cdot d\mathbf{x}$ is an exact differential. Consequently, Stokes' theorem means that the circulation of an irrotational velocity field computed around an arbitrary closed loop vanishes (Section 33.3)

$$\mathcal{C} \equiv \oint_{\partial S} \mathbf{v} \cdot d\mathbf{x} = \int_S \boldsymbol{\omega} \cdot \hat{\mathbf{n}} dS = 0. \quad (2.90)$$

Another way to see this result is to note that a vanishing curl means that the velocity field can be expressed as the gradient of a scalar, $\mathbf{v} = \nabla\psi$, so that $d\Psi = \nabla\psi \cdot d\mathbf{x}$, which is manifestly exact.

2.8.5 Heuristic physics of exact and inexact differential operations

Consider a hiker climbing a mountain. The mechanical work, which is force applied over a distance, is a function of the path taken. Some paths are smooth and well marked, whereas others are rough and poorly marked. Likewise, the frictional heating (of the hiker's feet, for example) depend on details of the path (and the shoes!). So although the start and finish points are fixed, the work exerted and heat generated in going between these points is a function of the path.

In contrast, the change in gravitational potential energy between the start and finish points is a function only of the relative elevation; it does not depend on the path between the points. So the gravitational potential energy increment between the two points is an exact differential, with the potential energy for each point a function of the elevation at the point. Analogously, the First Law of thermodynamics says that the sum of path-dependent processes (work and heat) used in going from one thermodynamic state to another equals to the difference in the internal energy between the two states. That is, the sum of the inexact differentials for heat and work equal to the exact differential for internal energy.

2.9 Helmholtz decomposition

In the development of conservation laws for continuum systems such as a fluid, it is common to express those laws in their flux-form whereby a scalar property (e.g., matter content) has a local time change determined by the convergence of a flux, \mathbf{J} . Since it is only the divergent portion of the flux that alters the time tendency, the non-divergent "rotational" portion has no impact on the time change; it is a "do nothing" portion of the flux. Therefore, it can be of use to decompose a general flux vector into its divergent and rotational components to focus on just that portion contributing to time changes.

As discussed in [Morse and Feshbach \(1953\)](#), the Helmholtz decomposition theorem states that any vector field, \mathbf{J} , can be decomposed as

$$\mathbf{J} = -\nabla\Psi + \nabla \wedge \mathbf{A}. \quad (2.91)$$

That is, we can decompose \mathbf{J} into a divergent vector plus a solenoidal/rotational vector field so that

$$\nabla \cdot \mathbf{J} = -\nabla^2\Psi \quad \text{and} \quad \nabla \wedge \mathbf{J} = \nabla \wedge (\nabla \wedge \mathbf{A}). \quad (2.92)$$

If we can determine this decomposition in a meaningful and robust manner, then our quest to study just the divergent component is met.

Indeed, the decomposition is unique in an infinite domain assuming suitable decay properties for \mathbf{F} (faster than r^{-2} for three-dimensions and faster than r^{-1} for two-dimensions). Unfortunately, the decomposition is not unique in finite domains since there are no unique choices for the boundary conditions. Furthermore, each of the potentials Ψ and \mathbf{A} are unique only up to a *gauge transformation* whereby

$$\Psi' = \Phi + C \quad \text{and} \quad \mathbf{A}' = \mathbf{A} + \nabla\lambda \quad (2.93)$$

lead to the same decomposition, with C an arbitrary constant and λ an arbitrary function. There are a variety of quasi-methods of use to partially address the ambiguity in a physically sensible manner, with the studies from [Fox-Kemper et al. \(2003\)](#) and [Maddison et al. \(2015\)](#) offering detailed discussions.

2.10 Exercises

Throughout these exercises we consider a point whose Cartesian position vector, relative to an arbitrary origin, is written

$$\mathbf{x} = x \hat{\mathbf{x}} + y \hat{\mathbf{y}} + z \hat{\mathbf{z}} \quad (2.94)$$

and whose squared distance from the origin is

$$r^2 = \mathbf{x} \cdot \mathbf{x} = x^2 + y^2 + z^2. \quad (2.95)$$

EXERCISE 2.1: PRACTICE WITH THE GRADIENT OPERATOR

Prove the following identities:

- (a) $\nabla(|\mathbf{x}|) = \mathbf{x} |\mathbf{x}|^{-1} \equiv \hat{\mathbf{r}}$
- (b) $\nabla \ln |\mathbf{x}| = \mathbf{x} |\mathbf{x}|^{-2} = \hat{\mathbf{r}} |\mathbf{x}|^{-1}$
- (c) $\nabla |\mathbf{x}|^{-1} = -\mathbf{x} |\mathbf{x}|^{-3} = -\hat{\mathbf{r}} |\mathbf{x}|^{-2}$.

EXERCISE 2.2: PRACTICE WITH THE LAPLACIAN OPERATOR

Show that the Laplacian of the function

$$\psi = \frac{z x^2}{r^2} \quad (2.96)$$

is given by

$$\nabla^2 \psi = \frac{2 z (r^2 - 5 x^2)}{r^4}. \quad (2.97)$$

Perform the proof using both Cartesian coordinates as well as spherical coordinates (see Figure 8.1), making use of the following expressions for Laplacian operator acting on a scalar field

$$\nabla^2 \psi(x, y, z) = \frac{\partial^2 \psi}{\partial x^2} + \frac{\partial^2 \psi}{\partial y^2} + \frac{\partial^2 \psi}{\partial z^2} \quad (2.98a)$$

$$\nabla^2 \psi(\lambda, \phi, r) = \frac{1}{r^2 \cos \phi} \left[\frac{1}{\cos \phi} \frac{\partial^2 \psi}{\partial \lambda^2} + \frac{\partial}{\partial \phi} \left(\cos \phi \frac{\partial \psi}{\partial \phi} \right) + \cos \phi \frac{\partial}{\partial r} \left(r^2 \frac{\partial \psi}{\partial r} \right) \right]. \quad (2.98b)$$

EXERCISE 2.3: MORE PRACTICE WITH OPERATORS

Prove the following identities with $r \neq 0$:

- (a) $\nabla^2 r^{-1} = 0$

- (b) $\nabla \cdot (\mathbf{x}/r^3) = 0$
- (c) $\nabla \cdot (\mathbf{A} \wedge \mathbf{x}) = \mathbf{x} \cdot (\nabla \wedge \mathbf{A})$ for an arbitrary vector field $A(\mathbf{x})$.
- (d) $\nabla \wedge [\mathbf{x} f(r)] = 0$ for an arbitrary function $f(r) = f(|\mathbf{x}|)$.

EXERCISE 2.4: SOLID-BODY ROTATION

Define a velocity field according to

$$\mathbf{v} = \boldsymbol{\Omega} \wedge \mathbf{x} \quad (2.99)$$

with $\boldsymbol{\Omega}$ a spatially constant angular rotation velocity (e.g., rotation of the earth). This velocity field describes solid-body rotation as discussed in Section 33.5. Show that $2\boldsymbol{\Omega} = \nabla \wedge \mathbf{v}$. See also Exercise 33.1.

EXERCISE 2.5: DIVERGENCE-FREE AND IRROTATIONAL VECTOR

Let Φ be a harmonic function so that $\nabla^2 \Phi = 0$. Show that $\mathbf{v} = -\nabla \Phi$ satisfies

(a) $\nabla \cdot \mathbf{v} = 0$

(b) $\nabla \wedge \mathbf{v} = 0$.

In this way we prove that all harmonic scalar fields correspond to a divergence-free and curl-free vector field.

EXERCISE 2.6: CONSERVATIVE VECTOR FIELD AND SCALAR POTENTIAL

Show that the curl, $\nabla \wedge \mathbf{F}$, of the following vector field vanishes

$$\mathbf{F} = 2x z \hat{\mathbf{x}} + 2y z^2 \hat{\mathbf{y}} + (x^2 + 2y^2 z - 1) \hat{\mathbf{z}}. \quad (2.100)$$

Hence, deduce that \mathbf{F} is a *conservative vector field*, meaning that it can be written as the gradient of a scalar potential ψ according to $\mathbf{F} = -\nabla \psi$, where (to within an arbitrary constant)

$$\psi = -[x^2 z + (y z)^2 - z]. \quad (2.101)$$

EXERCISE 2.7: PRODUCT RULE IDENTITIES

Prove the following identities for vectors in \mathbb{R}^3 :

(a) $\mathbf{F} = \partial_n (F_n \mathbf{x}) - \mathbf{x} \nabla \cdot \mathbf{F}$

(b) $2F_m = [\mathbf{x} \wedge (\nabla \wedge \mathbf{F})]_m - \partial_m (\mathbf{x} \cdot \mathbf{F}) + \nabla \cdot (\mathbf{x} F_m)$.

Make use of Cartesian tensors and show all relevant steps, including use of the Levi-Civita tensor from Section 1.4.1 for the cross-product.

As discussed in Section 12.4.1 of [Bühler \(2014a\)](#), these product rule identities have use for the study of impulses imparted by a body force per volume, \mathbf{F} , to a fluid on an unbounded domain where the force has compact support (i.e., the force vanishes outside a finite domain). In that case the above product rule identities allow us to make use of the corresponding integral identities

$$\int \mathbf{F} dV = - \int \mathbf{x} \nabla \cdot \mathbf{F} dV = \frac{1}{2} \int \mathbf{x} \wedge (\nabla \wedge \mathbf{F}) dV. \quad (2.102)$$

EXERCISE 2.8: BELTRAMI FLOW

Beltrami flow is defined by velocity and vorticity fields satisfying

$$\nabla \cdot \mathbf{v} = 0 \quad (2.103a)$$

$$\boldsymbol{\omega} = \nabla \wedge \mathbf{v} = \lambda \mathbf{v} \quad (2.103b)$$

where λ is a constant. Show that the following velocity field is a Beltrami flow

$$\mathbf{v} = (A \sin z + C \cos y) \hat{x} + (B \sin x + A \cos z) \hat{y} + (C \sin y + B \cos x) \hat{z}, \quad (2.104)$$

where A, B, C are constants. Hint: the solution follows directly from computing

$$u = \frac{\partial w}{\partial y} - \frac{\partial v}{\partial z} \quad v = \frac{\partial u}{\partial z} - \frac{\partial w}{\partial x} \quad w = \frac{\partial v}{\partial x} - \frac{\partial u}{\partial y}. \quad (2.105)$$

EXERCISE 2.9: PRACTICE WITH PATH INTEGRALS

Consider the vector field

$$\mathbf{F} = x y^2 \hat{x} + 2 \hat{y} + x \hat{z}. \quad (2.106)$$

Let \mathcal{L} be a path parameterized by

$$x = c t \quad y = c/t \quad z = d \quad t \in [1, 2], \quad (2.107)$$

where c and d are constants. Show that the following identities hold

$$\int_{\mathcal{L}} \mathbf{F} dt = c^3 \ln 2 \hat{x} + 2 \hat{y} + \frac{3c}{2} \hat{z} \quad (2.108a)$$

$$\int_{\mathcal{L}} \mathbf{F} dy = -\frac{3c^4}{8} \hat{x} - c \hat{y} - c^2 \ln 2 \hat{z} \quad (2.108b)$$

$$\int_{\mathcal{L}} \mathbf{F} \cdot d\mathbf{x} = c^4 \ln 2 - c, \quad (2.108c)$$

where $d\mathbf{x} = \hat{x} dx + \hat{y} dy + \hat{z} dz$. Although all three integrals are computed along the same path, they are not necessarily of the same type. In particular, the first two integrals are vector fields, whereas the third integral is a scalar.

EXERCISE 2.10: STOKES' THEOREM ON A PLANE

Show that

$$\mathcal{I} = \oint_{\partial\mathcal{S}} [y(4x^2 + y^2) dx + x(2x^2 + 3y^2) dy] = \frac{\pi}{2} b a^3 \quad (2.109)$$

when integrating around the boundary of an ellipse \mathcal{S} defined by

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1, \quad (2.110)$$

where a, b are constants. Hint: make use of Stokes' Theorem on a plane, otherwise known as Green's Theorem. Also make use of the substitution $x = a \cos \phi$ and the identity

$$\int_{\pi}^0 \sin^2(2\phi) d\phi = -\frac{\pi}{2}. \quad (2.111)$$

EXERCISE 2.11: PRACTICE WITH GAUSS'S DIVERGENCE THEOREM

We here demonstrate the validity of Gauss's divergence theorem for a particular vector field

$$\mathbf{F} = \frac{\alpha \mathbf{x}}{(r^2 + a^2)^{3/2}}, \quad (2.112)$$

where α and a are constants and $r^2 = \mathbf{x} \cdot \mathbf{x}$ is the squared radial distance to a point. Using fluid mechanics jargon, we think of \mathbf{F} as a matter flux with physical dimensions of $M L^{-2} T^{-1}$ (mass length $^{-2}$ time $^{-1}$). Now compute the transport of \mathbf{F} through a spherical surface, \mathcal{S} , of radius $|\mathbf{x}| = a\sqrt{3}$

$$\Phi = \oint_{|\mathbf{x}|=a\sqrt{3}} \mathbf{F} \cdot \hat{\mathbf{n}} d\mathcal{S} = \frac{3\pi\alpha\sqrt{3}}{2}. \quad (2.113)$$

With \mathbf{F} a matter flux then Φ has physical dimensions of $M T^{-1}$, so that it is the *mass transport* through the spherical surface. Next, show that this transport is equal to the integral of the divergence over the volume of the sphere

$$\Phi = \int_{|\mathbf{x}|=a\sqrt{3}} \nabla \cdot \mathbf{F} dV. \quad (2.114)$$

We thus verify, for this particular vector field, the divergence theorem

$$\int_{\mathcal{R}} \nabla \cdot \mathbf{F} dV = \oint_{\partial\mathcal{R}} \mathbf{F} \cdot \hat{\mathbf{n}} d\mathcal{S}, \quad (2.115)$$

where $\hat{\mathbf{n}}$ is the outward normal on the bounding surface \mathcal{S} .

EXERCISE 2.12: MORE PRACTICE WITH GAUSS'S DIVERGENCE THEOREM

Prove the following identities, which are readily shown using Gauss's divergence theorem.

- (a) $\oint_{\partial\mathcal{R}} \mathbf{x} \cdot \hat{\mathbf{n}} d\mathcal{S} = 3 \int_{\mathcal{R}} dV = 3V$, where \mathcal{R} is a closed region bounded by $\partial\mathcal{R}$ and with volume $\int_{\mathcal{R}} dV = V$.
- (b) $\oint_{\partial\mathcal{R}} (\hat{\mathbf{n}} \wedge \mathbf{F}) d\mathcal{S} = \int_{\mathcal{R}} \nabla \wedge \mathbf{F} dV$, for an arbitrary vector field \mathbf{F} and with $\hat{\mathbf{n}}$ the outward normal on the bounding surface $\partial\mathcal{R}$. Hint: in a manner similar to the result shown in Section 2.7.2, make use of Gauss's theorem with $\mathbf{A} = \mathbf{F} \wedge \mathbf{C}$ where \mathbf{C} is a constant vector.
- (c) Let $\partial\mathcal{R}$ be a closed surface bounding a volume \mathcal{R} , and let \mathbf{x} denote the position vector of a point measured from an arbitrary origin. Prove the following

$$\oint_{\partial\mathcal{R}} \frac{\hat{\mathbf{n}} \cdot \mathbf{x}}{r^3} d\mathcal{S} = \begin{cases} 0 & \text{if the origin lies outside of } \partial\mathcal{R} \\ 4\pi & \text{if the origin lies inside of } \partial\mathcal{R}. \end{cases} \quad (2.116)$$

EXERCISE 2.13: HELMHOLTZ DECOMPOSITION FOR CORIOLIS ACCELERATION

Consider a vector field

$$\mathbf{F} = 2\boldsymbol{\Omega} \wedge \mathbf{v}, \quad (2.117)$$

where $\boldsymbol{\Omega}$ is a spatial constant and $\nabla \cdot \mathbf{v} = 0$. As seen in Part IV of this book, the vector \mathbf{F} is minus the Coriolis acceleration in the presence of an incompressible velocity. Since \mathbf{v} is non-divergent, there exists a vector potential so that

$$\mathbf{v} = \nabla \wedge \mathbf{B}. \quad (2.118)$$

Show that we can perform a Helmholtz decomposition

$$\mathbf{F} = 2\boldsymbol{\Omega} \wedge \mathbf{v} = -\nabla\Psi + \nabla \wedge \mathbf{A}, \quad (2.119)$$

where

$$-\nabla^2 \Psi = -2 \boldsymbol{\Omega} \cdot (\nabla \wedge \mathbf{v}) \quad (2.120a)$$

$$\nabla \wedge \mathbf{A} = (\boldsymbol{\Omega} \cdot \nabla) \mathbf{B} + \nabla \lambda, \quad (2.120b)$$

with λ an arbitrary gauge function.



Partial differential equations

Fluid mechanics is a classical field theory whose mathematical description involves partial differential equations (PDEs). Although most of these PDEs are nonlinear, some are linear. Regardless, an understanding of linear PDEs provides useful insights into the physics and maths of geophysical fluids. For this purpose, we here study linear PDEs.

READER'S GUIDE TO THIS CHAPTER

One can penetrate much of this book without reading this chapter. What this chapter adds, however, is an appreciation for the mathematical structure and physical behavior of the equations describing geophysical fluid mechanics. Indeed, it is remarkable how useful it is, both mathematically and physically, to develop a rudimentary understanding of linear PDEs. Furthermore, those aiming to develop solution methods, either analytical or numerical, should have a working knowledge of this chapter. Additionally, the Green's function method detailed in Chapter 4 assumes a firm understanding of the material in this chapter.

There are many resources devoted to the theory and application of PDEs throughout physics, engineering, and applied mathematics. Chapter 8 of *Hildebrand* (1976) offers a pedagogical starting point whereas *Stakgold* (2000a,b) thoroughly develops the theory and methods available for boundary value problems encountered in physics. *Duchateau and Zachmann* (1986) concisely summarize PDEs and provide nearly 300 worked exercises, with much of the presentation in this chapter following their treatment.

3.1	Loose threads	46
3.2	The advection equation	46
3.2.1	Constant advection velocity	46
3.2.2	Specifying the arbitrary functions resulting from PDEs	47
3.2.3	Further study	47
3.3	Characteristic curves for first order PDEs	48
3.3.1	General formulation	48
3.3.2	Examples	49
3.4	Classifying second order PDEs	50
3.5	Elliptic partial differential equations	50
3.5.1	Some general properties of Laplace's equation	51
3.5.2	Mean-value property of harmonic functions	51
3.5.3	Laplace's boundary value problem	52
3.5.4	Poisson's equation	53
3.5.5	Extended max-min principle for Poisson's equation	53
3.5.6	Poisson's boundary value problem	53

3.6	Parabolic partial differential equations	54
3.6.1	Initial and initial-boundary value problems	54
3.6.2	Smoothing property	55
3.6.3	Duhamel's superposition integral for the heat equation	56
3.6.4	Further study	57
3.7	Hyperbolic partial differential equations	57
3.7.1	Initial value problem for the infinite-domain wave equation	57
3.7.2	Domain of influence for wave signals	59
3.7.3	Helmholtz equation	60
3.7.4	Duhamel's superposition integral for the wave equation	60
3.7.5	Further study	61
3.8	Evolution of time averages	61
3.8.1	Time averages	62
3.8.2	Massaging the double time integrals	63
3.8.3	Making use of a double integral identity	63

3.1 Loose threads

- Basics of wave kinematics in Section 3.7

3.2 The advection equation

Consider a tracer concentration, C , which for present purposes is a scalar field that is a function of space and time. As derived in Section 48.5, the tracer concentration in the absence of diffusion satisfies the advection equation

$$\frac{\partial C}{\partial t} + \mathbf{v} \cdot \nabla C = 0. \quad (3.1)$$

The highest derivatives in both space and time are first order, indicating that the advection equation is a first order partial differential equation. It is a nonlinear PDE for *active* tracers such as temperature, where active tracers affect the velocity through changes to density and hence to pressure. In contrast, the advection equation is linear for *passive* tracers (e.g., colored dye, dust), which are tracers that do not significantly alter velocity (Section 16.2). We limit the present discussion to passive tracers so that the advection equation is linear.

3.2.1 Constant advection velocity

To expose the gist of the advection equation, consider one space dimension and let the advection velocity be constant in space and time,

$$\frac{\partial C}{\partial t} + U \frac{\partial C}{\partial x} = 0 \quad (3.2)$$

where U is a constant velocity in the $\hat{\mathbf{x}}$ direction. An inspired guess reveals that

$$C(x, t) = \Gamma(x - Ut) \quad (3.3)$$

is a general solution to equation (3.2), where Γ is an arbitrary differentiable function that is determined by the initial conditions of the tracer field. Verification of this result is readily found by

noting

$$\frac{\partial C}{\partial x} = \frac{d\Gamma(x - Ut)}{d(x - Ut)} \frac{\partial(x - Ut)}{\partial x} = \Gamma' \quad (3.4a)$$

$$\frac{\partial C}{\partial t} = \frac{d\Gamma(x - Ut)}{d(x - Ut)} \frac{\partial(x - Ut)}{\partial t} = -\Gamma' U. \quad (3.4b)$$

The functional dependence $x - Ut$ reveals that as time progresses with $U > 0$, an observer that moves in the positive \hat{x} direction with a speed U maintains a constant argument to the tracer concentration. This behavior means that the tracer concentration is transported by advection with a speed U without changing its structure. We illustrate this behavior in Figure 3.1.

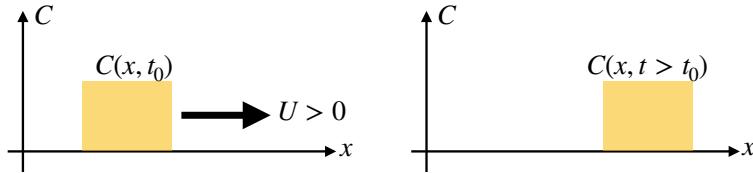


FIGURE 3.1: Illustrating the advection of a scalar field resulting from a constant advection velocity $v = U\hat{x}$ with $U > 0$. The initially square pulse of tracer concentration is translated, unchanged, by the constant advection velocity.

3.2.2 Specifying the arbitrary functions resulting from PDEs

As revealed from the above example, the solution to a PDE is typically given in terms of an arbitrary function with a specified dependence on the independent variables. The function itself is unspecified without additional information from initial and/or boundary conditions. For example, consider an initial tracer concentration in the form of a sine-wave

$$C(x, t = 0) = C_0 \sin x, \quad (3.5)$$

and allow the domain to be infinite in extent (no boundaries). When advected by a constant advection velocity, the solution to the advection equation is a tracer concentration in the form of a sine-wave moving in the positive \hat{x} direction with speed U

$$C(x, t) = C_0 \sin(x - Ut). \quad (3.6)$$

The arbitrary functional degree of freedom is reminiscent of ordinary differential equations, whose solutions are specified up to unknown constants with values set by initial and/or boundary conditions.

3.2.3 Further study

Advection plays a fundamental role in the transport of matter, energy, and momentum within fluids. As seen in our discussion of fluid kinematics in Chapter 13, advection appears in the fluid mechanical equations when viewing the fluid from the fixed laboratory or *Eulerian* reference frame. We thus encounter advection throughout this book, with further development of the mathematics and physics provided in Sections 48.5 and 48.6.

3.3 Characteristic curves for first order PDEs

The advection equation is the canonical first order PDE commonly found in fluid mechanics. A more general form for a first order PDE in one space dimension is given by¹

$$P(x, t, \psi) \frac{\partial \psi}{\partial x} + Q(x, t, \psi) \frac{\partial \psi}{\partial t} = R(x, t, \psi), \quad (3.7)$$

where P , Q , and R are arbitrary smooth functions. This PDE is linear if P , Q , and R are independent of ψ , and quasi-linear if P and Q are independent of ψ and R is at most a linear function of ψ . In this section we develop a formalism that determines the functional dependence of solutions to these PDEs. This *method of characteristics* is quite useful for exposing general properties of the solutions, even for those cases where the solution is not analytically available.

3.3.1 General formulation

In the first order PDE given by equation (3.7), assume there is a functional relation

$$\Upsilon(x, t, \psi) = \text{constant} \quad (3.8)$$

that determines ψ consistent with the PDE (3.7). We refer to Υ as an *integral surface* that specifies a solution to the PDE. For Υ to specify an integral surface it must satisfy

$$d\Upsilon = 0, \quad (3.9)$$

so that

$$\frac{d\Upsilon}{dt} = 0 = \frac{\partial \Upsilon}{\partial \psi} \frac{\partial \psi}{\partial t} + \frac{\partial \Upsilon}{\partial t} \quad (3.10a)$$

$$\frac{d\Upsilon}{dx} = 0 = \frac{\partial \Upsilon}{\partial \psi} \frac{\partial \psi}{\partial x} + \frac{\partial \Upsilon}{\partial x}. \quad (3.10b)$$

Assuming $\partial \Upsilon / \partial \psi \neq 0$ we can write

$$\frac{\partial \psi}{\partial t} = -\frac{\partial \Upsilon / \partial t}{\partial \Upsilon / \partial \psi} \quad \text{and} \quad \frac{\partial \psi}{\partial x} = -\frac{\partial \Upsilon / \partial x}{\partial \Upsilon / \partial \psi} \quad (3.11)$$

so that the first order PDE (3.7) takes on the equivalent form

$$P \frac{\partial \Upsilon}{\partial x} + Q \frac{\partial \Upsilon}{\partial t} + R \frac{\partial \Upsilon}{\partial \psi} = 0. \quad (3.12)$$

Through this construct we have brought the functions P, Q, R into a symmetric form. Consistency of equation (3.12) with

$$d\Upsilon = \frac{\partial \Upsilon}{\partial x} dx + \frac{\partial \Upsilon}{\partial t} dt + \frac{\partial \Upsilon}{\partial \psi} d\psi = 0, \quad (3.13)$$

requires

$$P = \mu dx \quad \text{and} \quad Q = \mu dt \quad \text{and} \quad R = \mu d\psi, \quad (3.14)$$

for some non-dimensional function $\mu(x, t, \psi)$. Eliminating the unknown function μ renders the ordinary differential equations

$$\frac{dx}{P} = \frac{dt}{Q} = \frac{d\psi}{R}. \quad (3.15)$$

Paths in (x, t, ψ) space that satisfy these differential equations are known as *characteristic curves*. Note that if any one of the functions P , Q , or R vanish, then we merely remove that term from these relations.

¹We here summarize the presentation in Section 8.2 of [Hildebrand \(1976\)](#).

3.3.2 Examples

We now illustrate the method of characteristics by considering some examples. First consider the linear homogeneous advection equation with constant advection speed

$$U \frac{\partial \psi}{\partial x} + \frac{\partial \psi}{\partial t} = 0, \quad (3.16)$$

in which we identify $P = U$, $Q = 1$, and $R = 0$. The single ordinary differential equation defining the characteristic curve is given by

$$\frac{dx}{U} = \frac{dt}{1}, \quad (3.17)$$

so that characteristics are given by the family of space-time lines

$$x - Ut = \alpha \quad (3.18)$$

with α an arbitrary constant. These lines determine the paths in space-time along which advective signals are transmitted.

Now add a constant source to the linear advection equation

$$U \frac{\partial \psi}{\partial x} + \frac{\partial \psi}{\partial t} = R. \quad (3.19)$$

The two ODEs defining the characteristic curve are

$$\frac{dx}{U} = \frac{dt}{1} = \frac{d\psi}{R}. \quad (3.20)$$

In addition to the relation $x - Ut = \alpha_1$ determined from the homogeneous case, we also have $\psi - Rt = \alpha_2$ for α_2 an arbitrary constant. Hence, the characteristic equations render the general solution of the form

$$\Gamma[x - Ut, \psi - Rt] = 0, \quad (3.21)$$

for Γ an arbitrary function. One example solution is given by

$$\psi = f(x - Ut) + Rt \quad (3.22)$$

for an arbitrary smooth function f . This solution has the form of a traveling signal, $f(x - Ut)$, plus a growing source, Rt .

For the final example, consider the linear advection equation with non-constant coefficients and non-constant source

$$x \frac{\partial \psi}{\partial x} + t \frac{\partial \psi}{\partial t} = \psi, \quad (3.23)$$

in which the ODEs determining the characteristics are given by

$$\frac{dx}{x} = \frac{dt}{t} = \frac{d\psi}{\psi}. \quad (3.24)$$

We are thus led to the relations

$$\frac{t}{x} = \alpha_1 \quad \frac{\psi}{x} = \alpha_2. \quad (3.25)$$

Hence, the general solution of the PDE (3.23) is given by

$$\Gamma(t/x, \psi/x) = 0 \Rightarrow \psi = x F(t/x) \quad (3.26)$$

for an arbitrary smooth function F .

3.4 Classifying second order PDEs

There are many second order PDEs appearing in fluid mechanics, a general form of which in one space dimension is given by

$$A \frac{\partial^2 \psi}{\partial x^2} + B \frac{\partial^2 \psi}{\partial x \partial t} + C \frac{\partial^2 \psi}{\partial t^2} = \Lambda. \quad (3.27)$$

For linear PDEs, A, B, C are arbitrary functions of space and time that are independent of ψ . Furthermore, Λ is a function of space and time and at most a linear function of ψ and its derivatives. The most general solution to a linear PDE consists of the sum of any *particular solution* and a solution to the homogeneous problem (where $\Lambda = 0$).

The terms involving second derivatives in equation (3.27) are of principle importance for determining the character of the solutions, with importance placed on the sign of the discriminant $B^2 - 4AC$. By analogy with conic sections we classify 2nd order PDEs as follows

$$\text{PDE form} = \begin{cases} \text{hyperbolic} & B^2 - 4AC > 0 \\ \text{elliptic} & B^2 - 4AC < 0 \\ \text{parabolic} & B^2 - 4AC = 0. \end{cases} \quad (3.28)$$

We can further motivate this terminology by considering the case of a homogeneous constant coefficient PDE and an assumed solution of the form

$$\psi(x, t) = f(mx + t). \quad (3.29)$$

Plugging into the 2nd order PDE (3.27) with $\Lambda = 0$ leads to

$$A m^2 + B m + C = 0. \quad (3.30)$$

The two solutions m_1 and m_2 are both real for the hyperbolic case, conjugate complex for the elliptic case, and a perfect square for the parabolic case.

3.5 Elliptic partial differential equations

The elliptic case from Section 3.4 has discriminant $B^2 - 4AC < 0$, in which case there are two imaginary characteristics. The canonical example elliptic equation is Laplace's equation

$$\frac{\partial^2 \psi}{\partial x^2} + \frac{\partial^2 \psi}{\partial y^2} = 0, \quad (3.31)$$

where we converted from t as an independent variable to the space coordinate y . Formally, this transition is realized by setting $t = iy$, where $i = \sqrt{-1}$. Laplace's equation is satisfied by time-independent (i.e., *steady state*) solutions to the heat/diffusion equation (Section 3.6). Another common elliptic equation is Poisson's equation, which results from adding a source to Laplace's equation

$$\frac{\partial^2 \psi}{\partial x^2} + \frac{\partial^2 \psi}{\partial y^2} = \Lambda. \quad (3.32)$$

As there is no time present in the Laplace and Poisson equations, their solutions are global. That is, “signals” propagate instantaneously so that their structure is fully determined by specified boundary conditions or boundary *data*. Strictly speaking this behavior is not physical since all signals have a finite propagation speed no greater than the speed of light. However, an infinite

speed can be a useful mathematical construct. For example, acoustic signals in fluids propagate much faster than other waves and particle speeds, and they carry a very small energy.² For many purposes it is thus suitable to assume acoustic speeds are infinite, and in so doing we *filter* them out of the dynamical equations. In the process, the hyperbolic equation describing acoustic signals is converted into an elliptic equation.

3.5.1 Some general properties of Laplace's equation

Solutions to Laplace's equation, $\nabla^2\psi = 0$, are known as *harmonic functions*. Here are some example harmonic functions for two space dimensions:

$$\psi = x^3 - 3xy^2 \quad \psi = \ln(x^2 + y^2) \quad \psi = e^{\gamma x} \cos(\gamma y) \quad \psi = ax + by, \quad (3.33)$$

for arbitrary constants a, b, γ . Furthermore, with

$$\nabla^2(\psi\phi) = \psi\nabla^2\phi + 2\nabla\phi \cdot \nabla\psi + \phi\nabla^2\psi, \quad (3.34)$$

we see that the product of two harmonic functions ($\nabla^2\psi = \nabla^2\phi = 0$) is itself harmonic if and only if their gradients are orthogonal, $\nabla\psi \cdot \nabla\phi = 0$. In the remainder of this section we present, without proof, some general properties of harmonic functions and develop self-consistency conditions for the boundary data appearing in the Laplacian boundary value problem defined below by equations (3.37a)-(3.37b).

3.5.2 Mean-value property of harmonic functions

Harmonic functions possess a remarkable *mean-value* property. This property says that the value of a harmonic function, ψ , at a point \mathbf{x}_o within an open region of \mathcal{R} equals to the average of ψ taken over the surface of a sphere within \mathcal{R} centered at \mathbf{x}_o . In equations this property states that

$$\psi(\mathbf{x}_o) = \frac{\oint_{S_R} \psi(\mathbf{x}) dS}{\oint_{S_R} dS}, \quad (3.35)$$

where S_R is a sphere with radius R centered at \mathbf{x}_o with “area” given by

$$\oint_{S_R} dS = \begin{cases} 2\pi R & n = 2 \text{ space dimensions} \\ 4\pi R^2 & n = 3 \text{ space dimensions.} \end{cases} \quad (3.36)$$

We illustrate this property in Figure 3.2.

The mean-value property of harmonic functions holds anywhere within the domain where $\nabla^2\psi = 0$, so long as the sphere is fully contained within that domain. It implies that there can be no extrema of ψ within the domain, since if there was an extrema then it could not satisfy the mean-value property. Hence, all extrema of harmonic functions must exist on the domain boundary. These properties lend mathematical support for considering harmonic functions to be solutions to continuous physical systems that are in equilibrium or a steady state. As a physical example, consider a temperature field, $T(\mathbf{x})$, in a region with zero heat sources and zero fluid flow. As shown in Section 48.3.4, the steady state temperature satisfies $\nabla^2T = 0$, and as such it is harmonic and hence has no extrema within the domain.

²A scuba diver feeling the beat of a ship underwater, or an audience member at a rock concert may question this statement. However, acoustic energy is in fact tiny relative to planetary waves and gravity waves, and utterly negligible for studies of large scale geophysical fluid motions.

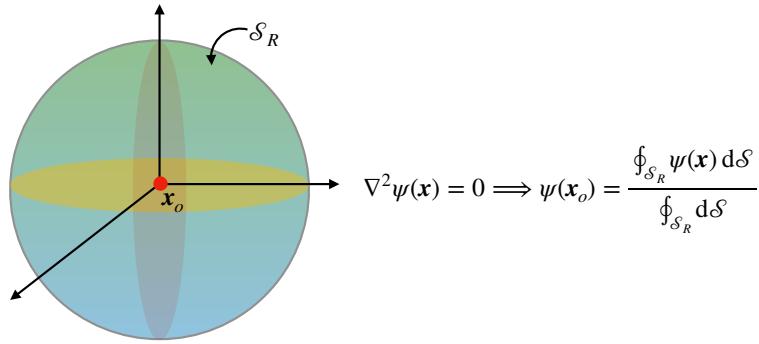


FIGURE 3.2: The value of a harmonic function at a point \mathbf{x}_o equals to the area average of the function over a sphere centered at \mathbf{x}_o . We here illustrate this property for 3-dimensions, but it holds for arbitrary space dimensions.

3.5.3 Laplace's boundary value problem

Laplace's equation requires boundary conditions to fully specify the unknown harmonic function. We here consider the Laplacian *boundary value problem* in the form

$$\nabla^2 \psi = 0 \quad \mathbf{x} \in \mathcal{R} \quad (3.37a)$$

$$\alpha \psi + \beta \hat{\mathbf{n}} \cdot \nabla \psi = \sigma \quad \mathbf{x} \in \partial \mathcal{R}, \quad (3.37b)$$

where \mathcal{R} is a smooth and simply connected volume, $\partial \mathcal{R}$ is the boundary surface enclosing \mathcal{R} and with outward normal $\hat{\mathbf{n}}$, and α , β , and σ are given boundary data functions.

We can establish constraints on the boundary conditions that lead to a self-consistent Laplacian boundary value problem (3.37a)-(3.37b). We do so through the use of Gauss's divergence theorem (Section 2.7) in which integration over the full domain leads to

$$0 = \int_{\mathcal{R}} \nabla^2 \psi \, dV = \oint_{\partial \mathcal{R}} \hat{\mathbf{n}} \cdot \nabla \psi \, dS. \quad (3.38)$$

In physical applications the boundary condition (3.37b) usually appears with either $\alpha = 0$ or $\beta = 0$, and these two cases are associated with distinct self-consistency constraints.

Dirichlet boundary condition

The case with $\beta = 0$ is referred to as a *Dirichlet* boundary condition whereby

$$\psi = \sigma \quad \mathbf{x} \in \partial \mathcal{R}, \quad (3.39)$$

where we set $\alpha = 1$ without loss of generality. In this case all boundary data result in a self-consistent Laplacian boundary value problem so there is no constraint on σ . Thinking again about temperature, this boundary condition specifies the temperature on the domain boundary to equal $T = \sigma$, with all boundary functions σ consistent with a harmonic temperature distribution within the domain interior.

Neumann boundary condition

The case with $\alpha = 0$ results in a *Neumann* boundary condition. Without loss of generality we set $\beta = 1$ and reach the following self-consistency condition

$$\oint_{\partial \mathcal{R}} \sigma \, dS = 0. \quad (3.40)$$

That is, a self-consistent boundary condition for Laplace's equation with a Neumann boundary condition must satisfy this surface integral constraint. For the temperature example, this boundary condition says that to realize a steady state harmonic temperature distribution within a region, we can at most apply a zero area averaged boundary heating. If the boundary constraint (3.40) is not satisfied, then the interior temperature field cannot be harmonic so that it will not be in a steady state.

3.5.4 Poisson's equation

The generic boundary value problem for Poisson's equation takes on the form

$$\nabla^2\psi = \Lambda \quad \mathbf{x} \in \mathcal{R} \quad (3.41a)$$

$$\alpha\psi + \beta\hat{\mathbf{n}} \cdot \nabla\psi = \sigma \quad \mathbf{x} \in \partial\mathcal{R}, \quad (3.41b)$$

where $\Lambda(\mathbf{x})$ is a specified source function. We here present, without proof, some general properties of solutions to Poisson's equation and develop self-consistency conditions for the boundary data appearing in equation (3.41b).

3.5.5 Extended max-min principle for Poisson's equation

A subharmonic function is one where

$$\nabla^2\psi = \Lambda \geq 0 \quad \mathbf{x} \in \mathcal{R}. \quad (3.42)$$

Here, the source function makes the curvature of a subharmonic function positive. Correspondingly, every point within \mathcal{R} satisfies the minimum principle

$$\psi(\mathbf{x}_o) \leq \frac{\oint_{\mathcal{S}_R} \psi(\mathbf{x}) d\mathcal{S}}{\oint_{\mathcal{S}_R} d\mathcal{S}}, \quad (3.43)$$

for spheres, \mathcal{S}_R , that are fully within \mathcal{R} . The signs switch for superharmonic functions whereby $\nabla^2\psi \leq 0$ for $\mathbf{x} \in \mathcal{R}$.

Returning to the temperature example, consider a temperature field in a region with a positive heat source, $\Lambda > 0$. The steady state temperature in the presence of zero fluid flow satisfies Poisson's equation $\nabla^2T = \Lambda \geq 0$ for regions with the heat source. The minimum principle (3.43) means that the temperature at any point within the heating region is less than the temperature averaged over a sphere centered on the point, so long as the sphere remains within the region of heating. It is only in the absence of a heat source or sink, where $\nabla^2T = 0$, that we recover the mean-value property of harmonic functions given by equation (3.35).

3.5.6 Poisson's boundary value problem

We follow the method in Section 3.5.3 to develop constraints on the boundary conditions applied as part of the Poisson boundary value problem (3.41a)-(3.41b). Use of Gauss's divergence theorem leads to the constraint

$$\oint_{\partial\mathcal{R}} \hat{\mathbf{n}} \cdot \nabla\psi d\mathcal{S} = \int_{\mathcal{R}} \Lambda dV. \quad (3.44)$$

We separately consider the Dirichlet and Neumann cases.

Dirichlet boundary condition

The Dirichlet condition with $\beta = 0$ leads to

$$\psi = \sigma \quad \mathbf{x} \in \partial\mathcal{R}. \quad (3.45)$$

Just as for Laplace's boundary value problem, all boundary data result in a self-consistent Poisson boundary value problem so there is no constraint on σ . Thinking again about temperature, this boundary condition specifies the temperature on the domain boundary to equal $T = \sigma$, with all boundary functions σ consistent with the interior heating Λ and a temperature field satisfying $\nabla^2 T = \Lambda$ within the interior.

Neumann boundary condition

The Neumann boundary condition leads to the following self-consistency condition

$$\oint_{\partial\mathcal{R}} \sigma \, d\mathcal{S} = \int_{\mathcal{R}} \Lambda \, dV. \quad (3.46)$$

That is, a self-consistent boundary condition for Poisson's equation with a Neumann boundary condition must satisfy this surface integral constraint. For the temperature example, this boundary condition says that the area integrated boundary data must be consistent with the volume integrated source function in order for the temperature to satisfy Poisson's equation. Otherwise, the temperature field will evolve in time and thus not be in a steady state.

3.6 Parabolic partial differential equations

The parabolic case from Section 3.4, $B^2 - 4AC = 0$, contains a single real characteristic. The canonical example is the *heat equation*, which is also known as the *diffusion equation*

$$\frac{\partial \psi}{\partial t} = \kappa \frac{\partial^2 \psi}{\partial x^2}, \quad (3.47)$$

where $\kappa > 0$ is the kinematic diffusivity with dimensions of squared length per time.

3.6.1 Initial and initial-boundary value problems

The *Cauchy Problem* is the name given to the initial value problem for the heat equation over all space, here given by Euclidean space

$$\frac{\partial \psi}{\partial t} = \kappa \nabla^2 \psi \quad \mathbf{x} \in \mathbb{R}^n, t > 0 \quad (3.48a)$$

$$\psi(\mathbf{x}, t = 0) = \sigma(\mathbf{x}) \quad \mathbf{x} \in \mathbb{R}^n \quad (3.48b)$$

$$|\psi(\mathbf{x}, t)| < \infty \quad \mathbf{x} \in \mathbb{R}^n, t > 0. \quad (3.48c)$$

The general initial-boundary value problem over a finite domain \mathcal{R} takes the form

$$\frac{\partial \psi}{\partial t} = \kappa \nabla^2 \psi \quad \mathbf{x} \in \mathcal{R}, t > 0 \quad (3.49a)$$

$$\psi(\mathbf{x}, t = 0) = \sigma(\mathbf{x}) \quad \mathbf{x} \in \mathcal{R} \quad (3.49b)$$

$$\alpha(\mathbf{x}) \psi(\mathbf{x}, t) + \beta(\mathbf{x}) \hat{\mathbf{n}} \cdot \nabla \psi(\mathbf{x}, t) = g(\mathbf{x}, t) \quad \mathbf{x} \in \partial\mathcal{R}, t > 0, \alpha \beta \geq 0. \quad (3.49c)$$

Following from the discussion of Laplace's and Poisson's boundary value problems, choices for the boundary functions α and β impact on the character of the boundary conditions. The Neumann condition is most commonly applied to set the flux of a tracer or temperature at the boundaries. The alternative use of the Dirichlet condition is commonly applied for idealized "diagnostic" tracers in geophysical fluid applications.

3.6.2 Smoothing property

The extended max-min principle from Section 3.5.5 holds also for the heat equation, which is consistent with solutions to the heat equation generally decaying their initial condition towards zero by reducing the amplitude of all extrema. Hence, no extrema are introduced in the interior of the domain by the heat equation; extrema only arise via boundary and/or initial conditions. Furthermore, the steady state limit of the heat equation is a harmonic function, and so solves Laplace's equation whereby the mean-value property holds (Section 3.5.2). These characteristics of the heat equation are generally shared by all linear parabolic PDEs.

Illustrating the smoothing property in a finite domain

We illustrate the smoothing property for the specific case of the one-dimensional initial-boundary value problem with homogeneous Dirichlet boundary conditions

$$\frac{\partial \psi}{\partial t} = \kappa \frac{\partial^2 \psi}{\partial x^2} \quad 0 < x < L, t > 0 \quad (3.50a)$$

$$\psi(x, t) = I(x) \quad 0 < x < L, t = 0 \quad (3.50b)$$

$$\psi(0, t) = \psi(L, t) = 0 \quad t > 0. \quad (3.50c)$$

A variety of methods, such as separation of variables, can be used to derive the following Fourier series solution

$$\psi(x, t) = \sum_{n=1}^{\infty} I_n e^{-\kappa t (n\pi/L)^2} \sin(n\pi x/L) \quad \text{with} \quad I_n = \frac{2}{L} \int_0^L I(x) \sin(n\pi x/L) dx. \quad (3.51)$$

As per the smoothing property of diffusion, note how the amplitude of each Fourier mode decays exponentially in time.

Smoothing property for an initial value problem on the real line

Now consider the one-dimensional heat equation on the real line, with the only boundary conditions being regularity at infinity

$$\frac{\partial \psi}{\partial t} = \kappa \frac{\partial^2 \psi}{\partial x^2} \quad -\infty < x < \infty, t > 0 \quad (3.52a)$$

$$\psi(x, t) = I(x) \quad -\infty < x < \infty, t = 0. \quad (3.52b)$$

One can show by direct differentiation that the following Gaussian is a solution

$$\psi(x, t) = \frac{1}{\sqrt{4\pi\kappa t}} \int_{-\infty}^{\infty} I(\xi) \exp\left[-\frac{(x-\xi)^2}{4\kappa t}\right] d\xi. \quad (3.53)$$

Again, this function smooths/damps the initial condition function $I(x)$ as time increases.

3.6.3 Duhamel's superposition integral for the heat equation

Consider a scalar field that starts from zero initial conditions and evolves in the presence of a source

$$\frac{\partial \Psi}{\partial t} = \kappa \nabla^2 \Psi + f(\mathbf{x}) \quad \mathbf{x} \in \mathbb{R}^n, t > 0 \quad (3.54a)$$

$$\Psi(\mathbf{x}, t) = 0 \quad \mathbf{x} \in \mathbb{R}^n, t = 0. \quad (3.54b)$$

Now consider the converse, in which another scalar field evolves without a source and yet is initialized according to the source

$$\frac{\partial \psi}{\partial t} = \kappa \nabla^2 \psi \quad \mathbf{x} \in \mathbb{R}^n, t > 0 \quad (3.55a)$$

$$\psi(\mathbf{x}, t) = f(\mathbf{x}) \quad \mathbf{x} \in \mathbb{R}^n, t = 0. \quad (3.55b)$$

The two scalar fields are related by *Duhamel's superposition integral*

$$\Psi(\mathbf{x}, t) = \int_0^t \psi(\mathbf{x}, t - \tau) d\tau. \quad (3.56)$$

We verify the connection by direct differentiation

$$\frac{\partial \Psi(\mathbf{x}, t)}{\partial t} = \psi(\mathbf{x}, 0) + \int_0^t \frac{\partial \psi(\mathbf{x}, t - \tau)}{\partial t} d\tau = f(\mathbf{x}) + \kappa \nabla^2 \Psi(\mathbf{x}, t). \quad (3.57)$$

Duhamel's superposition integral allows us to move the source from the partial differential operator into the initial conditions, and vice-versa. It says that the forced solution $\Psi(\mathbf{x}, t)$ is built by time integrating the “retarded” values of the unforced solution ψ from the initial time $t = 0$ to the current time t . Note that a more general presentation allows for the source function to be a function of time, $f(\mathbf{x}, t)$, in which case we must develop a family of solutions, $\psi_f(\mathbf{x}, t; \tau)$, generated by reinitializing $\psi_f(\mathbf{x}, t = \tau; \tau) = f(\mathbf{x}, \tau)$ and then superposing the members of this family to generate $\Psi(\mathbf{x}, t)$.

As an example, consider the initial value problem for the heat equation on a line as given by equations (3.52a)-(3.52b)

$$\frac{\partial \psi}{\partial t} = \kappa \frac{\partial^2 \psi}{\partial x^2} \quad -\infty < x < \infty, t > 0 \quad (3.58a)$$

$$\psi(x, t) = f(x) \quad -\infty < x < \infty, t = 0, \quad (3.58b)$$

whose solution is given by the Gaussian function in equation (3.53). Duhamel's superposition integral (3.56) says that

$$\Psi(x, t) = \int_0^t \psi(x, t - \tau) d\tau = \int_0^t \frac{1}{\sqrt{4\pi\kappa(t-\tau)}} \int_{-\infty}^{\infty} f(\xi) \exp\left[-\frac{(x-\xi)^2}{4\kappa(t-\tau)}\right] d\xi d\tau \quad (3.59)$$

satisfies the forced (inhomogeneous) initial value problem with zero initial condition

$$\frac{\partial \Psi}{\partial t} = \kappa \frac{\partial^2 \Psi}{\partial x^2} + f(x) \quad -\infty < x < \infty, t > 0 \quad (3.60a)$$

$$\Psi(x, t) = 0 \quad -\infty < x < \infty, t = 0. \quad (3.60b)$$

We make use of this result in our discussion of Green's functions in Chapter 4.

3.6.4 Further study

We examine physical and mathematical properties of the heat/diffusion equation in Sections 48.3 and 48.4. Section 9.11 of [Hildebrand \(1976\)](#) offers a lucid discussion of Duhamel's superposition integral.

3.7 Hyperbolic partial differential equations

The hyperbolic case from Section 3.4 has $B^2 - 4AC > 0$ and thus contains two real characteristics. The canonical example of a hyperbolic PDE is the linear homogeneous wave equation

$$\frac{\partial^2 \psi}{\partial t^2} - U^2 \frac{\partial^2 \psi}{\partial x^2} = 0. \quad (3.61)$$

Solutions have the form of a moving wave in both directions (the two wave characteristics)

$$\psi(x, t) = \mathcal{F}(x - Ut) + \mathcal{G}(x + Ut), \quad (3.62)$$

where \mathcal{F} and \mathcal{G} are differentiable functions whose form is determined by the initial conditions. Note that we can factor the differential operator into the form

$$(\partial_t - U \partial_x)(\partial_t + U \partial_x)\psi = 0. \quad (3.63)$$

Consequently, if either one of the linear first-order PDEs are satisfied

$$(\partial_t - U \partial_x)\psi = 0 \quad (3.64a)$$

$$(\partial_t + U \partial_x)\psi = 0 \quad (3.64b)$$

then ψ will satisfy the full wave equation. These first-order PDEs are the one-dimensional advection equations considered in Section 3.3 with opposite advection direction, and each of which has a single characteristic. In this manner, we can think of advection by constant velocity as the square root of the wave equation. Similarly, some disciplines refer to the linear advection equation (3.2), with constant advection speed, as the *one-way wave equation*.

3.7.1 Initial value problem for the infinite-domain wave equation

Since there are two time derivatives, specification of a solution requires initial conditions for the field and its first time derivative. To illustrate the structure of a solution to the wave equation, we develop a solution to the *Cauchy problem*, which is the initial value problem for the one-dimensional wave equation on the real line (infinite spatial domain so no boundary conditions)

$$\frac{\partial^2 \psi}{\partial t^2} = c^2 \frac{\partial^2 \psi}{\partial x^2} \quad -\infty < x < \infty, t > 0 \quad \text{wave equation on a line} \quad (3.65a)$$

$$\psi = F(x) \quad -\infty < x < \infty, t = 0 \quad \text{initial condition} \quad (3.65b)$$

$$\frac{\partial \psi}{\partial t} = G(x) \quad -\infty < x < \infty, t = 0 \quad \text{initial tendency}, \quad (3.65c)$$

where the initial condition data, F, G , are arbitrary functions of space and c is a constant wave speed. Following from the discussion of characteristics in Section 3.3, we are motivated to transform the wave equation into wave characteristic coordinates

$$\xi = x + ct \quad \text{and} \quad \eta = x - ct \implies \frac{\xi + \eta}{2} = x \quad \text{and} \quad \frac{\xi - \eta}{2c} = t. \quad (3.66)$$

Wave signals propagate in directions defined by constant ξ and η , so that these coordinates isolate the signal transmission. Furthermore, as we will see, this coordinate transformation facilitates a direct integration of the wave equation.

Transformation to characteristic coordinates

To help organize the transformation to characteristic coordinates, we write equation (3.66) as a matrix-vector equation

$$\begin{bmatrix} \xi \\ \eta \end{bmatrix} = \begin{bmatrix} c & 1 \\ -c & 1 \end{bmatrix} \begin{bmatrix} t \\ x \end{bmatrix} \iff \begin{bmatrix} t \\ x \end{bmatrix} = \frac{1}{2} \begin{bmatrix} c^{-1} & -c^{-1} \\ 1 & 1 \end{bmatrix} \begin{bmatrix} \xi \\ \eta \end{bmatrix}. \quad (3.67)$$

Furthermore, define

$$x^n = (x^0, x^1) = (x, t) \quad \text{and} \quad x^{\bar{n}} = (x^{\bar{0}}, x^{\bar{1}}) = (\xi, \eta) \quad (3.68)$$

and use index notation from Chapter 1 so that the transformation (3.67) takes the tidy form

$$x^{\bar{n}} = \Lambda^{\bar{n}}_m x^m \iff x^n = \Lambda^n_{\bar{m}} x^{\bar{m}}, \quad (3.69)$$

where the transformation matrices are given by

$$\Lambda^{\bar{n}}_m = \begin{bmatrix} c & 1 \\ -c & 1 \end{bmatrix} \quad \text{and} \quad \Lambda^n_{\bar{m}} = \frac{1}{2} \begin{bmatrix} c^{-1} & -c^{-1} \\ 1 & 1 \end{bmatrix}. \quad (3.70)$$

Note the use of an upstairs position for the row index on the transformation matrix, which conforms to the use with general tensors from Chapter 6. For present purposes there is no significance to this position placement; it is only used here to anticipate the general tensor machinery. The coordinate transformation (3.69) and the transformation matrices (3.70) then lead to the partial derivative relationship

$$\partial_{\bar{n}} = \Lambda^m_{\bar{n}} \partial_m \iff \partial_n = \Lambda^{\bar{m}}_n x_{\bar{m}}, \quad (3.71)$$

so that

$$\frac{\partial^2}{\partial x^2} = \left[\frac{\partial}{\partial \xi} + \frac{\partial}{\partial \eta} \right]^2 = \frac{\partial^2}{\partial \xi^2} + 2 \frac{\partial}{\partial \xi} \frac{\partial}{\partial \eta} + \frac{\partial^2}{\partial \eta^2} \quad (3.72a)$$

$$c^{-2} \frac{\partial^2}{\partial t^2} = \left[\frac{\partial}{\partial \xi} - \frac{\partial}{\partial \eta} \right]^2 = \frac{\partial^2}{\partial \xi^2} - 2 \frac{\partial}{\partial \xi} \frac{\partial}{\partial \eta} + \frac{\partial^2}{\partial \eta^2}. \quad (3.72b)$$

General solution for the initial value problem

The operator transformations (3.72a) and (3.72b) bring the initial value problem (3.65a)-(3.65c) into

$$\frac{\partial^2 \psi}{\partial \xi \partial \eta} = 0 \quad -\infty < \eta < \xi < \infty \quad (3.73a)$$

$$\psi(\xi, \eta) = F(\xi) \quad -\infty < \xi < \infty, \xi = \eta \quad (3.73b)$$

$$\frac{\partial \psi}{\partial \xi} - \frac{\partial \psi}{\partial \eta} = c^{-1} G(\xi) \quad -\infty < \xi < \infty, \xi = \eta. \quad (3.73c)$$

Integrating equation (3.73a) in two steps leads to $\partial_\xi \psi = \theta(\xi)$ so that

$$\psi(\xi, \eta) = \Phi(\eta) + \int^\xi \theta(s) ds \equiv \Phi(\eta) + \Theta(\xi), \quad (3.74)$$

for two functions $\Phi(\eta)$ and $\Theta(\xi)$. The initial conditions (3.73b) and (3.73c) determine relations between $\Phi(\eta)$ and $\Theta(\xi)$ and the initial data

$$\Theta(\xi) = \frac{1}{2} \left[F(\xi) + \frac{1}{c} \int^\xi G(s) ds \right] \quad (3.75a)$$

$$\Phi(\eta) = \frac{1}{2} \left[F(\eta) - \frac{1}{c} \int^\eta G(s) ds \right], \quad (3.75b)$$

in which case the general solution to the initial value problem (3.65a)-(3.65c) takes the form

$$\psi(x, t) = \frac{1}{2} [F(x + ct) + F(x - ct)] + \frac{1}{2c} \int_{x-ct}^{x+ct} G(s) ds, \quad (3.76)$$

where we reintroduced the variables (x, t) . This solution is known as the *D'Alembert formula* for the wave equation. Note how the initial profile, $F(x)$, is propagated along the two characteristics, $\xi = x + ct$ and $\eta = x - ct$, without any change. In contrast, the initial tendency, $\partial_t \psi(x, t = 0) = G(x)$, is smoothed through the time integration. This behavior contrasts to the heat equation in Section 3.6, with its single time derivative resulting in a smoothing of the full solution.

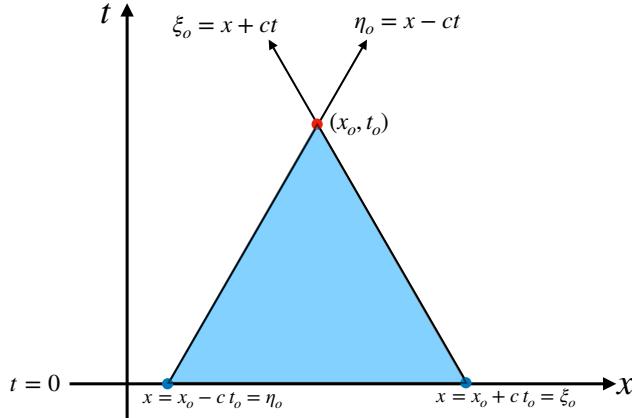


FIGURE 3.3: According to the wave solution (3.76) to the initial value problem on a line, an arbitrary space-time point, (x_o, t_o) , is causally connected via wave signals to all space-time points within the blue region. This *domain of influence* is bounded by the two wave characteristics, $\xi_o = x_o + ct_o$ and $\eta_o = x_o - ct_o$, with these characteristics the pathway for propagating information about the initial wave profile, $\psi(x, t = 0) = F(x)$. Points in between the characteristics are causally connected via the initial time tendency, $\partial_t \psi(x, t = 0) = G(x)$, which is integrated over the region $x_o - ct_o \leq x \leq x_o + ct_o$. Points outside the domain of influence are causally disconnected from the point (x_o, t_o) .

3.7.2 Domain of influence for wave signals

The wave solution (3.76) at a point in space time, (x_o, t_o) , depends on data to its past within a causality triangle, or *domain of influence*, as shown in Figure 3.3. The domain of influence is bounded by the two wave characteristics, $\xi_o = x_o + ct_o$ and $\eta_o = x_o - ct_o$. These characteristics are

the pathways for propagating information about the initial wave profile, $\psi(x, t = 0) = F(x)$. Points in between the characteristics are causally connected via the initial time tendency, $\partial_t \psi(x, t = 0) = G(x)$, which is integrated over the region $x_o - c t_o \leq x \leq x_o + c t_o$. Points outside the domain of influence are causally disconnected from the point (x_o, t_o) .

3.7.3 Helmholtz equation

Consider the wave equation with a constant wave speed

$$\frac{\partial^2 \psi}{\partial t^2} - c^2 \nabla^2 \psi = 0. \quad (3.77)$$

Assuming a wave ansatz³ of the form

$$\psi(\mathbf{x}, t) = e^{i\omega t} \Psi(\mathbf{x}) \quad (3.78)$$

results in the *Helmholtz equation* for the amplitude function

$$(\nabla^2 + k^2) \Psi = 0, \quad (3.79)$$

where $k = \omega/c$ is the wavenumber. The Helmholtz equation thus plays a central role in wave theory.

3.7.4 Duhamel's superposition integral for the wave equation

We here present Duhamel's superposition integral for the wave equation, following from the similar discussion for the heat equation in Section 3.6.3. For this purpose, consider the forced wave equation with time independent forcing

$$\frac{\partial^2 \Psi}{\partial t^2} = c^2 \nabla^2 \Psi + G(\mathbf{x}) \quad \mathbf{x} \in \mathbb{R}^n, t > 0 \quad (3.80a)$$

$$\Psi(\mathbf{x}, t) = 0 \quad \mathbf{x} \in \mathbb{R}^n, t = 0, \quad (3.80b)$$

and the corresponding unforced wave equation with inhomogenous initial time tendency

$$\frac{\partial^2 \psi}{\partial t^2} = c^2 \nabla^2 \psi \quad \mathbf{x} \in \mathbb{R}^n, t > 0 \quad (3.81a)$$

$$\psi(\mathbf{x}, t) = 0 \quad \mathbf{x} \in \mathbb{R}^n, t = 0. \quad (3.81b)$$

$$\frac{\partial \psi(\mathbf{x}, t)}{\partial t} = G(\mathbf{x}) \quad \mathbf{x} \in \mathbb{R}^n, t = 0. \quad (3.81c)$$

The two scalar fields are related by Duhamel's superposition integral

$$\Psi(\mathbf{x}, t) = \int_0^t \psi(\mathbf{x}, t - \tau) d\tau. \quad (3.82)$$

We can verify this formula through direct differentiation

$$\frac{\partial \Psi(\mathbf{x}, t)}{\partial t} = \int_0^t \frac{\partial \psi(\mathbf{x}, t - \tau)}{\partial t} d\tau \quad (3.83a)$$

$$\frac{\partial^2 \Psi(\mathbf{x}, t)}{\partial t^2} = \frac{\partial \psi(\mathbf{x}, 0)}{\partial t} + \int_0^t \frac{\partial^2 \psi(\mathbf{x}, t - \tau)}{\partial t^2} d\tau = G(\mathbf{x}) + c^2 \nabla^2 \Psi(\mathbf{x}, t). \quad (3.83b)$$

³Ansatze is a German word meaning an educated guess for the form of the solution.

As an example, consider the initial value problem for wave equation on a line with time independent forcing

$$\frac{\partial^2 \Psi}{\partial t^2} = c^2 \frac{\partial^2 \Psi}{\partial x^2} + G(x) \quad -\infty < x < \infty, t > 0 \quad (3.84a)$$

$$\Psi(x, t) = 0 \quad -\infty < x < \infty, t = 0 \quad (3.84b)$$

$$\frac{\partial \Psi(x, t)}{\partial t} = 0 \quad -\infty < x < \infty, t = 0. \quad (3.84c)$$

Duhamel's superposition integral says that Ψ is related to the solution of the unforced wave equation with initial time tendency given by the forcing

$$\frac{\partial^2 \psi}{\partial t^2} = c^2 \frac{\partial^2 \psi}{\partial x^2} \quad -\infty < x < \infty, t > 0 \quad (3.85a)$$

$$\psi(x, t) = 0 \quad -\infty < x < \infty, t = \tau \quad (3.85b)$$

$$\frac{\partial \psi(x, t)}{\partial t} = G(x) \quad -\infty < x < \infty, t = \tau > 0. \quad (3.85c)$$

We know from Section 3.7.1 that the solution ψ is given by the D'Alembert formula in equation (3.76), only here with the initial condition function set to zero

$$\psi(x, t) = \frac{1}{2c} \int_{x-ct}^{x+ct} G(s) ds. \quad (3.86)$$

Hence, D'Alembert's formula says that the solution to the forced wave equation (3.84a)-(3.84c) is given by the superposition integral

$$\Psi(x, t) = \frac{1}{2c} \int_0^t \int_{x-c(t-\tau)}^{x+c(t-\tau)} G(s) ds. \quad (3.87)$$

Introducing the *antiderivative* function via

$$\mathcal{G}(s) = \int^s G(s') ds' \iff \frac{\partial \mathcal{G}(s)}{\partial s} = G(s) \quad (3.88)$$

allows us to interpret the solution (3.87) as the superposition of two oppositely traveling waves

$$\Psi(x, t) = \frac{1}{2c} \int_0^t \left[\mathcal{G}[x + c(t - \tau)] - \mathcal{G}[x - c(t - \tau)] \right] d\tau. \quad (3.89)$$

3.7.5 Further study

[Stakgold \(2000a,b\)](#) provides a thorough discussion of the wave equation and the related Helmholtz equation.

3.8 Evolution of time averages

In geophysical fluid mechanics, we generically refer to an equation with a time derivative, such as a parabolic or hyperbolic equation, as a *prognostic equation* or an *evolution equation*. In the analysis of such equations, for example when analyzing simulation output or time series data, it is common

to take the time average in order to focus on lower frequency behavior. This section provides a technical discussion concerning this time averaging operation.

For this purpose, ignore all space coordinates and write a generic prognostic equation in the form

$$\frac{dA}{dt} = \mathcal{B}. \quad (3.90)$$

For example, the quantity A might be the velocity or temperature at a point in space, and \mathcal{B} might be the acceleration due to pressure or the heating due to temperature diffusion. We term dA/dt the *time tendency* of the quantity A whereas \mathcal{B} is the “forcing” that gives rise to the time tendency. In the analysis of fluid flows, we commonly wish to diagnose terms appearing in the evolution equations for the purpose of ascribing physical understanding to the flow regime; e.g., what forces are more active in certain regions. Although sitting a bit outside the scope of a chapter on PDEs, the material in this section exposes some common questions that arise when time averaging terms appearing in the prognostic equations of geophysical fluid mechanics.

Time integration of equation (3.90) leads to

$$A(t) = A(t_0) + \int_{t_0}^t \mathcal{B}(s) ds, \quad (3.91)$$

thus providing an expression for the instantaneous value of A at an arbitrary time t , assuming knowledge of the initial value, $A(t_0)$, as well as the time integral of \mathcal{B} . In practice, particularly when working with numerical models, we typically have access to time averages over some time interval (e.g., days, months, years, decades) rather than instantaneous (snapshot) values of A . Furthermore, instantaneous snapshots can be prone to relatively large fluctuations that expose the diagnostic calculations to numerical precision errors (e.g., small differences between relatively large fluctuating values). We are thus interested in relating time averages of A to time averages of \mathcal{B} .

3.8.1 Time averages

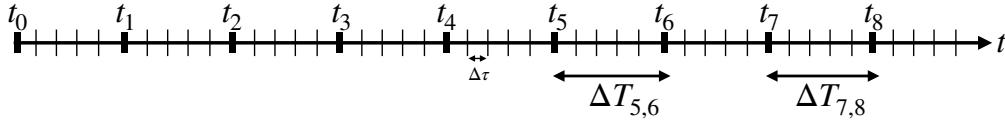


FIGURE 3.4: Example time axis for the discussion of time averaging. The labeled times, t_n , can represent, for example, days, months or years with the time interval, $\Delta T_{n,n+1} = t_{n+1} - t_n$, not necessarily the same (e.g., different number of days in a month or a leap year versus non-leap year). The smaller unlabeled time steps represent the time steps for the model’s prognostic equations (e.g., days, hours, seconds, etc.), with fixed time step $\Delta\tau$.

Introduce a discrete partitioning of the time axis as in Figure 3.4 and define an unweighted time average over a chosen time interval $\Delta T_{n,n+1} = t_{n+1} - t_n > 0$

$$\bar{A}_{n,n+1} = \frac{1}{\Delta T_{n,n+1}} \int_{t_n}^{t_{n+1}} A(t) dt \quad (3.92a)$$

$$\bar{\mathcal{B}}_{n,n+1} = \frac{1}{\Delta T_{n,n+1}} \int_{t_n}^{t_{n+1}} \mathcal{B}(t) dt. \quad (3.92b)$$

These integrals are realized in practice as a discrete sum over the model time steps, with only the lower limit inclusive so as to not double-count endpoints; i.e., $[t_n, t_{n+1})$. We allow for non-constant time intervals, $\Delta T_{n,n+1}$, as arises in monthly and yearly (with leap-years) time averages.

Substituting expression (3.91) into the time mean (3.92a) renders

$$\bar{A}_{n,n+1} - A(t_0) = \frac{1}{\Delta T_{n,n+1}} \int_{t_n}^{t_{n+1}} \left[\int_{t_0}^t \mathcal{B}(s) ds \right] dt, \quad (3.93)$$

and a similar expression over a later time interval $[t_p, t_{p+1}]$ with $p \geq n + 1$ leads to the difference between time averages

$$\bar{A}_{p,p+1} - \bar{A}_{n,n+1} = \frac{1}{\Delta T_{p,p+1}} \int_{t_p}^{t_{p+1}} \left[\int_{t_0}^t \mathcal{B}(s) ds \right] dt - \frac{1}{\Delta T_{n,n+1}} \int_{t_n}^{t_{n+1}} \left[\int_{t_0}^t \mathcal{B}(s) ds \right] dt. \quad (3.94)$$

The formalism allows us to take differences between time averages over intervals that are separated, such as might be of interest in taking decadal means between the beginning and end of a century, for example. Importantly, the initial value, $A(t_0)$, is absent from the difference in time means so that there are only time integrated quantities appearing in equation (3.94).

3.8.2 Massaging the double time integrals

The double time integrals in equation (3.94) can be massaged into a simpler form. We start by making the following decomposition and noting that $t_n \leq t \leq t_{n+1}$

$$\int_{t_n}^{t_{n+1}} \left[\int_{t_0}^t \mathcal{B}(s) ds \right] dt = \int_{t_n}^{t_{n+1}} \left[\int_{t_0}^{t_n} \mathcal{B}(s) ds + \int_{t_n}^t \mathcal{B}(s) ds \right] dt \quad (3.95a)$$

$$= \Delta T_{n,n+1} \int_{t_0}^{t_n} \mathcal{B}(s) ds + \int_{t_n}^{t_{n+1}} \left[\int_{t_n}^t \mathcal{B}(s) ds \right] dt. \quad (3.95b)$$

We are thus led to the difference

$$\begin{aligned} \bar{A}_{p,p+1} - \bar{A}_{n,n+1} &= \left[\int_{t_0}^{t_p} \mathcal{B}(s) ds - \int_{t_0}^{t_n} \mathcal{B}(s) ds \right] \\ &+ \frac{1}{\Delta T_{p,p+1}} \int_{t_p}^{t_{p+1}} \left[\int_{t_p}^t \mathcal{B}(s) ds \right] dt - \frac{1}{\Delta T_{n,n+1}} \int_{t_n}^{t_{n+1}} \left[\int_{t_n}^t \mathcal{B}(s) ds \right] dt \\ &= \int_{t_n}^{t_p} \mathcal{B}(t) dt + \frac{1}{\Delta T_{p,p+1}} \int_{t_p}^{t_{p+1}} \left[\int_{t_p}^t \mathcal{B}(s) ds \right] dt - \frac{1}{\Delta T_{n,n+1}} \int_{t_n}^{t_{n+1}} \left[\int_{t_n}^t \mathcal{B}(s) ds \right] dt. \end{aligned} \quad (3.96)$$

The double integrals in equation (3.96) take place over triangular time domains, such as shown in Figure 3.5.

3.8.3 Making use of a double integral identity

We here derive an identity that reduces the double integral in equation (3.96) to a single integral to expose the underlying geometry of the time windowing. For this purpose we make use of the following identity

$$\int_{t_n}^{t_{n+1}} \left[\int_{t_n}^t \mathcal{B}(s) ds \right] dt = \int_{t_n}^{t_{n+1}} (t_{n+1} - t) \mathcal{B}(t) dt. \quad (3.97)$$

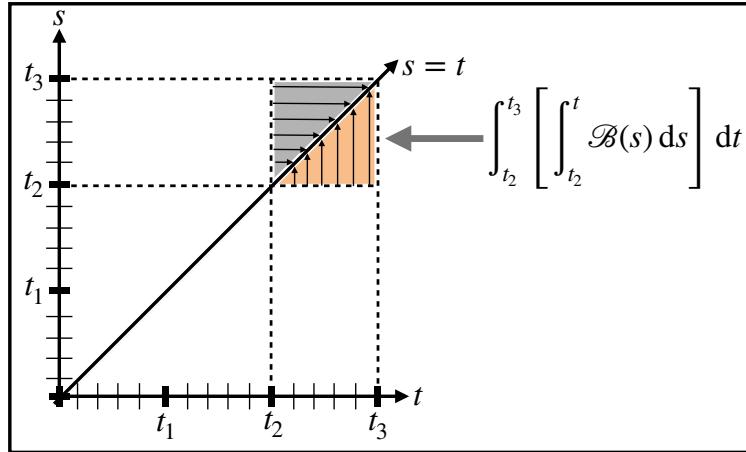


FIGURE 3.5: Gold region depicts the time integration domain used in one of the double integrals from equation (3.96) for the special case of $n = 2$. Note that the gray triangular region generally leads to a distinct integral.

To prove this identity we make the substitution $\mathcal{B}(s) = dA/ds$ from equation (3.90) and then show that both sides to equation (3.97) yield the same result. For the left hand side we have

$$\int_{t_n}^{t_{n+1}} \left[\int_{t_n}^t \mathcal{B}(s) ds \right] dt = \int_{t_n}^{t_{n+1}} \left[\int_{t_n}^t \frac{dA(s)}{ds} ds \right] dt \quad (3.98a)$$

$$= \int_{t_n}^{t_{n+1}} \left[\int_{t_n}^t dA(s) \right] dt \quad (3.98b)$$

$$= \int_{t_n}^{t_{n+1}} [A(t) - A(t_n)] dt \quad (3.98c)$$

$$= \int_{t_n}^{t_{n+1}} A(t) dt - (t_{n+1} - t_n) A(t_n), \quad (3.98d)$$

whereas the right hand side is

$$\int_{t_n}^{t_{n+1}} (t_{n+1} - t) \mathcal{B}(t) dt = \int_{t_n}^{t_{n+1}} (t_{n+1} - t) \frac{dA(t)}{dt} dt \quad (3.99a)$$

$$= \int_{t_n}^{t_{n+1}} (t_{n+1} - t) dA(t) \quad (3.99b)$$

$$= \int_{t_n}^{t_{n+1}} d[A(t) (t_{n+1} - t)] + A(t) dt \quad (3.99c)$$

$$= -A(t_n) (t_{n+1} - t_n) + \int_{t_n}^{t_{n+1}} A(t) dt, \quad (3.99d)$$

which is identical to the left hand side given by equation (3.98d). We have thus proven the double integral formula (3.97).

The right hand side of the double integral formula (3.97) can be written

$$\int_{t_n}^{t_{n+1}} (t_{n+1} - t) \mathcal{B}(t) dt = t_{n+1} \Delta T_{n,n+1} \bar{\mathcal{B}}_{n,n+1} - \int_{t_n}^{t_{n+1}} t \mathcal{B}(t) dt, \quad (3.100)$$

which might be useful in some contexts. However, it is awkward for our purposes since it exposes the absolute time, t_{n+1} , in the first term on the right hand side and the time, t , within the integral.

Since we generally do not hold any initial time as special (i.e., the initial time, t_0 , is arbitrary), it is preferable to retain the time differences throughout the formulation. Hence, when making use of the double integral identity (3.97) we bring equation (3.96) into the form

$$\bar{A}_{p,p+1} - \bar{A}_{n,n+1} = \int_{t_n}^{t_p} \mathcal{B}(t) dt + \int_{t_p}^{t_{p+1}} \frac{(t_{p+1} - t) \mathcal{B}(t)}{t_{p+1} - t_p} dt - \int_{t_n}^{t_{n+1}} \frac{(t_{n+1} - t) \mathcal{B}(t)}{t_{n+1} - t_n} dt \quad (3.101a)$$

$$= \int_{t_n}^{t_{n+1}} \frac{(t - t_n) \mathcal{B}(t)}{t_{n+1} - t_n} dt + \int_{t_{n+1}}^{t_p} \mathcal{B}(t) dt + \int_{t_p}^{t_{p+1}} \frac{(t_{p+1} - t) \mathcal{B}(t)}{t_{p+1} - t_p} dt. \quad (3.101b)$$

The first right hand side term is a weighted integral with a linearly increasing weight from zero to unity, whereas the final right hand side term has a linearly decreasing weight from unity to zero. The middle term has a unity weight throughout and it vanishes if $p = n + 1$, as when the averaging regions are adjacent. Figure 3.6. illustrates the time windowing used for equation (3.101b).

As a final means to write equation (3.101b), extend the middle term to the end of the time period and then subtract the extra piece and recombine to render

$$\bar{A}_{p,p+1} - \bar{A}_{n,n+1} = \int_{t_{n+1}}^{t_{p+1}} \mathcal{B}(t) dt + \int_{t_n}^{t_{n+1}} \frac{(t - t_n) \mathcal{B}(t)}{t_{n+1} - t_n} dt - \int_{t_p}^{t_{p+1}} \frac{(t - t_p) \mathcal{B}(t)}{t_{p+1} - t_p} dt. \quad (3.102)$$

The first term on the right hand side is an unweighted integral from the end of the first interval to the end of the final interval, whereas the other two terms both have increasing weights over their respective integration intervals. This form allows for some advantages diagnostically since we only need to save unweighted integrals plus linearly increasing weighted integrals; there is no need to save decreasing weighted integrals.

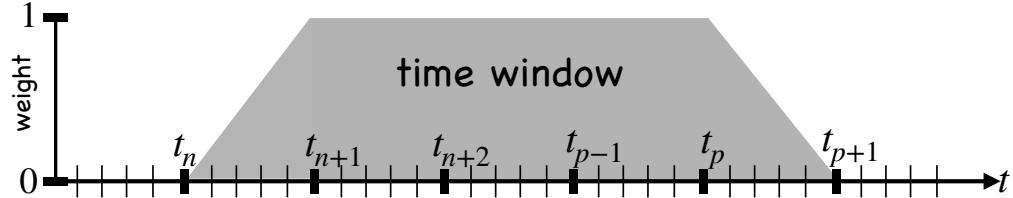


FIGURE 3.6: Illustrating the time window weighting used in computing the right hand side of equation (3.101b). Note that if $p = n + 1$ then there is no plateau region of unit weight; instead there are two adjacent triangular regions.



Green's function methods

There are many methods to solve linear partial differential equations (PDEs). Some methods, such as the separation of variables, Fourier transforms, and Laplace transforms, are suited to idealized domains with relatively simple geometries that enable closed form analytic solutions. However, for many physical applications we encounter domains and boundary conditions where analytical methods are unavailable. It is thus often useful to have a general closed-form solution to study properties of the solution either analytically or numerically. The Green's function method provides this solution.

In this chapter we develop the Green's function method for the elliptic and parabolic linear partial differential equations of mathematical physics introduced in Chapter 3. We are most interested in the Green's function method as a conceptual framework and as such we are relatively unconcerned with how to determine the Green's function for any particular problem. Even without an explicit form of the Green's function, an appreciation of Green's function methods deepens both physical and mathematical understanding of the various linear PDEs encountered in this book, such as the Poisson equation in Newtonian gravity (Section 10.12), and quasi-geostrophic theory (Chapter 42), as well as the diffusion equation and the advection-diffusion equation of Chapter 48.

READER'S GUIDE TO THIS CHAPTER

We encounter a variety of Green's functions that differ according to the differential operator and boundary conditions. Use of the common symbol, G , minimizes the adornments otherwise needed to distinguish Green's functions. Confusion is avoided by noting that properties of any particular Green's function are specific to the section where the function is discussed. Much in this chapter is based on the treatments found in Chapter 7 of *Morse and Feshbach* (1953) as well as *Stakgold* (2000a,b). Both of these books are classics whose presentations have stood the test of time.

4.1	Conceptual foundations	68
4.2	Dirac delta	69
4.2.1	A point mass source	70
4.2.2	Sifting property	70
4.2.3	Cartesian and spherical coordinates	71
4.2.4	Example $\delta^{(\epsilon)}(x)$ functions	71
4.2.5	Connection to the Heaviside step function	72
4.2.6	Sifting property for the Dirac dipole	73
4.2.7	Temporal Dirac delta and impulses	74
4.2.8	Shifting the space-time position of the source	74

4.3	Laplace's free space Green's function	75
4.3.1	Gravitational potential from a point mass source	75
4.3.2	Free space Green's function for Laplace's equation	75
4.3.3	Details of the three-dimensional free space Green's function	76
4.4	One-dimensional Poisson's equation	77
4.4.1	Integral expression for ψ in terms of G	77
4.4.2	Preliminary steps for obtaining the Green's function	78
4.4.3	Green's function satisfying Dirichlet boundary conditions	78
4.4.4	Green's function satisfying Neumann boundary conditions	79
4.4.5	Uniqueness of the solution to the Dirichlet problem	80
4.5	Poisson's equation with Dirichlet boundaries	80
4.5.1	Constraints on the source and boundary normal derivative	81
4.5.2	Uniqueness of the solution	81
4.5.3	The Green's function problem	82
4.5.4	Reciprocity of the Green's function	83
4.5.5	The integral solution	84
4.5.6	Properties of the solution	85
4.5.7	Boundary Green's function	87
4.6	Poisson's equation with Neumann boundaries	88
4.6.1	Constraints on the source and boundary data	88
4.6.2	Uniqueness of the solution up to a constant	89
4.6.3	The Green's function problem	89
4.6.4	Properties of the solution	89
4.6.5	Verifying the solution	91
4.7	The diffusion equation	92
4.7.1	Causal free space Green's function	92
4.7.2	Causal Green's function	92
4.7.3	Adjoint causal Green's function	93
4.7.4	Reciprocity of the Green's function and its adjoint	94
4.7.5	Integral solution	96
4.7.6	Properties of the solution	99
4.7.7	Boundary propagator	100
4.8	Initial value problems and response functions	102
4.8.1	Impulse response function	103
4.8.2	Step response function	104
4.8.3	Reciprocity relation	105
4.8.4	Response function for general forcing	106
4.8.5	Connection to the boundary propagator	107
4.8.6	Comments and further study	107

4.1 Conceptual foundations

Our treatment of the Dirac delta and Green's functions is physically formal, meaning that our priority is to be satisfying from a physical perspective. This physically formal treatment is generally not mathematically rigorous, with a rigorous treatment requiring an array of mathematical apparatus outside our scope.¹ The physically formal treatment found here is generally supported by heuristic arguments taken from Newtonian gravity (see Section 10.12); electrostatics (e.g., [Jackson](#)

¹The term *physically formal* is often used in the mathematical physics literature as a complement to *mathematically rigorous*.

(1975)); and the diffusion of temperature or matter within a continuous media.² The tracer and temperature examples are directly relevant to geophysical fluid mechanics and will be developed in this book. We also make use of the electrostatic and gravity examples since they only require elementary physics and historically they provided motivation for development of Green's function methods in the 19th and 20th centuries.

The basic mathematical problem of classical continuum field theory concerns an arbitrary field, $\psi(\mathbf{x}, t)$, that represents a physical property such as material tracer concentration, velocity stream-function, temperature in a continuous media, Newtonian gravitational potential, or classical electrostatic potential. The space-time structure of the field is the result of a space-time distributed source (e.g., tracer source, potential vorticity distribution, mass distribution, or electric charge distribution) along with differential operators (e.g., time derivative, Laplacian, diffusion operator, wave operator) that connect the field across points in space-time as well as initial and boundary conditions that fix the behavior of ψ in selected space-time regions.³

For our studies, the differential equations describing ψ are generally nonlinear. However, in some important cases it is possible to linearize the governing equations and learn much about the underlying mathematical and physical properties of the system. Furthermore, the advection-diffusion equation describing a passive tracer (Chapter 48) is linear even when the dynamical flow field is nonlinear. The conceptual foundation of the Green's function method is based on observing that if the field equations are linear, then ψ can be constructed by accumulating contributions from point-sized portions of the distributed source, including sources on the region boundaries. Let us write $G(\mathbf{x}, t | \mathbf{x}_0, t_0)$ for the field at an observation space-time point (\mathbf{x}, t) (the *field point*) caused by a point source at (\mathbf{x}_0, t_0) (the *source point*). We then observe that $\psi(\mathbf{x}, t)$ caused by a source that is distributed through space and time is the integral of G with the source computed over the space-time domain occupied by the source. Similar superpositions are made for the initial and boundary conditions. The function G is referred to as the *Green's function* in honor of the 19th century mathematician who first developed these functions for studies of electrostatics.

The above arguments rely on the superposition property of linear systems. It does so by finding a particular solution (the Green's function) to a linear initial-boundary value problem with a singular point source (Dirac delta) and homogeneous boundary conditions. The solution to the original problem is found by integrating (convolving) the Green's function with the boundary conditions, initial conditions, and distributed sources. The Green's function is generally simpler to determine than the solution to the original initial-boundary value problem. Furthermore, the Green's function provides a formal inverse to the linear partial differential operator in a manner reminiscent of matrix inversion used to solve a matrix-vector problem. Just as for the matrix-vector problem, once we have the Green's function we can write the solution to any of the associated initial-boundary value problems regardless the details of the distributed source, initial data, or boundary data. Herein lies the power of the Green's function method and why it has found much use across mathematical physics.

4.2 Dirac delta

The *Dirac delta* provides an idealization of a point source and it serves a core role in the study of Green's functions. In mathematics, the Dirac delta is known as a *generalized function* (e.g., chapter

²A heuristic technique or argument employs a practical method not guaranteed to be fully rational or deductive from all perspectives, but is sufficient for establishing a self-consistent formalism. The study of the Dirac delta and Green's functions are examples where physical heuristics established a formalism whose mathematical rigor followed later.

³In some cases, sources are present only along spatial boundaries.

5 of [Stakgold \(2000b\)](#)). We use the terse nomenclature *Dirac delta*, rather than the commonly used “delta function”, to emphasize its special nature as a generalized function.

4.2.1 A point mass source

In Newtonian gravity we encounter the Poisson equation for the gravitational potential, Φ , arising from an arbitrary mass density, ρ

$$\nabla^2 \Phi = 4\pi G^{\text{grv}} \rho, \quad (4.1)$$

with G^{grv} Newton’s gravitational constant (see Section 10.12.1).⁴ The gravitational potential for an arbitrary spherically symmetric mass, when sampled at a point outside the mass, equals to the potential of a point mass located at the origin. Making precise the notion of a “point mass” provides a venue to introduce the Dirac delta.

For that purpose, consider a mass, M , distributed uniformly within a sphere of radius ϵ and volume

$$V_\epsilon = \frac{4}{3}\pi\epsilon^3, \quad (4.2)$$

and let the sphere be centered at the origin of a coordinate system. The mass distribution thus has a density

$$\rho(\mathbf{x}) = M \delta^{(\epsilon)}(\mathbf{x}), \quad (4.3)$$

where we introduced the ϵ -distribution

$$\delta^{(\epsilon)}(\mathbf{x}) \equiv \begin{cases} V_\epsilon^{-1} & \text{if } |\mathbf{x}| \leq \epsilon \\ 0 & \text{if } |\mathbf{x}| > \epsilon. \end{cases} \quad (4.4)$$

By construction, an integral over a domain fully encompassing the sphere yields the mass

$$\int_{\mathcal{R}} \rho dV = M \int_{\mathcal{R}} \delta^{(\epsilon)}(\mathbf{x}) dV = M, \quad (4.5)$$

with this result holding even as the radius of the sphere becomes arbitrarily small, $\epsilon \rightarrow 0$. We define the Dirac delta as the limiting ϵ -distribution

$$\delta(\mathbf{x}) \equiv \lim_{\epsilon \rightarrow 0} \delta^{(\epsilon)}(\mathbf{x}). \quad (4.6)$$

By construction, the Dirac delta arises from taking the limit of a mass source that is zero everywhere in space except at a single point, at which point the Dirac delta is formally infinite. Concerns with how to interpret the infinite value of $\delta(\mathbf{x})$ at the origin are ameliorated by recognizing that $\delta(\mathbf{x})$ is evaluated only within an integral, with these integral properties well defined and described in the remainder of this section. Connecting to the other physical analogs, the Dirac delta corresponds to a point charge in electrostatics or to a point source of trace matter within a fluid.

4.2.2 Sifting property

Now consider multiplying an ϵ -distribution by an arbitrary smooth function, $\psi(\mathbf{x})$. Since the ϵ -distribution has support only within the ϵ -sphere surrounding the origin, an integral of $\delta^{(\epsilon)}(\mathbf{x}) \psi(\mathbf{x})$ over the sphere, in the limit that $\epsilon \rightarrow 0$, leads to the *sifting property*

$$\lim_{\epsilon \rightarrow 0} \int_{\mathcal{R}} \delta^{(\epsilon)}(\mathbf{x}) \psi(\mathbf{x}) dV = \int_{\mathcal{R}} \delta(\mathbf{x}) \psi(\mathbf{x}) dV = \psi(\mathbf{x} = 0). \quad (4.7)$$

Hence, the Dirac delta acts to “sift” out the value of the function at the location of the Dirac delta source, and it is the key defining feature of the Dirac delta for our purposes.

⁴Elsewhere in this book we write $G = G^{\text{grv}}$. We add the “grv” label in this chapter and reserve G for the Green’s function.

4.2.3 Cartesian and spherical coordinates

The definition (4.6) was constructed in three space dimensions, with the Dirac delta having dimensions of inverse volume. We can consider the same procedure for a single space dimension, with the one-dimensional Dirac delta having dimension of inverse length and thus satisfying the normalization condition

$$\int_{\mathcal{R}} \delta(x) dx = 1, \quad (4.8)$$

where the integration domain, \mathcal{R} , includes the origin. Correspondingly, we can decompose the three-dimensional Dirac delta according to

$$\delta(\mathbf{x}) = \delta(x) \delta(y) \delta(z), \quad (4.9)$$

so that, with the domain \mathcal{R} containing the origin, we have

$$\int_{\mathcal{R}} \delta(\mathbf{x}) dV = \int_{\mathcal{R}} \delta(x) \delta(y) \delta(z) dx dy dz = 1. \quad (4.10)$$

Rather than Cartesian coordinates, we sometimes find it useful to make use of the spherical coordinates⁵ to render

$$\delta(\mathbf{x}) = \frac{1}{r^2 \cos \phi} \delta(r) \delta(\phi) \delta(\lambda), \quad (4.11)$$

so that

$$\int_{\mathcal{R}} \delta(\mathbf{x}) dx dy dz = \int_{\mathcal{R}} \delta(\mathbf{x}) r^2 \cos \phi dr d\phi d\lambda = \int_{\mathcal{R}} \delta(r) \delta(\phi) \delta(\lambda) dr d\phi d\lambda = 1. \quad (4.12)$$

Notice how the dimensions of a particular Dirac delta equals to the inverse dimensions of its argument, so that both $\delta(\phi)$ and $\delta(\lambda)$ are dimensionless whereas $\delta(r)$ has dimensions of inverse length. Similar treatments hold for other coordinates.

4.2.4 Example $\delta^{(\epsilon)}(x)$ functions

The particular construction (4.6) is not unique. That is, there are many other suitable ϵ -distributions whose limiting behavior also result in a Dirac delta as defined by the unit normalization and sifting properties. We here list a few that appear in applications

$$\delta^{(\epsilon)}(x) = \epsilon^{-1} \quad \text{for } |x| < \epsilon/2 \text{ and } 0 \text{ for } |x| > \epsilon/2 \quad (4.13a)$$

$$\delta^{(\epsilon)}(x) = \frac{e^{-|x|/\epsilon}}{2\epsilon} \quad (4.13b)$$

$$\delta^{(\epsilon)}(x) = \frac{\epsilon}{\pi(x^2 + \epsilon^2)} \quad (4.13c)$$

$$\delta^{(\epsilon)}(x) = \frac{e^{-x^2/\epsilon^2}}{\epsilon \pi} \quad (4.13d)$$

$$\delta^{(\epsilon)}(x) = \frac{\sin(x/\epsilon)}{x \epsilon} \quad (4.13e)$$

$$\delta^{(\epsilon)}(x) = \frac{\epsilon \sin^2(x/\epsilon)}{\pi x^2}. \quad (4.13f)$$

⁵We discuss spherical coordinates in Section 8.2.

With x corresponding to a spatial position, note how each of these one-dimensional functions has dimensions of inverse length. Figure 4.1 depicts the rectangular function (4.13a) as a canonical example.

Consider the multiplication of each of the above functions $\delta^{(\epsilon)}(x)$, by a non-dimensional, continuous and bounded function, $F(x)$, that is unity at the origin, $F(0) = 1$. In this case, $F(x) \delta^{(\epsilon)}(x)$ still satisfies the normalization and sifting properties required of a Dirac delta

$$\lim_{\epsilon \rightarrow 0} \int_{\mathcal{R}} F(x) \delta^{(\epsilon)}(x) dx = F(0) = 1 \quad \text{and} \quad \lim_{\epsilon \rightarrow 0} \int_{\mathcal{R}} \psi(x) F(x) \delta^{(\epsilon)}(x) dx = \psi(0) F(0) = \psi(0), \quad (4.14)$$

so that

$$\lim_{\epsilon \rightarrow 0} F(x) \delta^{(\epsilon)}(x) = \delta(x). \quad (4.15)$$

Correspondingly, whenever working with a Dirac delta, if it is multiplied by a non-dimensional function $F(x)$ with $F(0) = 1$, then we can disregard this function since it does not modify the Dirac source.

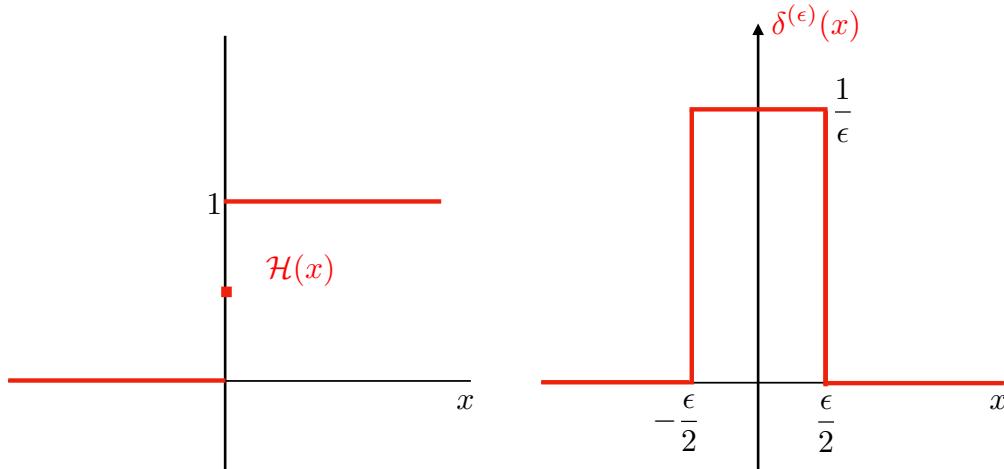


FIGURE 4.1: Left panel: Heaviside step function, $\mathcal{H}(x)$, as given by equation (4.16). Right panel: the square pulse function $\delta^{(\epsilon)}(x)$ given by equation (4.13a), with $\lim_{\epsilon \rightarrow 0} \delta^{(\epsilon)}(x) = \delta(x) = d\mathcal{H}(x)/dx$.

4.2.5 Connection to the Heaviside step function

The Heaviside step function (Figure 4.1) is given by⁶

$$\mathcal{H}(x) = \begin{cases} 0 & \text{if } x < 0 \\ 1/2 & \text{if } x = 0 \\ 1 & \text{if } x > 0, \end{cases} \quad (4.16)$$

and it is related to the sgn function

$$\operatorname{sgn}(x) = \begin{cases} -1 & \text{if } x < 0 \\ 0 & \text{if } x = 0 \\ 1 & \text{if } x > 0, \end{cases} \quad (4.17)$$

⁶In some treatments, $\mathcal{H}(x)$ is undefined at $x = 0$. For our purposes, the properties of the Heaviside step function remain unchanged whether it is defined at $x = 0$ or not. See footnote on page 20 of [Stakgold \(2000a\)](#) for more details.

according to

$$\text{sgn}(x) = 2\mathcal{H}(x) - 1. \quad (4.18)$$

Both the Heaviside and sgn functions are piecewise continuous and have infinite derivatives at $x = 0$. In particular, the derivative of the Heaviside step function equals to the Dirac delta

$$\frac{d\mathcal{H}(x)}{dx} = \delta(x). \quad (4.19)$$

This identity is most apparent by considering the square pulse approximation to the Dirac delta given by equation (4.13a), which can be written

$$\delta(x) = \lim_{\epsilon \rightarrow 0} \delta^{(\epsilon)}(x) = \lim_{\epsilon \rightarrow 0} \frac{\mathcal{H}(x + \epsilon/2) - \mathcal{H}(x - \epsilon/2)}{\epsilon} = \frac{d\mathcal{H}(x)}{dx}. \quad (4.20)$$

Equivalently, we see that the Heaviside step function is the cumulative distribution of the Dirac delta

$$\int_{-\infty}^x \delta(x') dx' = \int_{-\infty}^x \frac{d\mathcal{H}(x')}{dx'} dx' = \mathcal{H}(x) - \mathcal{H}(-\infty) = \mathcal{H}(x). \quad (4.21)$$

Connecting to the language of probability theory, the Dirac delta corresponds to a probability density function peaked over an infinitesimal region, whereas the Heaviside step function is the corresponding probability distribution function.

4.2.6 Sifting property for the Dirac dipole

The derivative of a Dirac delta represents an idealization of a dipole (e.g., see exercise 1.14 of [Stakgold \(2000a\)](#)). There are occasions when it is useful to know how a dipole acts on a function inside of an integral. For this purpose we make use of the identity

$$\int_{-\epsilon}^{\epsilon} \frac{d[\psi(x) \delta(x)]}{dx} dx = \psi(\epsilon) \delta(\epsilon) - \psi(-\epsilon) \delta(-\epsilon). \quad (4.22)$$

Each term on the right hand side vanishes when $\epsilon > 0$ since the Dirac delta never fires. Applying the product rule employed inside the integral renders the identity

$$\int_{-\epsilon}^{\epsilon} \frac{d[\psi(x) \delta(x)]}{dx} dx = \int_{-\epsilon}^{\epsilon} \left[\psi(x) \frac{d\delta(x)}{dx} + \delta(x) \frac{d\psi(x)}{dx} \right] dx = 0, \quad (4.23)$$

so that

$$\int_{-\epsilon}^{\epsilon} \psi(x) \frac{d\delta(x)}{dx} dx = - \left[\frac{d\psi(x)}{dx} \right]_{x=0}. \quad (4.24)$$

For example, let $\psi(x) = x$, in which case

$$\int_{-\epsilon}^{\epsilon} x \frac{d\delta(x)}{dx} dx = -1, \quad (4.25)$$

which can be formally written

$$x (d\delta(x)/dx) = -\delta(x). \quad (4.26)$$

4.2.7 Temporal Dirac delta and impulses

Poisson's equation for the gravitational potential is an elliptic partial differential equation (see Section 3.5), in which there is no time derivative. We now introduce the temporal Dirac delta to build towards our study of evolution equations, such as the diffusion equation of Section 4.7 and the advection-diffusion equation of Section 48.10. The temporal Dirac delta is a point source that is turned on just at one time instance and it is normalized according to

$$\int_{\mathcal{T}} \delta(t) dt = 1, \quad (4.27)$$

where \mathcal{T} is a time interval containing the source time, $t = 0$. This normalization means that $\delta(t)$ has dimensions of inverse time. The temporal Dirac delta also possesses the sifting property from Section 4.2.2, in which

$$\int_{\mathcal{T}} \delta(t) \psi(t) dt = \psi(t = 0). \quad (4.28)$$

In the study of transient behavior of dynamical systems, it is often of interest to examine the response of the system to an idealized force, $\mathcal{F}(t)$, where the force occurs over a small time increment. The time integral of this force is referred to as the *impulse*

$$I(\tau) = \int_{-\tau}^{\tau} \mathcal{F}(t) dt. \quad (4.29)$$

If the force is further idealized to occur just at a single moment in time, and it is normalized to unity, then we have the *unit impulse*, which is just the integral of the Dirac delta

$$I(\tau) = 1 = \int_{-\tau}^{\tau} \delta(t) dt. \quad (4.30)$$

The corresponding response of the dynamical system is referred to as the *impulse response function*. If the dynamical system is linear, then the impulse response function equals to the Green's function for the initial value problem. We further discuss the response function in Section 4.8.

4.2.8 Shifting the space-time position of the source

Thus far we have assumed the point source is located at the origin in space and time. But there is nothing special about that point. Correspondingly, we can arbitrarily place the source at (\mathbf{x}_0, t_0) , in which case the Dirac delta is written

$$\delta(\mathbf{x} - \mathbf{x}_0) \delta(t - t_0) = \delta(x - x_0) \delta(y - y_0) \delta(z - z_0) \delta(t - t_0). \quad (4.31)$$

Defining the region, \mathcal{R} , to now encompass the source point in space, $\mathbf{x} = \mathbf{x}_0$, and the time increment \mathcal{T} to encompasses the source time, $t = t_0$, we have the normalization condition

$$\int_{\mathcal{R}} \int_{\mathcal{T}} \delta(\mathbf{x} - \mathbf{x}_0) \delta(t - t_0) dV dt = 1, \quad (4.32)$$

as well as the sifting property

$$\int_{\mathcal{R}} \int_{\mathcal{T}} \psi(\mathbf{x}, t) \delta(\mathbf{x} - \mathbf{x}_0) \delta(t - t_0) dV dt = \psi(\mathbf{x}_0, t_0). \quad (4.33)$$

4.3 Laplace's free space Green's function

We introduced the Dirac delta by considering the Newtonian gravitational potential in the presence of a point mass source. This physical example also serves to introduce the notion of a Green's function, in this case a particularly simple Green's function known as a *free space Green's function* for Laplace's equation. The free space Green's function serves an important role in the analytical theory of Green's functions, and it offers us a pedagogical introduction to Green's functions.

4.3.1 Gravitational potential from a point mass source

Consider the gravitational potential, Φ , in the presence of a point mass source at $\mathbf{x} = \mathbf{x}_0$ in the absence of boundaries (i.e., in *free space*), and with an assumed decay of the potential when moving away from the source. In the study of Green's functions this potential is referred to as the *fundamental solution* to Laplace's equation. In other contexts it is referred to as the *free space Green's function*, which is the terminology we choose, where “free space” refers to the absence of any spatial boundaries. From the discussion in Section 4.2.1, the free space Green's function for Newtonian gravity satisfies the Poisson equation with a point mass source

$$\nabla_{\mathbf{x}}^2 \Phi(\mathbf{x}|\mathbf{x}_0) = 4\pi G^{\text{grav}} M \delta(\mathbf{x} - \mathbf{x}_0) \quad \text{and} \quad \Phi(\mathbf{x}|\mathbf{x}_0) \rightarrow 0 \text{ as } |\mathbf{x}| \rightarrow \infty, \quad (4.34)$$

where we wrote the point mass density in terms of the Dirac delta

$$\rho(\mathbf{x}|\mathbf{x}_0) = M \delta(\mathbf{x} - \mathbf{x}_0). \quad (4.35)$$

The first argument to the potential, \mathbf{x} , is the point where the field is sampled and is referred to as the *field point* or sometimes the *observation point*. The second argument, \mathbf{x}_0 , is where the source is located and is thus referred to as the *source point*. Figure 4.2 summarizes the notation. The Laplacian operator acts on the field point and is written as $\nabla_{\mathbf{x}}^2$ for clarity. The dimensional multiplier, $4\pi G^{\text{grav}} M$, acting on the Dirac delta in equation (4.34) is specific to the physical problem, here being for Newtonian gravity. In this case the Green's function has dimensions of $L^2 T^{-2}$ since it is a gravitational potential. In other cases, the Green's function will have distinct dimensions, whereas the Dirac delta remains with the same dimensions.⁷

The gravitational potential for a point mass source, in the absence of any boundaries, is the free space Green's function for Newtonian gravity. From Section 10.12.2 we can write this free space Green's function in the familiar form

$$\Phi(\mathbf{x}|\mathbf{x}_0) = -\frac{M G^{\text{grav}}}{|\mathbf{x} - \mathbf{x}_0|}. \quad (4.36)$$

This function is singular when sampling the field at the source location, $\mathbf{x} = \mathbf{x}_0$, and it decreases according to the inverse distance when moving away from the source.

4.3.2 Free space Green's function for Laplace's equation

Abstracting the previous discussion motivates us to define the free space Green's function, $\mathcal{G}(\mathbf{x}|\mathbf{x}_0)$, for the Laplace operator as the solution to the singular Poisson equation

$$-\nabla_{\mathbf{x}}^2 \mathcal{G}(\mathbf{x}|\mathbf{x}_0) = \delta(\mathbf{x} - \mathbf{x}_0) \quad \text{and} \quad \mathcal{G}(\mathbf{x}|\mathbf{x}_0) \rightarrow 0 \text{ as } |\mathbf{x}| \rightarrow \infty. \quad (4.37)$$

⁷As emphasized throughout this book, checking for physical dimensional consistency offers a powerful means to ensure that a mathematical expression makes physical sense. If the equation is not dimensionally consistent, then something is wrong either with the maths or the physics.

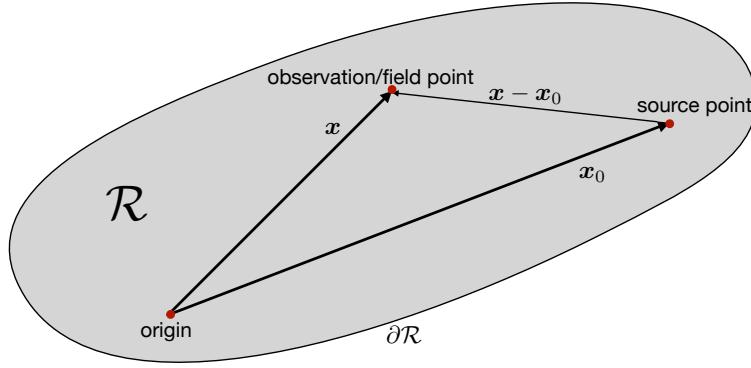


FIGURE 4.2: Depicting the geometry of a typical Green's function problem in space, shown here for a finite domain whereas the free-space Green's function has no boundaries. The field or observation point, \mathbf{x} , is where the Green's function is sampled, whereas the source point, \mathbf{x}_0 , is where the Dirac delta source is located. The origin is at an arbitrary position within the domain, \mathcal{R} , whose boundary is written $\partial\mathcal{R}$.

It is conventional in many treatments to place a minus sign on the Laplacian operator to correspond to how it appears in the diffusion equation of Section 4.7. In one, two and three space dimensions the free space Green's function is given by

$$\mathcal{G}(x|x_0) = -|x - x_0|/2 \quad \text{for } x, x_0 \in \mathbb{R}^1 \quad (4.38a)$$

$$\mathcal{G}(\mathbf{x}|\mathbf{x}_0) = -(2\pi)^{-1} \ln |\mathbf{x} - \mathbf{x}_0| \quad \text{for } \mathbf{x}, \mathbf{x}_0 \in \mathbb{R}^2 \quad (4.38b)$$

$$\mathcal{G}(\mathbf{x}|\mathbf{x}_0) = (4\pi |\mathbf{x} - \mathbf{x}_0|)^{-1} \quad \text{for } \mathbf{x}, \mathbf{x}_0 \in \mathbb{R}^3. \quad (4.38c)$$

We study the three-dimensional Green's function in Section 4.3.3 and encounter the one-dimensional Green's function in Section 4.4.

4.3.3 Details of the three-dimensional free space Green's function

We here verify that the expression (4.38c) is indeed the free space Green's function for \mathbb{R}^3 . This exercise provides some exposure to some of the formal manipulations encountered with Dirac deltas and Green's functions.

To simplify notation place the Dirac delta at $\mathbf{x}_0 = 0$ so that

$$\mathcal{G}(\mathbf{x}|0) = \frac{1}{4\pi r}, \quad (4.39)$$

where $|\mathbf{x}| = r$ is the radial distance from the origin. Introduce the continuous and non-singular function

$$G^{(\epsilon)}(\mathbf{x}) = \frac{1}{4\pi} \begin{cases} r^{-1} & \text{for } r > \epsilon \\ \epsilon^{-1} & \text{for } r \leq \epsilon, \end{cases} \quad (4.40)$$

with $\epsilon > 0$, in which case $G^{(\epsilon)}(\mathbf{x})$ has removed the singularity at $r = 0$. Now consider the integral

$$\mathcal{I}(\epsilon) = \int_{r \leq \epsilon} \psi(\mathbf{x}) \nabla^2 G^{(\epsilon)}(\mathbf{x}) dV, \quad (4.41)$$

for an arbitrary smooth function ψ and for a spherical region of radius $r = \epsilon$ centered on the origin. Since ψ is a smooth function, taking the limit as $\epsilon \rightarrow 0$ allows us to remove ψ from the integral

$$\lim_{\epsilon \rightarrow 0} \mathcal{I}(\epsilon) = \psi(r = 0) \lim_{\epsilon \rightarrow 0} \left[\int_{r \leq \epsilon} \nabla^2 G^{(\epsilon)}(\mathbf{x}) dV \right]. \quad (4.42)$$

Making use of the divergence theorem brings the volume integral to a surface integral over the ϵ -sphere

$$\lim_{\epsilon \rightarrow 0} \mathcal{I}(\epsilon) = \psi(r=0) \lim_{\epsilon \rightarrow 0} \left[4\pi \int_{r=\epsilon} \hat{\mathbf{r}} \cdot \nabla(1/r) r^2 dr \right] = -4\pi \psi(r=0), \quad (4.43)$$

where we introduced spherical coordinates from Section 8.2. This result establishes the sifting property for $\nabla^2|\mathbf{x}|^{-1}$ and hence allows us to write

$$-\nabla^2 \frac{1}{4\pi |\mathbf{x}|} = \delta(\mathbf{x}). \quad (4.44)$$

4.4 One-dimensional Poisson's equation

To illustrate the Green's function method for solving differential equations, consider the one-dimensional Poisson equation on a finite domain

$$-\frac{d^2\psi}{dx^2} = \Lambda \quad \text{for } -L \leq x \leq L, \quad (4.45)$$

with boundary conditions unspecified at this point. The corresponding Green's function satisfies

$$-\frac{d^2G(x|x_0)}{dx^2} = \delta(x - x_0) \quad \text{for } -L \leq x, x_0 \leq L, \quad (4.46)$$

with boundary conditions taken as the homogeneous form of the boundary conditions satisfied by ψ . For the Green's function problem, the source, $\Lambda(x)$, appearing in the ψ equation has been replaced by the Dirac delta, $\delta(x - x_0)$. Since the Dirac delta has dimensions of inverse length, equation (4.46) implies that the Green's function has dimensions of length. Keeping track of physical dimensions offers a powerful means to check the integrity of the mathematical manipulations.

4.4.1 Integral expression for ψ in terms of G

Multiplying the differential equation (4.45) with G and the Green's function equation (4.46) with ψ , and then subtracting, yields

$$-G(x|x_0) \frac{d^2\psi(x)}{dx^2} + \psi(x) \frac{d^2G(x|x_0)}{dx^2} = G(x|x_0) \Lambda(x) - \psi(x) \delta(x - x_0). \quad (4.47)$$

Integrating this equation over the domain, and use of the sifting property (4.7), yields

$$\psi(x_0) = \int_{-L}^L G(x|x_0) \Lambda(x) dx + \int_{-L}^L \left[G(x|x_0) \frac{d^2\psi(x)}{dx^2} - \psi(x) \frac{d^2G(x|x_0)}{dx^2} \right] dx \quad (4.48a)$$

$$= \int_{-L}^L G(x|x_0) \Lambda(x) dx + \int_{-L}^L \frac{d}{dx} \left[G(x|x_0) \frac{d\psi}{dx} - \psi(x) \frac{dG(x|x_0)}{dx} \right] dx \quad (4.48b)$$

$$= \int_{-L}^L G(x|x_0) \Lambda(x) dx + \left[G(x|x_0) \frac{d\psi}{dx} \right]_{x=-L}^{x=L} - \left[\psi(x) \frac{dG(x|x_0)}{dx} \right]_{x=-L}^{x=L}, \quad (4.48c)$$

where the final expression exposed the boundary contributions. Specializing to *Dirichlet boundary conditions*, with $G(x = \pm L|x_0) = 0$, renders

$$\psi^{\text{Dirichlet}}(x_0) = \int_{-L}^L G(x|x_0) \Lambda(x) dx - \left[\psi(x) \frac{dG(x|x_0)}{dx} \right]_{x=L} + \left[\psi(x) \frac{dG(x|x_0)}{dx} \right]_{x=-L}, \quad (4.49)$$

whereas the solution with *Neumann boundary conditions*, where $dG(x = \pm L|x_0)/dx = 0$, is

$$\psi^{\text{Neumann}}(x_0) = \int_{-L}^L G(x|x_0) \Lambda(x) dx + \left[G(x|x_0) \frac{d\psi(x)}{dx} \right]_{x=L} - \left[G(x|x_0) \frac{d\psi(x)}{dx} \right]_{x=-L}. \quad (4.50)$$

4.4.2 Preliminary steps for obtaining the Green's function

The solution to the Green's function equation (4.46) is given by the linear functions

$$G(x|x_0) = \begin{cases} Ax + B & \text{for } -L \leq x \leq x_0 \\ Cx + D & \text{for } x_0 \leq x \leq L. \end{cases} \quad (4.51)$$

The constants A, B, C, D are determined by the following conditions: (i) jump condition for $dG(x|x_0)/dx$ at $x = x_0$, (ii) continuity of $G(x|x_0)$ at $x = x_0$, and (iii) homogeneous boundary conditions at $x = \pm L$.

Jump condition for the Green's function derivative at $x = x_0$

The Green's function differential equation (4.46) shows that the second derivative to the Green's function equals to a Dirac delta. Integrating this equation over an arbitrary region straddling $x = x_0$ yields the finite jump condition for the first derivative of the Green's function

$$\lim_{\epsilon \rightarrow 0} \left[\frac{dG(x|x_0)}{dx} \right]_{x=x_0+\epsilon} - \lim_{\epsilon \rightarrow 0} \left[\frac{dG(x|x_0)}{dx} \right]_{x=x_0-\epsilon} = -1. \quad (4.52)$$

That is, the derivative of the Green's function is not continuous across the point $x = x_0$. Instead, it has a finite jump. Making use of this condition in equation (4.51) leads to $C = A - 1$ so that

$$G(x|x_0) = \begin{cases} Ax + B & \text{for } -L \leq x \leq x_0 \\ (A - 1)x + D & \text{for } x_0 \leq x \leq L. \end{cases} \quad (4.53)$$

Continuity of the Green's function at $x = x_0$

A finite jump in the derivative at $x = x_0$ means that the Green's function is continuous at this point

$$\lim_{\epsilon \rightarrow 0} G(x = x_0 + \epsilon|x_0) = \lim_{\epsilon \rightarrow 0} G(x = x_0 - \epsilon|x_0). \quad (4.54)$$

Making use of this condition in equation (4.53) yields

$$G(x|x_0) = \begin{cases} Ax - x_0 + D & \text{for } -L \leq x \leq x_0 \\ (A - 1)x + D & \text{for } x_0 \leq x \leq L. \end{cases} \quad (4.55)$$

4.4.3 Green's function satisfying Dirichlet boundary conditions

The two constants A and D in equation (4.55) are determined by specifying the boundary conditions. With Dirichlet conditions we set $G(x = \pm L|x_0) = 0$, which then leads to

$$G(x|x_0) = \frac{1}{2L} \begin{cases} (L - x_0)(L + x) & \text{for } -L \leq x \leq x_0 \\ (L - x)(L + x_0) & \text{for } x_0 \leq x \leq L, \end{cases} \quad (4.56)$$

which is depicted in Figure 4.3. The structure of this function motivates the name *impulse response function* or *influence function*. We identify the following properties of this Green's function, with these properties also appearing in the higher dimension elliptical problems discussed in Sections 4.5 and 4.6.

Reciprocity condition

The Green's function (4.56) satisfies the symmetry condition

$$G(x|x_0) = G(x_0|x), \quad (4.57)$$

so that the Green's function is invariant under interchange of the source point, x_0 , and field point, x . In Section 4.5.4, we show how this *reciprocity condition* generally holds for all Green's functions of the Poisson equation.

Sum of free-space Green's function plus harmonic function

We can write the Green's function (4.56) as the sum of the free space Green's function (4.38a) plus a harmonic function

$$G(x|x_0) = -|x - x_0|/2 + \frac{L^2 - x x_0}{2 L} = \mathcal{G}(x|x_0) + H(x|x_0), \quad (4.58)$$

where

$$-\frac{d^2 H(x|x_0)}{dx^2} = 0 \quad \text{and} \quad H(x = \pm L|x_0) = -\mathcal{G}(x = \pm L|x_0). \quad (4.59)$$

As shown in Section 4.5.3, the Green's function for an elliptic PDE generally takes on this form.

Boundary condition satisfied by Green's function derivative

The Green's function is itself independent of the Dirichlet boundary conditions placed on the function ψ , as well as the source function Λ . Hence, we can establish general properties of the Green's function by considering special cases for the boundary conditions and source function. For this purpose, set $\Lambda = 0$ and $\psi(x = -L) = \psi(x = L)$, in which case ψ is itself a constant throughout the domain. The Dirichlet solution (4.49) then leads to the condition placed on the Green's function derivatives at the boundaries

$$\left[\frac{dG(x|x_0)}{dx} \right]_{x=L} - \left[\frac{dG(x|x_0)}{dx} \right]_{x=-L} = -1. \quad (4.60)$$

Interpreting G as a tracer concentration, this result means that the positive tracer input by the Dirac delta source is exactly balanced by a negative tracer departing the region through its boundaries. It is by this balance that the Green's function is able to maintain its homogeneous Dirichlet boundary condition, $G(x = \pm L|x_0) = 0$. A similar interpretation holds when G is the gravitational potential induced by the Dirac source in the region interior. In this case, for the gravitational potential to vanish on the boundaries, the positive gravitational acceleration induced by the Dirac delta source at $x = x_0$ must be compensated by an opposing gravitational acceleration along the boundaries.

4.4.4 Green's function satisfying Neumann boundary conditions

A constant is the only solution to $-d^2 G/dx^2 = 0$ with homogeneous Neumann boundary conditions, $dG/dx = 0$ at $x = \pm L$. However, a constant cannot satisfy the derivative jump condition at $x = x_0$, and thus cannot satisfy the differential equation $-d^2 G/dx^2 = \delta(x - x_0)$. Hence, there is no Green's function satisfying Neumann boundary conditions for a one-dimensional bounded domain. As we will see, in higher space dimensions there are nontrivial Green's functions for the Neumann boundary case. Even so, this result identifies a key condition required for a Green's function to exist. Namely, if the completely homogeneous problem ($\Lambda = 0$ with homogeneous boundary

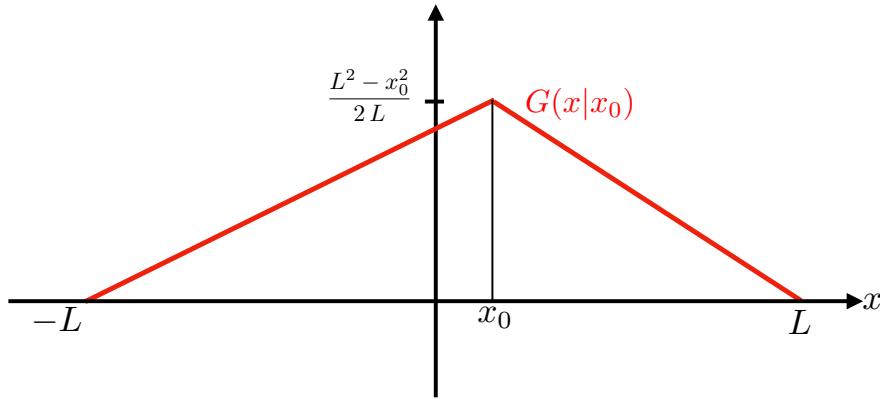


FIGURE 4.3: The Green's function (4.56) for the one-dimensional Poisson equation satisfying Dirichlet boundary conditions at $x = \pm L$. The Dirac delta source is located at $x = x_0$, which is the point where the Green's function has its maximum displacement of $(L^2 - x_0^2)/(2L)$.

conditions) only has the trivial solution, then the Green's function exists and is unique. For the case of Dirichlet boundary conditions

$$-\frac{d^2\psi}{dx^2} = 0 \quad \text{and} \quad \psi(x = \pm L) = 0 \implies \psi = 0, \quad (4.61)$$

whereas for Neumann boundary conditions

$$-\frac{d^2\psi}{dx^2} = 0 \quad \text{and} \quad \left[\frac{d\psi}{dx} \right]_{x=\pm L} = 0 \implies \psi = \text{constant}. \quad (4.62)$$

The existence of a nontrivial (an arbitrary constant) solution to the one-dimensional homogeneous Neumann boundary value problem signals the absence of a Green's function.

4.4.5 Uniqueness of the solution to the Dirichlet problem

Consider again the Dirichlet problem

$$-\frac{d^2\psi}{dx^2} = \Lambda \quad \text{for } -L \leq x \leq L \quad \text{with} \quad \psi(x = \pm L) = \text{prescribed constants}. \quad (4.63)$$

To prove that the solution to this problem is unique, assume there are two distinct solutions, ψ_a and ψ_b , and define $\Psi = \psi_a - \psi_b$. By construction, Ψ satisfies Laplace's equation with homogeneous Dirichlet boundaries

$$-\frac{d^2\Psi}{dx^2} = 0 \quad \text{for } -L \leq x \leq L, \quad \Psi(x = \pm L) = 0. \quad (4.64)$$

As discussed above, the only solution to this problem is $\Psi = 0$, which means that $\psi_a = \psi_b$, hence proving that the solution to equation (4.63) is unique. Uniqueness holds regardless the source function, so that the Green's function from Section 4.4.3 is also unique.

4.5 Poisson's equation with Dirichlet boundaries

In this section we develop the Green's function method for Poisson's equation in three-dimensions with a *Dirichlet boundary condition*

$$-\nabla^2\psi = \Lambda, \quad \text{for } \mathbf{x} \in \mathcal{R} \quad \text{and} \quad \psi = \sigma, \quad \text{for } \mathbf{x} \in \partial\mathcal{R}. \quad (4.65)$$

We separately consider the Neumann boundary condition in Section 4.6, whereby $\hat{\mathbf{n}} \cdot \nabla \psi$ is specified on the boundary rather than ψ itself. Although the Dirichlet and Neumann problems share many features, there are sufficient distinctions to warrant a separate treatment.

In the following, our mathematical goal is to determine an integral expression for ψ in terms of the distributed source and boundary data. Before doing so, we highlight the two elements determining the solution to any boundary value problem. First there is the partial differential operator (here the Laplacian) and the distributed source, Λ , both of which operate within the region interior, $\mathbf{x} \in \mathcal{R}$. Second, there is the boundary data that prescribes the solution (either its value or its boundary normal derivative) for points on the boundary, $\mathbf{x} \in \partial\mathcal{R}$. The distributed source can be defined everywhere, including the boundary. However, it only affects the solution for points within the region interior since the boundary conditions determine the boundary structure. Conversely, the boundary conditions are only prescribed along the boundary, and yet they affect the solution throughout the interior. A key feature of the Green's function method is that it clearly delineates these aspects of the solution, and in so doing it provides both a conceptual and computational framework.

4.5.1 Constraints on the source and boundary normal derivative

Assuming there exists a solution to the boundary value problem (4.65), we can derive a constraint on the normal derivative of ψ along the domain boundary.⁸ This constraint is revealed by integrating the Poisson equation (4.65) over the spatial domain, with the left hand side yielding

$$-\int_{\mathcal{R}} \nabla^2 \psi \, dV = -\oint_{\partial\mathcal{R}} \nabla \psi \cdot \hat{\mathbf{n}} \, dS, \quad (4.66)$$

where we made use of the divergence theorem. Equating this result to the integral of the source, Λ , leads to the constraint

$$\oint_{\partial\mathcal{R}} \nabla \psi \cdot \hat{\mathbf{n}} \, dS = -\int_{\mathcal{R}} \Lambda \, dV. \quad (4.67)$$

We can physically understand this constraint by invoking the steady state tracer diffusion interpretation of the Poisson equation. For a steady state solution to exist in the presence of specified boundary tracer concentration along with interior sources, there must be a balance between the normal gradient of the tracer concentration, as integrated over the boundary, plus the volume integrated source in the interior. This balance is required to maintain the specified Dirichlet boundary values in the presence of the interior source. Absent this balance, there will be depletion or accumulation of ψ that leads to a transient adjustment, thus breaking the steady state assumption. Notably, the constraint (4.67) can be realized for arbitrary Dirichlet boundary data, σ . An analogous interpretation holds when ψ is the gravitational potential resulting from the mass source, Λ . For ψ to be a specified value along the domain boundary requires the volume integrated mass source to balance a gravitational acceleration integrated along the boundary.

4.5.2 Uniqueness of the solution

Just like the one-dimensional case in Section 4.4.5, we can readily establish uniqueness of the solution to the Dirichlet problem by considering two functions, ψ_a and ψ_b , each satisfying the boundary value problem (4.65). In turn, their difference, $\Psi = \psi_a - \psi_b$, satisfies the homogeneous problem

$$-\nabla^2 \Psi = 0, \quad \text{for } \mathbf{x} \in \mathcal{R} \quad \text{and} \quad \Psi = 0, \quad \text{for } \mathbf{x} \in \partial\mathcal{R}. \quad (4.68)$$

⁸To be self-contained, we here repeat the discussion of the constraint (4.67) that is also provided in Section 3.5.6.

The solution to this Laplace equation is $\Psi = 0$, which then means that $\psi_a = \psi_b$ so that the solution to equation (4.65) is unique. This result also means that the Green's function considered in Section 4.5.3 is unique.

For the skeptic, let us provide a more formal proof that the solution to equation (4.68) is indeed zero. To do so, assume it is not zero and see what happens. First consider the case where Ψ is a constant, so that $\psi_a - \psi_b = \text{constant}$. But zero is the only constant that can satisfy the Dirichlet boundary condition, $\Psi = 0$. Next, assume Ψ has spatial dependence so that $\nabla\Psi \neq 0$ and consider the non-negative integral over the domain

$$\mathcal{I} = \int_{\mathcal{R}} \nabla\Psi \cdot \nabla\Psi \, dV. \quad (4.69)$$

Since $\nabla^2\Psi = 0$ we have $\nabla \cdot (\Psi \nabla\Psi) = \nabla\Psi \cdot \nabla\Psi$ so that

$$\mathcal{I} = \int_{\mathcal{R}} \nabla \cdot (\Psi \nabla\Psi) \, dV = \oint_{\partial\mathcal{R}} \Psi \nabla\Psi \cdot \hat{\mathbf{n}} \, dS, \quad (4.70)$$

where we used the divergence theorem. But since $\Psi = 0$ on $\partial\mathcal{R}$ we find that $\mathcal{I} = 0$, in which case $\nabla\Psi = 0$ everywhere. We are thus led to $\Psi = 0$ throughout the domain, thus allowing us to conclude that the solution to the boundary value problem (4.65) is indeed unique.

4.5.3 The Green's function problem

The Green's function corresponding to the boundary value problem (4.65) is the solution to the Poisson equation with a Dirac delta source (rather than Λ) and a homogeneous Dirichlet boundary condition (rather than σ)

$$-\nabla_{\mathbf{x}}^2 G(\mathbf{x}|\mathbf{x}_0) = \delta(\mathbf{x} - \mathbf{x}_0) \quad \text{for } \mathbf{x}, \mathbf{x}_0 \in \mathcal{R} \quad \text{and} \quad G(\mathbf{x}|\mathbf{x}_0) = 0, \quad \text{for } \mathbf{x} \in \partial\mathcal{R}. \quad (4.71)$$

For this three-dimensional problem, we see that the Green's function has dimensions of inverse length since the Dirac delta has dimensions of inverse volume. By construction, the Green's function, $G(\mathbf{x}|\mathbf{x}_0)$, is harmonic everywhere except at the location of the singular Dirac delta source, $\mathbf{x} = \mathbf{x}_0$, at which point the Green's function equals to minus the inverse Laplacian of the Dirac delta. Furthermore, the Green's function satisfies a homogenous Dirichlet boundary condition whenever the field point, \mathbf{x} , is on the boundary and for an arbitrary source position, \mathbf{x}_0 .

Connection to the free space Green's function

Since $G(\mathbf{x}|\mathbf{x}_0)$ satisfies a boundary condition, it is distinct from the free space Green's function, $\mathcal{G}(\mathbf{x}|\mathbf{x}_0)$, which has no concern for boundaries. Even so, linearity enables us to write the Green's function as the sum of the free space Green's function plus a harmonic function

$$G(\mathbf{x}|\mathbf{x}_0) = \mathcal{G}(\mathbf{x}|\mathbf{x}_0) + H(\mathbf{x}|\mathbf{x}_0), \quad (4.72)$$

where the free space Green's function and harmonic function satisfy

$$-\nabla_{\mathbf{x}}^2 \mathcal{G}(\mathbf{x}|\mathbf{x}_0) = \delta(\mathbf{x} - \mathbf{x}_0) \quad \text{for } \mathbf{x}, \mathbf{x}_0 \in \mathbb{R}^3 \quad (4.73a)$$

$$-\nabla_{\mathbf{x}}^2 H(\mathbf{x}|\mathbf{x}_0) = 0 \quad \text{for } \mathbf{x}, \mathbf{x}_0 \in \mathcal{R} \quad (4.73b)$$

$$-H(\mathbf{x}|\mathbf{x}_0) = \mathcal{G}(\mathbf{x}|\mathbf{x}_0) \quad \text{for } \mathbf{x} \in \partial\mathcal{R}. \quad (4.73c)$$

Assuming we know the free space Green's function from equation (4.38c), the mathematical problem of finding $G(\mathbf{x}|\mathbf{x}_0)$ reduces to finding the harmonic function, $H(\mathbf{x}|\mathbf{x}_0)$, satisfying the inhomogeneous Dirichlet boundary condition (4.73c).

Although the Green's function, G , cares about the existence of spatial boundaries, it is independent of the boundary data, σ , and the distributed source, Λ , that appear in the boundary value problem (4.65) for ψ . In that manner, the Green's function is connected to the original boundary value problem only through the differential operator (here the Laplacian) and the type of boundary condition (here the Dirichlet condition).

Jump condition induced by the Dirac source

Integrating the Green's function partial differential equation (4.71) over a volume, \mathcal{R}_0 , that encloses the Dirac delta source point at $\mathbf{x}_0 \in \mathcal{R}_0$ leads to

$$\int_{\mathcal{R}_0} \nabla_{\mathbf{x}} \cdot \nabla_{\mathbf{x}} G(\mathbf{x}|\mathbf{x}_0) dV = \oint_{\partial\mathcal{R}_0} \hat{\mathbf{n}} \cdot \nabla_{\mathbf{x}} G(\mathbf{x}|\mathbf{x}_0) dS = -1, \quad (4.74)$$

where we used the divergence theorem for the first equality. This result means that for any domain enclosing the Dirac source point, the normal derivative of the Green's function has a jump condition across the boundary of this region. We encountered this property in Section 4.4.2 for the one-dimensional Green's function, and will see it again when discussing the boundary Green's function in Section 4.5.6.

4.5.4 Reciprocity of the Green's function

The Green's function satisfies reciprocity

$$G(\mathbf{x}|\mathbf{x}_0) = G(\mathbf{x}_0|\mathbf{x}), \quad (4.75)$$

with this property holding for both Dirichlet and Neumann boundary conditions. As introduced in Section 4.4.3 for the one-dimensional Poisson equation, reciprocity says that the Green's function at the field point, \mathbf{x} , arising from a Dirac delta source at the source point, \mathbf{x}_0 is identical to the Green's function at the field point, \mathbf{x}_0 , arising from a Dirac delta source at point, \mathbf{x} . By inspection, the free space Green's functions (4.38a)-(4.38c) satisfy reciprocity. Hence, by implication the harmonic function, $H(\mathbf{x}|\mathbf{x}_0)$, also satisfies reciprocity. Even so, we find it useful pedagogically to offer a direct derivation of reciprocity by using steps similar to those used for establishing the second form of Green's integral identity (2.71). These steps are utilized for many purposes when working with Green's functions.

Derivation of reciprocity

Consider the two Green's functions, $G(\mathbf{x}|\mathbf{a})$ and $G(\mathbf{x}|\mathbf{b})$, arising from Dirac delta sources at source points $\mathbf{a} \in \mathcal{R}$ and $\mathbf{b} \in \mathcal{R}$

$$-\nabla_{\mathbf{x}}^2 G(\mathbf{x}|\mathbf{a}) = \delta(\mathbf{x} - \mathbf{a}) \quad (4.76a)$$

$$-\nabla_{\mathbf{x}}^2 G(\mathbf{x}|\mathbf{b}) = \delta(\mathbf{x} - \mathbf{b}). \quad (4.76b)$$

Multiplying the first equation by $G(\mathbf{x}, \mathbf{b})$ and the second by $G(\mathbf{x}, \mathbf{a})$, and subtracting leads to

$$-G(\mathbf{x}|\mathbf{b}) \nabla_{\mathbf{x}}^2 G(\mathbf{x}|\mathbf{a}) + G(\mathbf{x}|\mathbf{a}) \nabla_{\mathbf{x}}^2 G(\mathbf{x}|\mathbf{b}) = G(\mathbf{x}|\mathbf{b}) \delta(\mathbf{x} - \mathbf{a}) - G(\mathbf{x}|\mathbf{a}) \delta(\mathbf{x} - \mathbf{b}). \quad (4.77)$$

Now integrate this equation over the region \mathcal{R} , with the right hand side rendering

$$\int_{\mathcal{R}} [G(\mathbf{x}|\mathbf{b}) \delta(\mathbf{x} - \mathbf{a}) - G(\mathbf{x}|\mathbf{a}) \delta(\mathbf{x} - \mathbf{b})] dV = G(\mathbf{a}, \mathbf{b}) - G(\mathbf{b}, \mathbf{a}), \quad (4.78)$$

where we made use of the sifting property (4.7). Integrating the left hand side of equation (4.77) leads to

$$\int_{\mathcal{R}} [-G(\mathbf{x}|\mathbf{b}) \nabla_{\mathbf{x}}^2 G(\mathbf{x}|\mathbf{a}) + G(\mathbf{x}|\mathbf{a}) \nabla_{\mathbf{x}}^2 G(\mathbf{x}|\mathbf{b})] dV \quad (4.79a)$$

$$= \int_{\mathcal{R}} [-\nabla_{\mathbf{x}} \cdot [G(\mathbf{x}|\mathbf{b}) \nabla_{\mathbf{x}} G(\mathbf{x}|\mathbf{a})] + \nabla_{\mathbf{x}} \cdot [G(\mathbf{x}|\mathbf{a}) \nabla_{\mathbf{x}} G(\mathbf{x}|\mathbf{b})]] dV \quad (4.79b)$$

$$= \oint_{\partial\mathcal{R}} [-G(\mathbf{x}|\mathbf{b}) \nabla_{\mathbf{x}} G(\mathbf{x}|\mathbf{a}) + G(\mathbf{x}|\mathbf{a}) \nabla_{\mathbf{x}} G(\mathbf{x}|\mathbf{b})] \cdot \hat{\mathbf{n}} d\mathcal{S}, \quad (4.79c)$$

where we made use of the divergence theorem for the second equality. The final result vanishes when making use of either the homogeneous Dirichlet boundary condition (4.71) or the homogeneous Neumann boundary condition (4.102) considered in Section 4.6. We are thus led to the reciprocity relation (4.75).

Comments on self-adjoint operators

Reciprocity in the relatively simple form of equation (4.75) is a signature of *self-adjoint* differential operators. Self-adjointness reflects properties of both the differential operator and the boundary conditions, with the discussion in this section revealing that the Laplacian operator is self-adjoint with either Dirichlet or Neumann boundary conditions. Operators that are not self-adjoint, such as the diffusion operator in Section 4.7, satisfy a slightly more general reciprocity relation (see Section 4.7.4).

The absence of self-adjointness reflects some form of symmetry breaking either through the operator itself or through the boundary and/or initial conditions. The diffusion operator has a single time derivative, ∂_t , which breaks symmetry between past and future thus leading to the absence of self-adjointness. Furthermore, self-adjointness is a property that depends on the nature of the chosen inner product, with the inner product in the present discussion defined by integration over the domain \mathcal{R} . See [Stakgold \(2000a,b\)](#) for a thorough discussion accessible to physicists.

4.5.5 The integral solution

We have the elements in place to determine ψ as an integral expression involving $G(\mathbf{x}|\mathbf{x}_0)$ along with the prescribed source, Λ , and boundary data, σ . To do so we follow steps similar to those used to establish reciprocity. Recall the Poisson boundary value problem for ψ and the associated Green's function problem, here exposing arguments for clarity

$$-\nabla_{\mathbf{x}}^2 \psi(\mathbf{x}) = \Lambda(\mathbf{x}) \quad (4.80a)$$

$$-\nabla_{\mathbf{x}}^2 G(\mathbf{x}|\mathbf{x}_0) = \delta(\mathbf{x} - \mathbf{x}_0). \quad (4.80b)$$

Multiply the $\psi(\mathbf{x})$ equation by $G(\mathbf{x}|\mathbf{x}_0)$ and the $G(\mathbf{x}|\mathbf{x}_0)$ equation by $\psi(\mathbf{x})$ to find

$$-G(\mathbf{x}|\mathbf{x}_0) \nabla_{\mathbf{x}}^2 \psi(\mathbf{x}) = G(\mathbf{x}|\mathbf{x}_0) \Lambda(\mathbf{x}) \quad (4.81a)$$

$$-\psi(\mathbf{x}) \nabla_{\mathbf{x}}^2 G(\mathbf{x}|\mathbf{x}_0) = \psi(\mathbf{x}) \delta(\mathbf{x} - \mathbf{x}_0), \quad (4.81b)$$

and then subtract these two equations

$$-G(\mathbf{x}|\mathbf{x}_0) \nabla_{\mathbf{x}}^2 \psi(\mathbf{x}) + \psi(\mathbf{x}) \nabla_{\mathbf{x}}^2 G(\mathbf{x}|\mathbf{x}_0) = G(\mathbf{x}|\mathbf{x}_0) \Lambda(\mathbf{x}) - \psi(\mathbf{x}) \delta(\mathbf{x} - \mathbf{x}_0). \quad (4.82)$$

Now integrate over observational points, \mathbf{x} , sampled over the region \mathcal{R} , in which case the right hand side becomes

$$\int_{\mathcal{R}} [G(\mathbf{x}|\mathbf{x}_0) \Lambda(\mathbf{x}) - \psi(\mathbf{x}) \delta(\mathbf{x} - \mathbf{x}_0)] dV = -\psi(\mathbf{x}_0) + \int_{\mathcal{R}} G(\mathbf{x}|\mathbf{x}_0) \Lambda(\mathbf{x}) dV, \quad (4.83)$$

where we made use of the sifting property (4.7) to expose the function, ψ , at the location of the Dirac source, \mathbf{x}_0 . Integrating the left hand side of equation (4.82) yields

$$\int_{\mathcal{R}} [-G(\mathbf{x}|\mathbf{x}_0) \nabla_{\mathbf{x}}^2 \psi(\mathbf{x}) + \psi(\mathbf{x}) \nabla_{\mathbf{x}}^2 G(\mathbf{x}|\mathbf{x}_0)] dV \quad (4.84a)$$

$$= \int_{\mathcal{R}} \nabla_{\mathbf{x}} \cdot [-G(\mathbf{x}|\mathbf{x}_0) \nabla_{\mathbf{x}} \psi(\mathbf{x}) + \psi(\mathbf{x}) \nabla_{\mathbf{x}} G(\mathbf{x}|\mathbf{x}_0)] dV \quad (4.84b)$$

$$= \oint_{\partial\mathcal{R}} [-G(\mathbf{x}|\mathbf{x}_0) \nabla_{\mathbf{x}} \psi(\mathbf{x}) + \psi(\mathbf{x}) \nabla_{\mathbf{x}} G(\mathbf{x}|\mathbf{x}_0)] \cdot \hat{\mathbf{n}}_{\mathbf{x}} d\mathcal{S}. \quad (4.84c)$$

Bringing the two sides to equation (4.82) together leads to

$$\psi(\mathbf{x}_0) = \int_{\mathcal{R}} \Lambda(\mathbf{x}) G(\mathbf{x}|\mathbf{x}_0) dV + \oint_{\partial\mathcal{R}} [G(\mathbf{x}|\mathbf{x}_0) \nabla_{\mathbf{x}} \psi(\mathbf{x}) - \psi(\mathbf{x}) \nabla_{\mathbf{x}} G(\mathbf{x}|\mathbf{x}_0)] \cdot \hat{\mathbf{n}}_{\mathbf{x}} d\mathcal{S}. \quad (4.85)$$

As a final step, it is convenient to relabel $\mathbf{x}_0 \leftrightarrow \mathbf{x}$ and make use of reciprocity, $G(\mathbf{x}|\mathbf{x}_0) = G(\mathbf{x}_0|\mathbf{x})$, so that

$$\psi(\mathbf{x}) = \underbrace{\int_{\mathcal{R}} \Lambda(\mathbf{x}_0) G(\mathbf{x}|\mathbf{x}_0) dV_0}_{\text{volume integral over } \mathcal{R}} + \underbrace{\oint_{\partial\mathcal{R}} [G(\mathbf{x}|\mathbf{x}_0) \nabla_{\mathbf{x}_0} \psi(\mathbf{x}_0) - \psi(\mathbf{x}_0) \nabla_{\mathbf{x}_0} G(\mathbf{x}|\mathbf{x}_0)] \cdot \hat{\mathbf{n}}_{\mathbf{x}_0} d\mathcal{S}_0}_{\text{boundary area integral over } \partial\mathcal{R}}. \quad (4.86)$$

We have thus established ψ as a volume integral of the Green's function with the source, Λ , plus a boundary integral of the Green's function with the boundary data. This expression holds for either the Dirichlet condition or Neumann condition, whereby the distinction occurs only for the boundary integral. Furthermore, note that the Green's function is independent of the source, Λ , and the boundary data, so that $G(\mathbf{x}|\mathbf{x}_0)$ can be used to express ψ for arbitrary source functions and boundary data.

4.5.6 Properties of the solution

Specializing the general solution (4.86) to Dirichlet boundary conditions results in

$$\psi(\mathbf{x}) = \int_{\mathcal{R}} \Lambda(\mathbf{x}_0) G(\mathbf{x}|\mathbf{x}_0) dV_0 - \oint_{\partial\mathcal{R}} \sigma(\mathbf{x}_0) \nabla_{\mathbf{x}_0} G(\mathbf{x}|\mathbf{x}_0) \cdot \hat{\mathbf{n}}_{\mathbf{x}_0} d\mathcal{S}_0. \quad (4.87)$$

Notice how the boundary integral involves the normal gradient of the Green's function, which is consistent with the one-dimensional case (4.49) derived in Section 4.4.1. We refer to the inward normal gradient of the Green's function as the *boundary Green's function*

$$G^{bd}(\mathbf{x}|\mathbf{x}_0) = -\nabla_{\mathbf{x}_0} G(\mathbf{x}|\mathbf{x}_0) \cdot \hat{\mathbf{n}}_{\mathbf{x}_0} = -\frac{\partial G(\mathbf{x}|\mathbf{x}_0)}{\partial \hat{\mathbf{n}}_{\mathbf{x}_0}}, \quad (4.88)$$

in which case the Dirichlet solution takes the form

$$\psi(\mathbf{x}) = \int_{\mathcal{R}} \Lambda(\mathbf{x}_0) G(\mathbf{x}|\mathbf{x}_0) dV_0 + \oint_{\partial\mathcal{R}} \sigma(\mathbf{x}_0) G^{\text{bd}}(\mathbf{x}|\mathbf{x}_0) dS_0. \quad (4.89)$$

This form for the solution emphasizes the role of the Green's function as the mediator between the distributed source, Λ , and all surrounding points $\mathbf{x} \in \mathcal{R}$, whereas the boundary Green's function mediates between information prescribed along the boundary, $\mathbf{x} \in \partial\mathcal{R}$, and interior points. In this subsection we summarize certain properties of the solution (4.87) and (4.89), and infer (through insisting on self-consistency) corresponding properties of the Green's function and boundary Green's function.

Linear superposition

The Dirichlet solution (4.87) manifests the linear superposition principle by writing $\psi = \psi_A + \psi_B$ as given by Table 4.1. By construction, ψ_A satisfies Poisson's equation with homogeneous Dirichlet boundary conditions, whereas ψ_B satisfies Laplace's equation with inhomogeneous Dirichlet boundary conditions.

PDE: $\mathbf{x} \in \mathcal{R}$	BC: $\mathbf{x} \in \partial\mathcal{R}$	SOLUTION $\psi = \psi_A + \psi_B$
$-\nabla^2 \psi_A = \Lambda$	$\psi_A = 0$	$\psi_A(\mathbf{x}) = \int_{\mathcal{R}} G(\mathbf{x} \mathbf{x}_0) \Lambda(\mathbf{x}_0) dV_0$
$-\nabla^2 \psi_B = 0$	$\psi_B = \sigma$	$\psi_B(\mathbf{x}) = \oint_{\partial\mathcal{R}} G^{\text{bd}}(\mathbf{x} \mathbf{x}_0) \sigma(\mathbf{x}_0) dS_0$

TABLE 4.1: Decomposing the Dirichlet solution (4.87) into $\psi = \psi_A + \psi_B$, with ψ_A and ψ_B satisfying the properties shown in this table. For the boundary contribution we made use of the boundary Green's function, $G^{\text{bd}}(\mathbf{x}|\mathbf{x}_0) = -\nabla_{\mathbf{x}_0} G(\mathbf{x}|\mathbf{x}_0) \cdot \hat{\mathbf{n}}_{\mathbf{x}_0}$, as given by equation (4.88).

Verifying the partial differential equation for $\mathbf{x} \in \mathcal{R}$

We derived the Dirichlet solution (4.87) with manipulations that are reversible; i.e., equal signs were used at every step. Hence, we know that the expression (4.87) indeed satisfies the Dirichlet boundary value problem (4.65). Even so, the exercise of verifying the solution reveals valuable insights into the Green's function.

To verify that the partial differential equation is satisfied for points within the interior of the region, $\mathbf{x} \in \mathcal{R}$, operate with $-\nabla_{\mathbf{x}}^2$ on equation (4.89) to find

$$-\nabla_{\mathbf{x}}^2 \psi(\mathbf{x}) = \int_{\mathcal{R}} \Lambda(\mathbf{x}_0) [-\nabla_{\mathbf{x}}^2 G(\mathbf{x}|\mathbf{x}_0)] dV_0 + \oint_{\partial\mathcal{R}} \sigma(\mathbf{x}_0) [-\nabla_{\mathbf{x}}^2 G^{\text{bd}}(\mathbf{x}|\mathbf{x}_0)] dS_0. \quad (4.90)$$

For the first term on the right hand side, make use of the Green's function identity $-\nabla_{\mathbf{x}}^2 G(\mathbf{x}|\mathbf{x}_0) = \delta(\mathbf{x} - \mathbf{x}_0)$ and then use the sifting property of the Dirac delta, $\int_{\mathcal{R}} \Lambda(\mathbf{x}_0) \delta(\mathbf{x} - \mathbf{x}_0) dV_0 = \Lambda(\mathbf{x})$. To show that the boundary contribution vanishes, make use of the following

$$\nabla_{\mathbf{x}}^2 G^{\text{bd}}(\mathbf{x}|\mathbf{x}_0) = \nabla_{\mathbf{x}}^2 [-\nabla_{\mathbf{x}_0} G(\mathbf{x}|\mathbf{x}_0) \cdot \hat{\mathbf{n}}_{\mathbf{x}_0}] \quad (4.91a)$$

$$= \hat{\mathbf{n}}_{\mathbf{x}_0} \cdot \nabla_{\mathbf{x}_0} [-\nabla_{\mathbf{x}}^2 G(\mathbf{x}|\mathbf{x}_0)] \quad (4.91b)$$

$$= \hat{\mathbf{n}}_{\mathbf{x}_0} \cdot \nabla_{\mathbf{x}_0} \delta(\mathbf{x} - \mathbf{x}_0), \quad (4.91c)$$

and, without loss of generality, let the normal direction be parallel to the vertical, $\hat{\mathbf{n}}_{\mathbf{x}_0} = \hat{\mathbf{z}}$, so that

$$\hat{\mathbf{n}}_{\mathbf{x}_0} \cdot \nabla_{\mathbf{x}_0} \delta(\mathbf{x} - \mathbf{x}_0) = \partial_{z_0} [\delta(x - x_0) \delta(y - y_0) \delta(z - z_0)] = \delta(x - x_0) \delta(y - y_0) \frac{d[\delta(z - z_0)]}{dz_0}. \quad (4.92)$$

With $\mathbf{x}_0 \in \partial\mathcal{R}$ yet $\mathbf{x} \notin \partial\mathcal{R}$, the Dirac delta never fires, thus eliminating the boundary contribution. We have thus verified that $-\nabla_{\mathbf{x}}^2 \psi = \Lambda$ for points $\mathbf{x} \in \mathcal{R}$.

Verifying the Dirichlet boundary condition for $\mathbf{x} \in \partial\mathcal{R}$

To verify that the boundary conditions are satisfied by the Dirichlet solution (4.87), bring the field point, \mathbf{x} , onto the boundary

$$\mathbf{x} \rightarrow \mathbf{x}_{\partial\mathcal{R}} \in \partial\mathcal{R}. \quad (4.93)$$

For such boundary points, the volume integral in the solution (4.87) vanishes since the Dirichlet Green's function vanishes on the boundary, $G(\mathbf{x}_{\partial\mathcal{R}}|\mathbf{x}_0) = 0$. Self-consistency with the Dirichlet boundary data, $\psi(\mathbf{x}_{\partial\mathcal{R}}) = \sigma(\mathbf{x}_{\partial\mathcal{R}})$, leads to

$$\sigma(\mathbf{x}) = \oint_{\partial\mathcal{R}} \sigma(\mathbf{x}_0) G^{\text{bd}}(\mathbf{x}|\mathbf{x}_0) dS_0 = - \oint_{\partial\mathcal{R}} \sigma(\mathbf{x}_0) \nabla_{\mathbf{x}_0} G(\mathbf{x}_{\partial\mathcal{R}}|\mathbf{x}_0) \cdot \hat{\mathbf{n}}_{\mathbf{x}_0} dS_0 \quad \text{for } \mathbf{x} \in \partial\mathcal{R}. \quad (4.94)$$

This integral equation is consistent so long as the boundary Green's function satisfies the boundary condition

$$G^{\text{bd}}(\mathbf{x}|\mathbf{x}_0) = -\nabla_{\mathbf{x}_0} G(\mathbf{x}|\mathbf{x}_0) \cdot \hat{\mathbf{n}}_{\mathbf{x}_0} = \delta^{(2)}(\mathbf{x} - \mathbf{x}_0) \quad \text{for } \mathbf{x}, \mathbf{x}_0 \in \partial\mathcal{R}, \quad (4.95)$$

where $\delta^{(2)}(\mathbf{x} - \mathbf{x}_0)$ is the surface Dirac delta with physical dimensions of inverse area. This property of the boundary Green's function is consistent with the jump condition (4.98) found for the one-dimensional Poisson equation. It furthermore leads to the corresponding integral identity

$$\oint_{\partial\mathcal{R}} G^{\text{bd}}(\mathbf{x}|\mathbf{x}_0) dS_0 = \oint_{\partial\mathcal{R}} \delta^{(2)}(\mathbf{x} - \mathbf{x}_0) dS_0 = 1 \quad \text{for } \mathbf{x}, \mathbf{x}_0 \in \partial\mathcal{R}. \quad (4.96)$$

We emphasize⁹ that the expressions (4.95) and (4.96) are found by first placing the source point, \mathbf{x}_0 , on the boundary and thereafter moving the field point to the boundary, $\mathbf{x} \rightarrow \mathbf{x}_{\partial\mathcal{R}} \in \partial\mathcal{R}$.

4.5.7 Boundary Green's function

Boundary value problem for the boundary Green's function

As part of the development in Section 4.5.6, we revealed that the boundary Green's function satisfies the following boundary value problem

$$-\nabla_{\mathbf{x}}^2 G^{\text{bd}}(\mathbf{x}|\mathbf{x}_0) = 0 \quad \text{for } \mathbf{x} \in \mathcal{R} \quad \text{and} \quad G^{\text{bd}}(\mathbf{x}|\mathbf{x}_0) = \delta^{(2)}(\mathbf{x} - \mathbf{x}_0) \quad \text{for } \mathbf{x}, \mathbf{x}_0 \in \partial\mathcal{R}. \quad (4.97)$$

As seen by equation (4.71), the Dirichlet Green's function, $G(\mathbf{x}|\mathbf{x}_0)$, feels the Dirac source in the interior of the domain, $\mathbf{x} \in \mathcal{R}$, and satisfies homogeneous Dirichlet boundary conditions for $\mathbf{x} \in \partial\mathcal{R}$. As a complement, the boundary Green's function, $G^{\text{bd}}(\mathbf{x}|\mathbf{x}_0)$, is harmonic everywhere in the domain interior and yet equals to the Dirac source along the boundary. By construction, the boundary Green's function incorporates boundary information into the region as part of the Dirichlet solution (4.89). We make further use of the boundary Green's function when studying the diffusion equation in Section 4.7 and the advection-diffusion equation in Section 48.10, at which point the boundary Green's function is referred to as the *boundary propagator*.

⁹As per page 801 of *Morse and Feshbach* (1953).

Normalization of the boundary Green's function

Consider the special case of constant boundary data, $\sigma = \sigma_{\text{const}}$, in which case the harmonic function in Table 4.1 is itself a constant, $\psi_B(\mathbf{x}) = \sigma_{\text{const}}$. Consequently, the boundary Green's function satisfies

$$-\oint_{\partial\mathcal{R}} \nabla_{\mathbf{x}_0} G(\mathbf{x}|\mathbf{x}_0) \cdot \hat{\mathbf{n}}_{\mathbf{x}_0} d\mathcal{S}_0 = \oint_{\partial\mathcal{R}} G^{\text{bd}}(\mathbf{x}|\mathbf{x}_0) d\mathcal{S}_0 = 1. \quad (4.98)$$

This relation is built by placing Dirac delta sources along the boundary, $\mathbf{x}_0 \in \partial\mathcal{R}$, and then area integrating over the boundary area. It holds for any field point, $\mathbf{x} \in \mathcal{R}$. Although derived by considering the special case of constant boundary data, equation (4.98) holds in general since the Green's function is independent of the boundary data.

To help understand the identity (4.98), consider the Green's function to be the steady state temperature or tracer concentration resulting from a Dirac delta source placed within the domain interior. The area integrated condition (4.98) acts to maintain the homogeneous Dirichlet boundary condition, $G(\mathbf{x}_{\partial\mathcal{R}}|\mathbf{x}_0) = 0$, for every point $\mathbf{x}_{\partial\mathcal{R}} \in \partial\mathcal{R}$. It does so by providing a boundary flux, via the normal gradient, whose area integral precisely cancels the unit positive source from the Dirac delta. We encountered this property for the one-dimensional equation studied in Section 4.4.3.

Equation (4.98) represents a normalization of the boundary Green's function at each point within the region interior, and it is consistent with the boundary condition given in equation (4.97). If we place Dirac delta sources along the boundary and integrate over the boundary, then every point within the domain feels a net unit source from these boundary sources, which is reflected by the normalization of the boundary Green's function. We return to this normalization when considering the boundar propagator for the diffusion equation in Section 4.7.7.

4.6 Poisson's equation with Neumann boundaries

We now switch from the Dirichlet boundary conditions in Section 4.5 to the Poisson equation with *Neumann boundary conditions*

$$-\nabla^2\psi = \Lambda, \quad \text{for } \mathbf{x} \in \mathcal{R} \quad \text{and} \quad \hat{\mathbf{n}} \cdot \nabla\psi = \Sigma, \quad \text{for } \mathbf{x} \in \partial\mathcal{R}. \quad (4.99)$$

Rather than specifying the value of ψ along the boundary, the Neumann condition specifies the normal derivative. Although much is shared between the Dirichlet and Neumann problems, there are distinctions that arise in the present section.

4.6.1 Constraints on the source and boundary data

As in our discussion of Dirichlet boundary conditions in Section 4.5.1, we can realize a solution to the boundary value problem (4.99) only so long as the constraint (4.67) is satisfied. For Neumann boundary conditions, we specify the normal derivative along the boundary so that the constraint (4.67) is now imposed on the volume source and boundary data¹⁰

$$\oint_{\partial\mathcal{R}} \Sigma dS = - \int_{\mathcal{R}} \Lambda dV. \quad (4.100)$$

If the source and boundary data do not satisfy this constraint, then this is no solution to the Poisson problem (4.99). If the Poisson problem arises physically from steady state tracer diffusion, then the constraint (4.100) imposes a balance between the diffusive flux integrated around the boundary

¹⁰To be self-contained, we here repeat the discussion of the constraint (4.100) that is also provided in Section 3.5.6.

(left hand side) with the volume integrated tracer source (right hand side). In the absence of this balance, there is no solution to the Poisson problem thus indicating the presence of transients (i.e., time dependent diffusion). If the Poisson problem arises from Newtonian gravity, then the condition (4.100) means that the area integrated gravitational acceleration specified on the boundary must be consistent with the volume integrated mass source distributed within the domain.

4.6.2 Uniqueness of the solution up to a constant

As in Section 4.5.2, we consider the uniqueness of the solution to the boundary value problem (4.99). We do so, again, by considering two functions, ψ_A and ψ_B , each satisfying the boundary value problem (4.99) and noting that their difference, $\Psi = \psi_A - \psi_B$, satisfies the homogeneous problem

$$-\nabla^2\Psi = 0, \quad \text{for } \mathbf{x} \in \mathcal{R} \quad \text{and} \quad \hat{\mathbf{n}} \cdot \nabla\Psi = 0, \quad \text{for } \mathbf{x} \in \partial\mathcal{R}. \quad (4.101)$$

The same arguments we used in Section 4.5.2 lead us to conclude that $\nabla\Psi = 0$, but for the Neumann problem this result is consistent with Ψ being an arbitrary spatial constant. We can understand this arbitrariness since the Neumann boundary condition involves a derivative, with the derivative of a constant vanishing. We thus conclude that the solution to the boundary value problem (4.99) is unique up to an arbitrary constant.

4.6.3 The Green's function problem

The Green's function corresponding to the Poisson boundary value problem (4.99) is the solution to the Poisson equation with a Dirac delta source and a homogeneous Neumann boundary condition

$$-\nabla_{\mathbf{x}}^2 G(\mathbf{x}|\mathbf{x}_0) = \delta(\mathbf{x} - \mathbf{x}_0) \quad \text{for } \mathbf{x}, \mathbf{x}_0 \in \mathcal{R} \quad \text{and} \quad \hat{\mathbf{n}} \cdot \nabla_{\mathbf{x}} G(\mathbf{x}|\mathbf{x}_0) = 0, \quad \text{for } \mathbf{x} \in \partial\mathcal{R}. \quad (4.102)$$

Linearity enables us to write the Green's function as the sum of the free space Green's function plus a harmonic function

$$G(\mathbf{x}|\mathbf{x}_0) = \mathcal{G}(\mathbf{x}|\mathbf{x}_0) + H(\mathbf{x}|\mathbf{x}_0), \quad (4.103)$$

where

$$-\nabla_{\mathbf{x}}^2 \mathcal{G}(\mathbf{x}|\mathbf{x}_0) = \delta(\mathbf{x} - \mathbf{x}_0) \quad \text{for } \mathbf{x}, \mathbf{x}_0 \in \mathbb{R}^3 \quad (4.104a)$$

$$-\nabla_{\mathbf{x}}^2 H(\mathbf{x}|\mathbf{x}_0) = 0 \quad \text{for } \mathbf{x}, \mathbf{x}_0 \in \mathcal{R} \quad (4.104b)$$

$$-\hat{\mathbf{n}} \cdot \nabla_{\mathbf{x}} H(\mathbf{x}|\mathbf{x}_0) = \hat{\mathbf{n}} \cdot \nabla_{\mathbf{x}} \mathcal{G}(\mathbf{x}|\mathbf{x}_0) \quad \text{for } \mathbf{x} \in \partial\mathcal{R}. \quad (4.104c)$$

4.6.4 Properties of the solution

Specializing the general solution (4.86) to Neumann boundary conditions results in

$$\psi(\mathbf{x}) = \int_{\mathcal{R}} \Lambda(\mathbf{x}_0) G(\mathbf{x}|\mathbf{x}_0) dV_0 + \oint_{\partial\mathcal{R}} \Sigma(\mathbf{x}_0) G(\mathbf{x}|\mathbf{x}_0) dS_0. \quad (4.105)$$

In this section we summarize certain properties of the solution (4.105) along with properties of the Green's function that are inferred (through insisting on self-consistency) from the solution.

Linear superposition

This Green's function solution (4.105) manifests the linear superposition principle by writing $\psi = \psi_A + \psi_B$ as given by Table 4.2. By construction, ψ_A satisfies Poisson's equation with homogeneous Neumann boundary conditions, whereas ψ_B satisfies Laplace's equation with inhomogeneous Neumann boundary conditions.

PDE: $\mathbf{x} \in \mathcal{R}$	BC: $\mathbf{x} \in \partial\mathcal{R}$	SOLUTION $\psi = \psi_A + \psi_B$
$-\nabla^2\psi_A = \Lambda$	$\psi_A = 0$	$\psi_A(\mathbf{x}) = \int_{\mathcal{R}} G(\mathbf{x} \mathbf{x}_0) \Lambda(\mathbf{x}_0) dV_0$
$-\nabla^2\psi_B = 0$	$\hat{\mathbf{n}} \cdot \nabla\psi_B = \Sigma$	$\psi_B(\mathbf{x}) = \oint_{\partial\mathcal{R}} \Sigma(\mathbf{x}_0) G(\mathbf{x} \mathbf{x}_0) dS_0$

TABLE 4.2: Decomposing the Neumann solution (4.105) into $\psi = \psi_A + \psi_B$, with ψ_A and ψ_B satisfying the properties shown here. The function ψ_A is identical to ψ_A appearing for the Dirichlet problem (see Table 4.1), whereas the function ψ_B is distinct since it is determined by the respective boundary conditions.

Transforming the boundary data to the interior

The Neumann solution (4.105) allows for a dual formulation of the boundary data. To motivate this formulation, assume the only nontrivial Neumann boundary data appears along the constant geopotential surface $z = z_b$, with all other boundaries maintaining the homogeneous boundary condition, $\hat{\mathbf{n}} \cdot \nabla\psi = 0$. In this special case the Neumann solution (4.105) takes the form

$$\psi(\mathbf{x}) = \int_{\mathcal{R}} \Lambda(\mathbf{x}_0) G(\mathbf{x}|\mathbf{x}_0) dV_0 + \int_{z=z_b} \Sigma(x_0, y_0) G(\mathbf{x}|\mathbf{x}_0) dx_0 dy_0 \quad (4.106a)$$

$$= \int_{\mathcal{R}} [\Lambda(\mathbf{x}_0) + \Sigma(x_0, y_0) \delta(z_0 - z_b)] G(\mathbf{x}|\mathbf{x}_0) dV_0 \quad (4.106b)$$

$$= \int_{\mathcal{R}} \Lambda^*(\mathbf{x}_0) G(\mathbf{x}|\mathbf{x}_0) dV_0. \quad (4.106c)$$

These manipulations have absorbed the non-homogeneous Neumann boundary condition into a modified source, $\Lambda^*(\mathbf{x})$. We are thus led to two equivalent formulations for the Poisson boundary value problem with Neumann conditions. The first is given by equation (4.99), whereby ψ is the solution to the Poisson equation with source Λ and with inhomogeneous Neumann boundary data, Σ . The second formulation considers ψ to be the solution to Poisson's equation with homogeneous Neumann boundary conditions yet with a modified source function

$$\Lambda^*(\mathbf{x}) = \Lambda(\mathbf{x}) + \Sigma(x, y) \delta(z - z_b) = \Lambda(\mathbf{x}) + \partial_z \psi(\mathbf{x}) \delta(z - z_b). \quad (4.107)$$

As a check, note that the physical dimensionality of $\Lambda^*(\mathbf{x})$ is indeed correct since the Dirac delta, $\delta(z - z_b)$, has dimensions of inverse length. We interpret the term $\Sigma(x, y) \delta(z - z_b)$ as a flux or *Dirac delta sheet* that sits just inside the boundary (at $z = z_b - \epsilon$ with $\epsilon \rightarrow 0$), which allows this data to be incorporated into the volume source data rather than be part of the boundary data. The specific form (4.107) can be generalized to the expression

$$\Lambda^*(\mathbf{x}) = \Lambda(\mathbf{x}) + \Sigma(\mathbf{x}) \delta[\hat{\mathbf{n}} \cdot (\mathbf{x} - \mathbf{x}_{\partial\mathcal{R}})] = \Lambda(\mathbf{x}) + \hat{\mathbf{n}} \cdot \nabla\psi(\mathbf{x}) \delta[\hat{\mathbf{n}} \cdot (\mathbf{x} - \mathbf{x}_{\partial\mathcal{R}})], \quad (4.108)$$

where the argument to the Dirac delta picks out the coordinates in the direction of the outward normal. The transformed Neumann problem thus takes the generic form

$$-\nabla^2\psi = \Lambda^*, \quad \text{for } \mathbf{x} \in \mathcal{R} \quad \text{and} \quad \hat{\mathbf{n}} \cdot \nabla\psi = 0, \quad \text{for } \mathbf{x} \in \partial\mathcal{R}. \quad (4.109)$$

Again, the solution to the boundary value problem (4.109) is identical to the solution of the original problem (4.99).

Notably, the same data, Σ , is needed for both equations (4.109) and (4.99), so there is nothing fundamentally special or efficient about one formulation or the other. Rather, it is a matter of convenience. For example, the formulation using Λ^* has found some favor in the study of quasi-geostrophic potential vorticity ([Bretherton, 1966](#)). We also make use of a similar construct in Section 52.5.3 for studies of boundary buoyancy fluxes in water mass transformation analysis.

4.6.5 Verifying the solution

We verify that the solution (4.105) solves the Poisson boundary value problem (4.99). First consider a point in the interior of the region, $\mathbf{x} \in \mathcal{R}$, and apply $-\nabla_{\mathbf{x}}^2$ on equation to find (4.105)

$$-\nabla_{\mathbf{x}}^2 \psi(\mathbf{x}) = \int_{\mathcal{R}} \Lambda(\mathbf{x}_0) [-\nabla_{\mathbf{x}}^2 G(\mathbf{x}|\mathbf{x}_0)] dV_0 + \oint_{\partial\mathcal{R}} \Sigma(\mathbf{x}_0) [-\nabla_{\mathbf{x}}^2 G(\mathbf{x}|\mathbf{x}_0)] d\mathcal{S}_0 = \Lambda(\mathbf{x}). \quad (4.110)$$

To reach this result we used $-\nabla_{\mathbf{x}}^2 G(\mathbf{x}|\mathbf{x}_0) = \delta(\mathbf{x} - \mathbf{x}_0)$ and noted that the boundary integral vanishes for $\mathbf{x} \notin \partial\mathcal{R}$.

To verify the boundary condition, act with the gradient, $\nabla_{\mathbf{x}}$, on equation (4.105)

$$\nabla_{\mathbf{x}} \psi(\mathbf{x}) = \int_{\mathcal{R}} \Lambda(\mathbf{x}_0) [\nabla_{\mathbf{x}} G(\mathbf{x}|\mathbf{x}_0)] dV_0 + \oint_{\partial\mathcal{R}} \Sigma(\mathbf{x}_0) [\nabla_{\mathbf{x}} G(\mathbf{x}|\mathbf{x}_0)] d\mathcal{S}_0. \quad (4.111)$$

Now move the field point onto the boundary, $\mathbf{x} \rightarrow \mathbf{x}_{\partial\mathcal{R}} \in \partial\mathcal{R}$, and project the gradient onto the normal direction, $\hat{\mathbf{n}}_{\mathbf{x}}$, at the point $\mathbf{x}_{\partial\mathcal{R}}$. This projection leads to $\hat{\mathbf{n}}_{\mathbf{x}} \cdot \nabla_{\mathbf{x}} \psi(\mathbf{x}) = \Sigma(\mathbf{x})$ on the left hand side, and it annihilates the volume term on the right hand side since $\hat{\mathbf{n}}_{\mathbf{x}} \cdot \nabla_{\mathbf{x}} G(\mathbf{x}|\mathbf{x}_0) = 0$ for $\mathbf{x} \in \partial\mathcal{R}$. We are thus left with

$$\hat{\mathbf{n}} \cdot \nabla_{\mathbf{x}} \psi(\mathbf{x}) = \Sigma(\mathbf{x}) = \oint_{\partial\mathcal{R}} \Sigma(\mathbf{x}_0) [\hat{\mathbf{n}}_{\mathbf{x}} \cdot \nabla_{\mathbf{x}} G(\mathbf{x}|\mathbf{x}_0)] d\mathcal{S}_0 \quad \text{for } \mathbf{x} \in \partial\mathcal{R}. \quad (4.112)$$

Again, both arguments of the Green's function are on the boundary, $\mathbf{x}, \mathbf{x}_0 \in \partial\mathcal{R}$, with the Dirac source point, \mathbf{x}_0 , integrated around the boundary whereas the field point, \mathbf{x} , is an arbitrary point on the boundary. Equation (4.94) is the analogous integral equation for the solution with Dirichlet boundary conditions. As for the Dirichlet case, we are ensured a solution to the integral equation (4.112) if the Green's function satisfies the property

$$\hat{\mathbf{n}}_{\mathbf{x}} \cdot \nabla_{\mathbf{x}} G(\mathbf{x}|\mathbf{x}_0) = \delta^{(2)}(\mathbf{x} - \mathbf{x}_0) \quad \text{for } \mathbf{x}, \mathbf{x}_0 \in \partial\mathcal{R}, \quad (4.113)$$

which takes on the integral expression

$$\oint_{\partial\mathcal{R}} \hat{\mathbf{n}}_{\mathbf{x}} \cdot \nabla_{\mathbf{x}} G(\mathbf{x}|\mathbf{x}_0) d\mathcal{S}_0 = \oint_{\partial\mathcal{R}} \delta^{(2)}(\mathbf{x} - \mathbf{x}_0) d\mathcal{S}_0 \quad \text{for } \mathbf{x} \in \partial\mathcal{R}. \quad (4.114)$$

Table 4.3 compares the boundary properties of the respective Green's functions.

BOUNDARY CONDITION	GREEN'S FUNCTION PROPERTY	EQUATION
Dirichlet	$\partial G / \partial \hat{\mathbf{n}}_0 = \hat{\mathbf{n}}_{\mathbf{x}_0} \cdot \nabla_{\mathbf{x}_0} G(\mathbf{x} \mathbf{x}_0) = -\delta^{(2)}(\mathbf{x} - \mathbf{x}_0)$	(4.95)
Neumann	$\partial G / \partial \hat{\mathbf{n}} = \hat{\mathbf{n}}_{\mathbf{x}} \cdot \nabla_{\mathbf{x}} G(\mathbf{x} \mathbf{x}_0) = \delta^{(2)}(\mathbf{x} - \mathbf{x}_0)$	(4.113)

TABLE 4.3: Comparing the boundary normal derivatives for the Poisson equation Green's function with Dirichlet and Neumann boundary conditions. Each point is on the boundary, $\mathbf{x}, \mathbf{x}_0 \in \partial\mathcal{R}$, and the Dirac delta is two-dimensional so it has dimensions of inverse area. The properties satisfied by these Green's functions are realized by first placing the source point, \mathbf{x}_0 , on the boundary and then moving the field point to the boundary, $\mathbf{x} \rightarrow \mathbf{x}_{\partial\mathcal{R}} \in \partial\mathcal{R}$. For the Dirichlet condition, the normal derivative is computed at the source point, whereas the Neumann condition computes the normal derivative at the field point.

4.7 The diffusion equation

We continue building our understanding of Green's function methods by now studying the *diffusion equation*, also known as the *heat equation*. We first encountered the diffusion equation in Section 3.6, where it served as the canonical parabolic partial differential equation. It is more thoroughly explored in Chapter 48, with much of the current chapter a foundation for the Green's function treatment of the advection-diffusion equation in Section 48.10. Time dependence appearing in the diffusion equation is the new feature relative to the elliptic Poisson equation encountered thus far in this chapter.

The traditional Green's function formalism for the diffusion equation, and the one pursued in this chapter, assumes that the spatial domain, \mathcal{R} , is static. Doing so ensures that space and time operations commute, such as differentiation and integration. Although time independent domains are suitable for many applications, certain geophysical applications of interest in this book make use of time dependent boundaries, with the ocean free surface the key example. We examine time dependent boundaries when studying the advection-diffusion equation in Section 48.10, with that generality best confronted only after further exercising our Green's function brain muscle in the current chapter.

4.7.1 Causal free space Green's function

Consider a Dirac delta tracer source at (\mathbf{x}_0, t_0) and assume a continuous media where there is no advection, such as found in a stagnant fluid or elastic solid. The simplest solution to the diffusion equation is known as the *causal free space Green's function*, which is defined for $t \in (-\infty, \infty)$ and satisfies

$$(\partial_t - \kappa \nabla_{\mathbf{x}}^2) \mathcal{G}(\mathbf{x}, t | \mathbf{x}_0, t_0) = \delta(t - t_0) \delta(\mathbf{x} - \mathbf{x}_0) \quad \mathbf{x} \in \mathbb{R}^n \quad (4.115a)$$

$$\mathcal{G}(\mathbf{x}, t | \mathbf{x}_0, t_0) \rightarrow 0 \quad |\mathbf{x} - \mathbf{x}_0| \rightarrow \infty \quad (4.115b)$$

$$\mathcal{G}(\mathbf{x}, t | \mathbf{x}_0, t_0) = 0 \quad \mathbf{x} \in \mathbb{R}^n, \quad t < t_0, \quad (4.115c)$$

where \mathbb{R}^n is Euclidean space with $n = 1, 2, 3$ for a line, plane, and volume, respectively, and where $\kappa > 0$ is a constant diffusivity (dimensions $L^2 T^{-1}$). Equation (4.115b) ensures that the free space Green's function, \mathcal{G} , decays as the field point gets further away from the source point. The *causality condition* (4.115c) means that the Green's function vanishes for times prior to the time, t_0 , at which the Dirac source occurs.

The causal free-space Green's function is given by

$$\mathcal{G}(\mathbf{x}, t | \mathbf{x}_0, t_0) = \frac{H(t - t_0)}{[4\pi\kappa(t - t_0)]^{n/2}} e^{-|\mathbf{x} - \mathbf{x}_0|^2/[4\kappa(t - t_0)]}, \quad (4.116)$$

where the Heaviside step function (equation (4.16)) enforces causality. The amplitude of the Green's function exponentially decays when moving away from the source location, thus satisfying the condition (4.115b). Additionally, as time progresses beyond the source time, the Green's function decays according to the pre-factor $(t - t_0)^{-n/2}$. Each of these properties of the causal free space Green's function are reflected in solutions to more general diffusion processes.

4.7.2 Causal Green's function

Moving beyond the free space solution, we next introduce the *causal Green's function* for the diffusion equation. This Green's function is defined for $t \in (-\infty, \infty)$ and satisfies the following

equations when assuming Neumann boundary conditions

$$\frac{\partial[G(\mathbf{x}, t|\mathbf{x}_0, t_0)]}{\partial t} - \nabla_{\mathbf{x}} \cdot [\mathbb{K} \cdot \nabla_{\mathbf{x}} G(\mathbf{x}, t|\mathbf{x}_0, t_0)] = \delta(t - t_0) \delta(\mathbf{x} - \mathbf{x}_0) \quad \mathbf{x} \in \mathcal{R} \quad (4.117a)$$

$$\hat{\mathbf{n}} \cdot \mathbb{K} \cdot \nabla_{\mathbf{x}} G(\mathbf{x}, t|\mathbf{x}_0, t_0) = 0 \quad \mathbf{x} \in \partial\mathcal{R} \quad (4.117b)$$

$$G(\mathbf{x}, t|\mathbf{x}_0, t_0) = 0 \quad \mathbf{x} \in \mathcal{R}, \quad t < t_0. \quad (4.117c)$$

The Neumann boundary condition is relevant when one knows the boundary flux of heat or tracer concentration. Alternatively, Dirichlet conditions are used when knowing the boundary values for the field, in which case the corresponding causal Green's function satisfies

$$\frac{\partial[G(\mathbf{x}, t|\mathbf{x}_0, t_0)]}{\partial t} - \nabla_{\mathbf{x}} \cdot [\mathbb{K} \cdot \nabla_{\mathbf{x}} G(\mathbf{x}, t|\mathbf{x}_0, t_0)] = \delta(t - t_0) \delta(\mathbf{x} - \mathbf{x}_0) \quad \mathbf{x} \in \mathcal{R} \quad (4.118a)$$

$$G(\mathbf{x}, t|\mathbf{x}_0, t_0) = 0 \quad \mathbf{x} \in \partial\mathcal{R} \quad (4.118b)$$

$$G(\mathbf{x}, t|\mathbf{x}_0, t_0) = 0 \quad \mathbf{x} \in \mathcal{R}, \quad t < t_0. \quad (4.118c)$$

The Dirichlet condition is particularly relevant for passive tracers in the atmosphere and ocean, such as those studied in Section 48.10. In both the Neumann and Dirichlet cases, we generalized the free space solution from Section 4.7.1 by introducing a space-time dependent diffusivity tensor, $\mathbb{K}(\mathbf{x}, t)$, which is a symmetric second order tensor with dimensions $L^2 T^{-1}$ (see Chapter 48 for discussion). The Green's function has dimensions of inverse volume, L^{-3} , which is implied since the Dirac delta source has dimensions of inverse volume times inverse time, $L^{-3} T^{-1}$.

In the presence of boundaries, the spatial position of the Dirac delta source impacts the value of the causal Green's function. In contrast, the causality condition (4.117c) means that the Green's function is dependent only on the time since the introduction of the source, $t - t_0$. Hence, there is no added generality afforded by setting the source time, t_0 , to be distinct from $t_0 = 0$. Even so, we retain t_0 to maintain symmetry with the spatial location \mathbf{x}_0 . Doing so also helps to distinguish the Dirac source time, t_0 , from the initial time, t_{init} , with the intitial time introduced in Section 4.7.5.

4.7.3 Adjoint causal Green's function

In using Green's function methods for solving initial-boundary value problems for the diffusion equation, we will make use of the *adjoint causal Green's function*. The adjoint causal Green's function is defined for $t \in (-\infty, \infty)$ and satisfies the following boundary value problem for the Neumann conditions, with these equations representing the adjoint to the Green's function equations (4.117a)-(4.117c)

$$-\frac{\partial[\tilde{G}(\mathbf{x}, t|\mathbf{x}_0, t_0)]}{\partial t} - \nabla_{\mathbf{x}} \cdot [\mathbb{K} \cdot \nabla_{\mathbf{x}} \tilde{G}(\mathbf{x}, t|\mathbf{x}_0, t_0)] = \delta(t - t_0) \delta(\mathbf{x} - \mathbf{x}_0) \quad \mathbf{x} \in \mathcal{R} \quad (4.119a)$$

$$\hat{\mathbf{n}} \cdot \mathbb{K} \cdot \nabla_{\mathbf{x}} \tilde{G}(\mathbf{x}, t|\mathbf{x}_0, t_0) = 0 \quad \mathbf{x} \in \partial\mathcal{R} \quad (4.119b)$$

$$\tilde{G}(\mathbf{x}, t|\mathbf{x}_0, t_0) = 0 \quad \mathbf{x} \in \mathcal{R}, \quad t > t_0. \quad (4.119c)$$

Similarly, the adjoint Green's function satisfying Dirichlet boundary conditions is determined by

$$-\frac{\partial[\tilde{G}(\mathbf{x}, t|\mathbf{x}_0, t_0)]}{\partial t} - \nabla_{\mathbf{x}} \cdot [\mathbb{K} \cdot \nabla_{\mathbf{x}} \tilde{G}(\mathbf{x}, t|\mathbf{x}_0, t_0)] = \delta(t - t_0) \delta(\mathbf{x} - \mathbf{x}_0) \quad \mathbf{x} \in \mathcal{R} \quad (4.120a)$$

$$\tilde{G}(\mathbf{x}, t|\mathbf{x}_0, t_0) = 0 \quad \mathbf{x} \in \partial\mathcal{R} \quad (4.120b)$$

$$\tilde{G}(\mathbf{x}, t|\mathbf{x}_0, t_0) = 0 \quad \mathbf{x} \in \mathcal{R}, \quad t > t_0. \quad (4.120c)$$

Note the sign change on the time derivative in equations (4.119a) and (4.120a) relative to equations (4.117a) and (4.118a). This change results since the single partial time derivative is not a self-adjoint operator, reflecting the lack of time symmetry of the diffusion equation (i.e., the diffusion equation distinguishes between past and future). Also note the backward causal condition, equations (4.119c) and (4.120c). Pondering the physics of these equations, we propose that the adjoint Green's function provides a solution to the *concentration equation*, which is the diffusion equation run backwards in time.

4.7.4 Reciprocity of the Green's function and its adjoint

When studying Poisson's equation we made use of reciprocity (4.75) satisfied by the Poisson equation Green's function, and we will require a corresponding reciprocity for the diffusion equation Green's function. Deriving reciprocity requires a bit more work for the diffusion equation due to the added time derivative term, which renders the adjoint diffusion operator distinct from the diffusion operator. That is, the diffusion operator is not *self-adjoint* due to sign change on the time derivative, whereas the Laplacian operator is self-adjoint (even with the symmetric diffusion tensor).

Before starting this derivation, note that we did not derive the adjoint in Section 4.7.3, instead we merely wrote it down. However, introduction of the adjoint Green's function is largely motivated by the following derivation of the reciprocity relation, where we see that the Green's function for the diffusion equation satisfies a reciprocity relation with the adjoint Green's function. Hence, as part of the following derivation we indirectly see how to construct the adjoint problem.

Setting up the derivation

To derive reciprocity, consider the partial differential equation (4.117a) with a Dirac delta source $\delta(t - t_1) \delta(\mathbf{x} - \mathbf{x}_1)$, and the adjoint partial differential equation (4.119a) with a distinct Dirac delta source $\delta(t - t_2) \delta(\mathbf{x} - \mathbf{x}_2)$. Multiply each of these equations by the complement Green's function and subtract

$$\begin{aligned} & \tilde{G}(\mathbf{x}, t | \mathbf{x}_2, t_2) \left[\frac{\partial G(\mathbf{x}, t | \mathbf{x}_1, t_1)}{\partial t} - \nabla_{\mathbf{x}} \cdot [\mathbb{K} \cdot \nabla_{\mathbf{x}} G(\mathbf{x}, t | \mathbf{x}_1, t_1)] \right] \\ & - G(\mathbf{x}, t | \mathbf{x}_1, t_1) \left[-\frac{\partial \tilde{G}(\mathbf{x}, t | \mathbf{x}_2, t_2)}{\partial t} - \nabla_{\mathbf{x}} \cdot [\mathbb{K} \cdot \nabla_{\mathbf{x}} \tilde{G}(\mathbf{x}, t | \mathbf{x}_2, t_2)] \right] \\ & = \tilde{G}(\mathbf{x}, t | \mathbf{x}_2, t_2) \delta(t - t_1) \delta(\mathbf{x} - \mathbf{x}_1) - G(\mathbf{x}, t | \mathbf{x}_1, t_1) \delta(t - t_2) \delta(\mathbf{x} - \mathbf{x}_2). \end{aligned} \quad (4.121)$$

The Dirac delta source locations, \mathbf{x}_1 and \mathbf{x}_2 , are arbitrary, so long as they are within the domain \mathcal{R} . Likewise, the source times, t_1 and t_2 , are arbitrary. In the following, we find it useful to introduce an arbitrarily large time, T , so that

$$-T < t_1, t_2 < T, \quad (4.122)$$

with T later dropping out from the results.

Integration and use of the sifting property

An integral of the right hand side of equation (4.121) over the domain \mathcal{R} and over time, both for the observational space-time points (\mathbf{x}, t) , leads to

$$\begin{aligned} \int_{-T}^T \int_{\mathcal{R}} & \left[\tilde{G}(\mathbf{x}, t | \mathbf{x}_2, t_2) \delta(t - t_1) \delta(\mathbf{x} - \mathbf{x}_1) - G(\mathbf{x}, t | \mathbf{x}_1, t_1) \delta(t - t_2) \delta(\mathbf{x} - \mathbf{x}_2) \right] dV dt \\ &= \tilde{G}(\mathbf{x}_1, t_1 | \mathbf{x}_2, t_2) - G(\mathbf{x}_2, t_2 | \mathbf{x}_1, t_1). \end{aligned} \quad (4.123)$$

As shown in the following, the same integral of the left hand side of equation (4.121) vanishes, which then establishes the reciprocity property between the Green's function and the adjoint Green's function.

Moving the time derivative from G to \tilde{G} and picking up a minus sign

The left hand side of equation (4.121) requires us to massage just the first term since, as we will show, this term equals to the second so that the left hand side of equation (4.121) vanishes. To prove this assertion, start by examining the time derivative. Since the spatial domain, \mathcal{R} , is assumed to be static, we can swap the time and space derivatives to find

$$\begin{aligned} & \int_{-T}^T \tilde{G}(\mathbf{x}, t | \mathbf{x}_2, t_2) \partial_t G(\mathbf{x}, t | \mathbf{x}_1, t_1) dt \\ &= \int_{-T}^T \left[\partial_t \left(\tilde{G}(\mathbf{x}, t | \mathbf{x}_2, t_2) G(\mathbf{x}, t | \mathbf{x}_1, t_1) \right) - \partial_t \tilde{G}(\mathbf{x}, t | \mathbf{x}_2, t_2) G(\mathbf{x}, t | \mathbf{x}_1, t_1) \right] dt. \end{aligned} \quad (4.124)$$

We now set

$$G(\mathbf{x}, t = -T | \mathbf{x}_1, t_1) = 0 \quad \text{and} \quad \tilde{G}(\mathbf{x}, t = +T | \mathbf{x}_2, t_2) = 0, \quad (4.125)$$

which result from the causality conditions (4.117c) and (4.119c). Hence, in moving the time derivative from G to \tilde{G} we pick up a minus sign, which means that the time derivative is not self-adjoint

$$\int_{-T}^T \tilde{G}(\mathbf{x}, t | \mathbf{x}_2, t_2) \partial_t G(\mathbf{x}, t | \mathbf{x}_1, t_1) dt = - \int_{-T}^T \partial_t \tilde{G}(\mathbf{x}, t | \mathbf{x}_2, t_2) G(\mathbf{x}, t | \mathbf{x}_1, t_1) dt. \quad (4.126)$$

The Laplacian is self-adjoint even with a symmetric diffusion tensor

Consider next the spatial derivative term on the left hand side of equation (4.121)

$$\begin{aligned} & - \int_{\mathcal{R}} \tilde{G}(\mathbf{x}, t | \mathbf{x}_2, t_2) \nabla_{\mathbf{x}} \cdot [\mathbb{K} \cdot \nabla_{\mathbf{x}} G(\mathbf{x}, t | \mathbf{x}_1, t_1)] dV = \\ & \int_{\mathcal{R}} \left[-\nabla_{\mathbf{x}} \cdot \left(\tilde{G}(\mathbf{x}, t | \mathbf{x}_2, t_2) \mathbb{K} \cdot \nabla_{\mathbf{x}} G(\mathbf{x}, t | \mathbf{x}_1, t_1) \right) + \nabla_{\mathbf{x}} \tilde{G}(\mathbf{x}, t | \mathbf{x}_2, t_2) \cdot \mathbb{K} \cdot \nabla_{\mathbf{x}} G(\mathbf{x}, t | \mathbf{x}_1, t_1) \right] dV. \end{aligned} \quad (4.127)$$

Use of the divergence theorem and either the homogeneous Neumann boundary condition (4.117b) or homogeneous Dirichlet condition (4.118b) allow us to drop the total derivative term. The same manipulation, with either the Neumann condition (4.119b) or Dirichlet condition (4.120b) satisfied by the adjoint Green's function $\tilde{G}(\mathbf{x}, t | \mathbf{x}_2, t_2)$, allows us to seamlessly move the Laplacian operator

from $G(\mathbf{x}, t | \mathbf{x}_1, t_1)$ onto $\tilde{G}(\mathbf{x}, t | \mathbf{x}_2, t_2)$, thus manifesting the self-adjoint nature of the Laplacian operator even in the presence of a symmetric diffusion tensor

$$\begin{aligned} \int_{\mathcal{R}} \tilde{G}(\mathbf{x}, t | \mathbf{x}_2, t_2) \nabla_{\mathbf{x}} \cdot [\mathbb{K} \cdot \nabla_{\mathbf{x}} G(\mathbf{x}, t | \mathbf{x}_1, t_1)] dV \\ = \int_{\mathcal{R}} \nabla_{\mathbf{x}} \cdot [\mathbb{K} \cdot \nabla_{\mathbf{x}} \tilde{G}(\mathbf{x}, t | \mathbf{x}_2, t_2)] G(\mathbf{x}, t | \mathbf{x}_1, t_1) dV. \end{aligned} \quad (4.128)$$

Reciprocity of the Green's function and the adjoint Green's function

The above manipulations show that the space-time integral for the left hand side of equation (4.121) vanishes. Consequently, we are left with the diffusion equation reciprocity

$$\tilde{G}(\mathbf{x}, t | \mathbf{x}_0, t_0) = G(\mathbf{x}_0, t_0 | \mathbf{x}, t). \quad (4.129)$$

In words, this identity means that the adjoint Green's function for the diffusion equation equals to the Green's function after swapping the space-time points for the field and source. Quite conveniently, reciprocity in the form of equation (4.129) means we have no need to explicitly solve the adjoint Green's function equations for $\tilde{G}(\mathbf{x}, t | \mathbf{x}_0, t_0)$. Instead, it is sufficient to determine the Green's function $G(\mathbf{x}, t | \mathbf{x}_0, t_0)$ and then use reciprocity to determine $\tilde{G}(\mathbf{x}, t | \mathbf{x}_0, t_0)$. The simplicity of the reciprocity relation (4.129) is central to the practical use of the Green's function method for the diffusion equation.

4.7.5 Integral solution

Having established reciprocity (4.129), we are now ready to derive an integral expression for the field, $\psi(\mathbf{x}, t)$, satisfying the diffusion equation initial-boundary value problem with either the Neumann boundary conditions

$$\frac{\partial \psi(\mathbf{x}, t)}{\partial t} - \nabla \cdot [\mathbb{K} \cdot \nabla_{\mathbf{x}} \psi(\mathbf{x}, t)] = \Lambda(\mathbf{x}, t) \quad \mathbf{x} \in \mathcal{R} \quad (4.130a)$$

$$\hat{\mathbf{n}} \cdot \mathbb{K} \cdot \nabla_{\mathbf{x}} \psi(\mathbf{x}, t) = \Sigma(\mathbf{x}, t) \quad \mathbf{x} \in \partial \mathcal{R} \quad (4.130b)$$

$$\psi(\mathbf{x}, t = t_{\text{init}}) = I(\mathbf{x}) \quad \mathbf{x} \in \mathcal{R}, \quad (4.130c)$$

or Dirichlet boundary conditions

$$\frac{\partial \psi(\mathbf{x}, t)}{\partial t} - \nabla \cdot [\mathbb{K} \cdot \nabla_{\mathbf{x}} \psi(\mathbf{x}, t)] = \Lambda(\mathbf{x}, t) \quad \mathbf{x} \in \mathcal{R} \quad (4.131a)$$

$$\psi(\mathbf{x}, t) = \sigma(\mathbf{x}, t) \quad \mathbf{x} \in \partial \mathcal{R} \quad (4.131b)$$

$$\psi(\mathbf{x}, t = t_{\text{init}}) = I(\mathbf{x}) \quad \mathbf{x} \in \mathcal{R}. \quad (4.131c)$$

In these equations we introduced the initial time, $t = t_{\text{init}}$, which is distinct from the Dirac delta source time, $t = t_0$. Correspondingly, the Dirac delta source is fired *after* the initial time,

$$t_{\text{init}} < t_0, \quad (4.132)$$

which follows since we are interested in the evolution of ψ after specification of the initial data, $\psi(\mathbf{x}, t = t_{\text{init}}) = I(\mathbf{x})$. Correspondingly, t_{init} defines the lower limit on time integrals in the following. Use of the $\nabla_{\mathbf{x}}$ notation is not needed for these equations, since there is no source point, \mathbf{x}_0 , in any of the expressions. However, this added detail helps us remain organized during the following

manipulations. Finally, note that the diffusion tensor appearing in these equations means that the source function, $\Lambda(\mathbf{x}, t)$, in equations (4.130a) and (4.131a), as well as the boundary data, $\Sigma(\mathbf{x}, t)$, in equation (4.130b), have different dimensions from their counterparts found in the Poisson boundary value problems (4.65) and (4.99).

The following derivation emulates that for the Poisson equation in Section 4.5.5, yet with distinct features arising from time evolution and the corresponding need to use the adjoint causal Green's function, \tilde{G} . We expose many details as doing so reveals general notions and tricks arising with Green's function methods for initial-boundary value problems.

Setting up the derivation

To start the derivation, multiply the diffusion equation (4.130a) by the adjoint Green's function, $\tilde{G}(\mathbf{x}, t|\mathbf{x}_0, t_0)$, and the adjoint Green's function equation (4.119a) by $\psi(\mathbf{x}, t)$. Subtracting and rearranging leads to

$$\partial_t(\tilde{G}\psi) + \nabla \cdot [\psi \mathbb{K} \cdot \nabla_{\mathbf{x}} \tilde{G} - \tilde{G} \mathbb{K} \cdot \nabla \psi] = \tilde{G}(\mathbf{x}, t|\mathbf{x}_0, t_0) \Lambda(\mathbf{x}, t) - \psi(\mathbf{x}, t) \delta(\mathbf{x} - \mathbf{x}_0) \delta(t - t_0), \quad (4.133)$$

where we temporarily suppressed arguments on the left hand side for brevity. Since the spatial domain is assumed to be static, we can integrate this equation over space and time without concern for the order of integration.

Time integration

A time integral of the first left hand side term in equation (4.133) leads to

$$\int_{t_{\text{init}}}^T \partial_t[\tilde{G}(\mathbf{x}, t|\mathbf{x}_0, t_0) \psi(\mathbf{x}, t)] dt = \tilde{G}(\mathbf{x}, T|\mathbf{x}_0, t_0) \psi(\mathbf{x}, T) - \tilde{G}(\mathbf{x}, t_{\text{init}}|\mathbf{x}_0, t_0) \psi(\mathbf{x}, t_{\text{init}}) \quad (4.134a)$$

$$= -\tilde{G}(\mathbf{x}, t_{\text{init}}|\mathbf{x}_0, t_0) I(\mathbf{x}), \quad (4.134b)$$

where we made use of the backward causal condition (4.119c) satisfied by the adjoint Green's function to set $\tilde{G}(\mathbf{x}, t = T|\mathbf{x}_0, t_0) = 0$, and used the initial condition (4.130c) to introduce the initial value data, $\psi(\mathbf{x}, t_{\text{init}}) = I(\mathbf{x})$.

Space integration

A space integral over all observation points, and use of the divergence theorem, brings the divergence term on the left side of equation (4.133) into

$$\begin{aligned} & \int_{\mathcal{R}} \nabla \cdot [\psi \mathbb{K} \cdot \nabla_{\mathbf{x}} \tilde{G} - \tilde{G} \mathbb{K} \cdot \nabla_{\mathbf{x}} \psi] dV \\ &= \oint_{\partial\mathcal{R}} \left[\psi(\mathbf{x}, t) \mathbb{K} \cdot \nabla_{\mathbf{x}} \tilde{G}(\mathbf{x}, t|\mathbf{x}_0, t_0) - \tilde{G}(\mathbf{x}, t|\mathbf{x}_0, t_0) \mathbb{K} \cdot \nabla_{\mathbf{x}} \psi \right] \cdot \hat{\mathbf{n}} d\mathcal{S}, \end{aligned} \quad (4.135)$$

where we kept both boundary terms pending specification of whether the fields satisfy Dirichlet or Neumann conditions. This boundary integral has the same appearance as found for the Poisson equation in Section 4.5.5, with the added feature here of the diffusion tensor.

Integrating the right hand side of equation (4.133)

A space and time integral for the right hand side of equation (4.133), along with the sifting properties of the Dirac delta, render

$$\begin{aligned} \int_{t_{\text{init}}}^T \left[\int_{\mathcal{R}} [\tilde{G}(\mathbf{x}, t | \mathbf{x}_0, t_0) \Lambda(\mathbf{x}, t) - \psi(\mathbf{x}, t) \delta(\mathbf{x} - \mathbf{x}_0) \delta(t - t_0)] dV \right] dt \\ = -\psi(\mathbf{x}_0, t_0) + \int_{t_{\text{init}}}^T \left[\int_{\mathcal{R}} \tilde{G}(\mathbf{x}, t | \mathbf{x}_0, t_0) \Lambda(\mathbf{x}, t) dV \right] dt. \quad (4.136) \end{aligned}$$

Notice how this integration over the observation space-time points, (\mathbf{x}, t) , serves to pick out the field, ψ , evaluated at the space-time point, (\mathbf{x}_0, t_0) , where the Dirac delta source is located.

Rearrangement and use of reciprocity

Bringing the above results together leads to the expression

$$\begin{aligned} \psi(\mathbf{x}_0, t_0) &= \int_{\mathcal{R}} \tilde{G}(\mathbf{x}, t_{\text{init}} | \mathbf{x}_0, t_0) I(\mathbf{x}) dV + \int_{t_{\text{init}}}^{t_0} \left[\int_{\mathcal{R}} \tilde{G}(\mathbf{x}, t | \mathbf{x}_0, t_0) \Lambda(\mathbf{x}, t) dV \right] dt \\ &+ \int_{t_{\text{init}}}^{t_0} \left[\oint_{\partial\mathcal{R}} [\tilde{G}(\mathbf{x}, t | \mathbf{x}_0, t_0) \mathbb{K} \cdot \nabla_{\mathbf{x}} \psi(\mathbf{x}, t) - \psi(\mathbf{x}, t) \mathbb{K} \cdot \nabla_{\mathbf{x}} \tilde{G}(\mathbf{x}, t | \mathbf{x}_0, t_0)] \cdot \hat{\mathbf{n}} d\mathcal{S} \right] dt. \quad (4.137) \end{aligned}$$

The time integrals are restricted to the range $t \in [t_{\text{init}}, t_0]$ through use of the causality condition (4.119c) for the adjoint Green's function. Hence, the arbitrary time, T , drops out from the solution and there is no dependence on fields at times later than t_0 nor before t_{init} .

Use of reciprocity (4.129) allows us to replace the adjoint Green's function with the Green's function to thus bring equation (4.137) to

$$\begin{aligned} \psi(\mathbf{x}_0, t_0) &= \int_{\mathcal{R}} G(\mathbf{x}_0, t_0 | \mathbf{x}, t_{\text{init}}) I(\mathbf{x}) dV + \int_{t_{\text{init}}}^{t_0} \left[\int_{\mathcal{R}} G(\mathbf{x}_0, t_0 | \mathbf{x}, t) \Lambda(\mathbf{x}, t) dV \right] dt \\ &+ \int_{t_{\text{init}}}^{t_0} \left[\oint_{\partial\mathcal{R}} [G(\mathbf{x}_0, t_0 | \mathbf{x}, t) \mathbb{K} \cdot \nabla_{\mathbf{x}} \psi(\mathbf{x}, t) - \psi(\mathbf{x}, t) \mathbb{K} \cdot \nabla_{\mathbf{x}} G(\mathbf{x}_0, t_0 | \mathbf{x}, t)] \cdot \hat{\mathbf{n}} d\mathcal{S} \right] dt. \quad (4.138) \end{aligned}$$

Finally, it is convenient to relabel $(\mathbf{x}_0, t_0) \leftrightarrow (\mathbf{x}, t)$ to write

$$\begin{aligned} \psi(\mathbf{x}, t) &= \underbrace{\int_{\mathcal{R}} G(\mathbf{x}, t | \mathbf{x}_0, t_{\text{init}}) I(\mathbf{x}_0) dV_0}_{\text{space integral of } G \text{ with } I \text{ on } \mathcal{R}} + \underbrace{\int_{t_{\text{init}}}^t \left[\int_{\mathcal{R}} G(\mathbf{x}, t | \mathbf{x}_0, t_0) \Lambda(\mathbf{x}_0, t_0) dV_0 \right] dt_0}_{\text{space-time integral of } G \text{ with } \Lambda \text{ over } \mathcal{R}} \\ &+ \underbrace{\int_{t_{\text{init}}}^t \left[\oint_{\partial\mathcal{R}} [G(\mathbf{x}, t | \mathbf{x}_0, t_0) \mathbb{K} \cdot \nabla_{\mathbf{x}_0} \psi(\mathbf{x}_0, t_0) - \psi(\mathbf{x}_0, t_0) \mathbb{K} \cdot \nabla_{\mathbf{x}_0} G(\mathbf{x}, t | \mathbf{x}_0, t_0)] \cdot \hat{\mathbf{n}}_{\mathbf{x}_0} d\mathcal{S}_0 \right] dt_0}_{\text{space-time integral over } \partial\mathcal{R}}. \quad (4.139) \end{aligned}$$

Specializing to Neumann boundary conditions leads to

$$\begin{aligned} \psi^{\text{Neumann}}(\mathbf{x}, t) &= \int_{\mathcal{R}} G(\mathbf{x}, t | \mathbf{x}_0, t_{\text{init}}) I(\mathbf{x}_0) dV_0 + \int_{t_{\text{init}}}^t \left[\int_{\mathcal{R}} G(\mathbf{x}, t | \mathbf{x}_0, t_0) \Lambda(\mathbf{x}_0, t_0) dV_0 \right] dt_0 \\ &+ \int_{t_{\text{init}}}^t \left[\oint_{\partial\mathcal{R}} G(\mathbf{x}, t | \mathbf{x}_0, t_0) \Sigma(\mathbf{x}_0, t_0) d\mathcal{S}_0 \right] dt_0, \quad (4.140) \end{aligned}$$

whereas the solution with Dirichlet conditions is

$$\begin{aligned}\psi^{\text{Dirichlet}}(\mathbf{x}, t) = & \int_{\mathcal{R}} G(\mathbf{x}, t | \mathbf{x}_0, t_{\text{init}}) I(\mathbf{x}_0) dV_0 + \int_{t_{\text{init}}}^t \left[\int_{\mathcal{R}} G(\mathbf{x}, t | \mathbf{x}_0, t_0) \Lambda(\mathbf{x}_0, t_0) dV_0 \right] dt_0 \\ & - \int_{t_{\text{init}}}^t \left[\oint_{\partial\mathcal{R}} \sigma(\mathbf{x}_0, t_0) \mathbb{K}(\mathbf{x}_0, t_0) \cdot \nabla_{\mathbf{x}_0} G(\mathbf{x}, t | \mathbf{x}_0, t_0) \cdot \hat{\mathbf{n}}_{\mathbf{x}_0} dS_0 \right] dt_0.\end{aligned}\quad (4.141)$$

4.7.6 Properties of the solution

Many of the properties of the solution (4.139) are also reflected in the Poisson equation solutions from Sections 4.5.6 and 4.6.4. In particular, the solution manifests the linear superposition principle, with the solution given by the sum of three terms arising from the initial conditions, distributed volume source, and spatial boundary conditions. We expect to have this connection given that the steady state diffusion equation satisfies a generalized Poisson equation (generalized by the presence of a diffusion tensor). A fundamentally new piece of physics and maths arises from time dependence. In this section we exhibit properties of the Green's function as inferred (through insisting on self-consistency) by the initial and boundary conditions.

Uniqueness of the solution

Uniqueness of a solution to the diffusion boundary value problem follows as in the discussion of the Poisson equation in Sections 4.5.2 and 4.6.2. Namely, consider two solutions to the diffusion equation and take their difference, $\Psi = \psi_A - \psi_B$. We readily see that Ψ satisfies the homogeneous diffusion equation with homogeneous boundary conditions along with a zero initial condition. Ψ thus remains zero for both the Dirichlet and Neumann cases, thus proving that the solution to both problems is unique.

Initial conditions

By sampling the solution (4.139) as time decreases towards the initial time, $t \rightarrow t_{\text{init}}$, and noting the initial condition $\psi(\mathbf{x}, t_{\text{init}}) = I(\mathbf{x})$, we are led to¹¹

$$\lim_{t \rightarrow t_{\text{init}}} \psi(\mathbf{x}, t) = I(\mathbf{x}) = \lim_{t \rightarrow t_{\text{init}}} \int_{\mathcal{R}} G(\mathbf{x}, t | \mathbf{x}_0, t_{\text{init}}) I(\mathbf{x}_0) dV_0.\quad (4.142)$$

This temporal sampling of the field time is distinguished from the source time, t_0 , which here is fixed at the initial time, $t_0 = t_{\text{init}}$. Self-consistency in equation (4.142) implies that the Green's function for both Neumann and Dirichlet boundary conditions satisfies the initial condition

$$\lim_{t \rightarrow t_{\text{init}}} G(\mathbf{x}, t | \mathbf{x}_0, t_{\text{init}}) = \delta(\mathbf{x} - \mathbf{x}_0) \quad \text{with } \mathbf{x}, \mathbf{x}_0 \in \mathcal{R}.\quad (4.143)$$

That is, the Green's function is initialized by a Dirac delta pulse at the source point, \mathbf{x}_0 , which then leads to

$$\lim_{t \rightarrow t_{\text{init}}} \int_{\mathcal{R}} G(\mathbf{x}, t | \mathbf{x}_0, t_{\text{init}}) I(\mathbf{x}_0) dV_0 = \int_{\mathcal{R}} \delta(\mathbf{x} - \mathbf{x}_0) I(\mathbf{x}_0) dV_0 = I(\mathbf{x}).\quad (4.144)$$

¹¹Since t_{init} is the initial time, the limit $t \rightarrow t_{\text{init}}$ is taken from above so that $t = t_{\text{init}} + \epsilon$ with $\epsilon \rightarrow 0$.

Neumann boundary conditions

Acting with $\mathbb{K}(\mathbf{x}, t) \cdot \nabla_{\mathbf{x}}$ on the Neumann solution (4.140); evaluating the expression on the boundary $\mathbf{x} = \mathbf{x}_{\partial\mathcal{R}} \in \partial\mathcal{R}$; and then projecting onto the outward normal $\hat{\mathbf{n}}_{\mathbf{x}}$, serves to annihilate the volume integrals as per the homogeneous Neumann conditions satisfied by the Green's function. We are thus left with

$$\hat{\mathbf{n}}_{\mathbf{x}} \cdot \mathbb{K}(\mathbf{x}, t) \cdot \nabla_{\mathbf{x}} \psi(\mathbf{x}, t) = \Sigma(\mathbf{x}, t) = \int_{t_{\text{init}}}^t \left[\oint_{\partial\mathcal{R}} \hat{\mathbf{n}}_{\mathbf{x}} \cdot \mathbb{K}(\mathbf{x}, t) \cdot \nabla_{\mathbf{x}} G(\mathbf{x}, t | \mathbf{x}_0, t_0) \Sigma(\mathbf{x}_0, t_0) d\mathcal{S}_0 \right] dt_0. \quad (4.145)$$

Self-consistency implies that the Green's function for the Neumann problem, when evaluated on the spatial boundary, satisfies

$$\hat{\mathbf{n}}_{\mathbf{x}} \cdot \mathbb{K}(\mathbf{x}, t) \cdot \nabla_{\mathbf{x}} G(\mathbf{x}, t | \mathbf{x}_0, t_0) = \delta(t - t_0) \delta^{(2)}(\mathbf{x} - \mathbf{x}_0) \quad \text{with } \mathbf{x}, \mathbf{x}_0 \in \partial\mathcal{R}, \quad (4.146)$$

which is a generalization of the property (4.113) holding for the Poisson equation Green's function.

Transforming the Neumann boundary data to the interior

In Section 4.6.4, we saw how to transform the Neumann boundary condition into the interior by modifying the source function. The diffusion equation Neumann solution (4.140) allows for the same transformation by writing

$$\psi^{\text{Neumann}}(\mathbf{x}, t) = \int_{\mathcal{R}} G(\mathbf{x}, t | \mathbf{x}_0, t_{\text{init}}) I(\mathbf{x}_0) dV_0 + \int_{t_{\text{init}}}^t \left[\int_{\mathcal{R}} G(\mathbf{x}, t | \mathbf{x}_0, t_0) \Lambda^*(\mathbf{x}_0, t_0) dV_0 \right] dt_0, \quad (4.147)$$

where the modified source function follows from that used for the Poisson equation (4.108)

$$\Lambda^*(\mathbf{x}_0, t_0) = \Lambda(\mathbf{x}_0, t_0) + \Sigma(\mathbf{x}_0, t_0) \delta[\hat{\mathbf{n}} \cdot (\mathbf{x}_0 - \mathbf{x}_{\partial\mathcal{R}})]. \quad (4.148)$$

Dirichlet boundary conditions

Evaluating the Dirichlet solution (4.141) on a spatial boundary, $\mathbf{x} = \mathbf{x}_{\partial\mathcal{R}} \in \partial\mathcal{R}$, eliminates both of the volume integrals so that we are left with

$$\psi^{\text{dirichlet}}(\mathbf{x}, t) = \sigma(\mathbf{x}, t) = - \int_{t_{\text{init}}}^t \left[\oint_{\partial\mathcal{R}} \sigma(\mathbf{x}_0, t_0) \mathbb{K} \cdot \nabla_{\mathbf{x}_0} G(\mathbf{x}, t | \mathbf{x}_0, t_0) \cdot \hat{\mathbf{n}}_{\mathbf{x}_0} d\mathcal{S}_0 \right] dt_0. \quad (4.149)$$

Self-consistency implies that the Green's function for the Dirichlet problem, when evaluated on the spatial boundary, satisfies

$$\hat{\mathbf{n}}_{\mathbf{x}_0} \cdot \mathbb{K}(\mathbf{x}_0, t_0) \cdot \nabla_{\mathbf{x}_0} G(\mathbf{x}, t | \mathbf{x}_0, t_0) = -\delta(t - t_0) \delta^{(2)}(\mathbf{x} - \mathbf{x}_0) \quad \text{with } \mathbf{x}, \mathbf{x}_0 \in \partial\mathcal{R}, \quad (4.150)$$

which is a generalization of the property (4.95) holding for the Poisson equation Green's function.

4.7.7 Boundary propagator

In Sections 4.5.6 and 4.5.7 we studied the boundary Green's function for the Poisson equation. Here we extend those ideas to the *boundary propagator* for the diffusion equation, with the boundary propagator mediating the transfer of Dirichlet boundary information into the interior. Boundary propagators for diffusion and advection-diffusion (Section 48.10) have extensive use in geophysical fluids given that many tracers have no interior sources.

Defining the boundary propagator

To focus on the role of the boundary propagator, consider a tracer in which the initial conditions and interior source both vanish: $I(\mathbf{x}) = 0$ and $\Lambda(\mathbf{x}, t) = 0$. Assuming Dirichlet boundary conditions, the initial-boundary value problem (4.130a)-(4.130c) simplifies to

$$\frac{\partial \psi(\mathbf{x}, t)}{\partial t} - \nabla_{\mathbf{x}} \cdot [\mathbb{K} \cdot \nabla_{\mathbf{x}} \psi(\mathbf{x}, t)] = 0 \quad \mathbf{x} \in \mathcal{R} \quad (4.151a)$$

$$\psi(\mathbf{x}, t) = \sigma(\mathbf{x}, t) \quad \mathbf{x} \in \partial\mathcal{R} \quad (4.151b)$$

$$\psi(\mathbf{x}, t = t_{\text{init}}) = 0 \quad \mathbf{x} \in \mathcal{R}, \quad (4.151c)$$

with the corresponding Dirichlet Green's function solution (4.141) taking the form

$$\psi(\mathbf{x}, t) = - \int_{t_{\text{init}}}^t \left[\oint_{\partial\mathcal{R}} \sigma(\mathbf{x}_0, t_0) \mathbb{K}(\mathbf{x}_0, t_0) \cdot \nabla_{\mathbf{x}_0} G(\mathbf{x}, t | \mathbf{x}_0, t_0) \cdot \hat{\mathbf{n}}_{\mathbf{x}_0} d\mathcal{S}_0 \right] dt_0. \quad (4.152)$$

We define the *boundary propagator* as the kernel in this equation

$$G^{\text{bp}}(\mathbf{x}, t | \mathbf{x}_0, t_0) \equiv -\mathbb{K}(\mathbf{x}_0, t_0) \cdot \nabla_{\mathbf{x}_0} G(\mathbf{x}, t | \mathbf{x}_0, t_0) \cdot \hat{\mathbf{n}}_{\mathbf{x}_0} \quad \text{with } \mathbf{x}_0 \in \partial\mathcal{R}, \quad (4.153)$$

with this definition giving G^{bp} the dimensions of $L^{-2} T^{-1}$. Use of the boundary propagator brings the solution (4.152) into the succinct form

$$\psi(\mathbf{x}, t) = \int_{t_{\text{init}}}^t \left[\oint_{\partial\mathcal{R}} \sigma(\mathbf{x}_0, t_0) G^{\text{bp}}(\mathbf{x}, t | \mathbf{x}_0, t_0) d\mathcal{S}_0 \right] dt_0. \quad (4.154)$$

Boundary value problem for the boundary propagator

If we know the Green's function, $G(\mathbf{x}, t | \mathbf{x}_0, t_0)$, then we can compute the boundary propagator through the definition (4.153). Alternatively, we can directly determine the boundary propagator by solving its boundary value problem. Following from the definition (4.153) and the boundary condition (4.150), we know that

$$G^{\text{bp}}(\mathbf{x}, t | \mathbf{x}_0, t_0) = \delta(t - t_0) \delta^{(2)}(\mathbf{x} - \mathbf{x}_0) \quad \text{with } \mathbf{x}, \mathbf{x}_0 \in \partial\mathcal{R}. \quad (4.155)$$

Hence, the boundary propagator, when evaluated along the boundary, is a Dirac delta source that fires at time $t = t_0$ at the location $\mathbf{x} = \mathbf{x}_0 \in \partial\mathcal{R}$. To determine the partial differential equation satisfied by the boundary propagator, make use of the solution (4.154) and compute its time derivative

$$\frac{\partial \psi(\mathbf{x}, t)}{\partial t} = \oint_{\partial\mathcal{R}} \sigma(\mathbf{x}_0, t) G^{\text{bp}}(\mathbf{x}, t | \mathbf{x}_0, t) d\mathcal{S}_0 + \int_{t_{\text{init}}}^t \left[\oint_{\partial\mathcal{R}} \sigma(\mathbf{x}_0, t_0) \frac{\partial G^{\text{bp}}(\mathbf{x}, t | \mathbf{x}_0, t_0)}{\partial t} d\mathcal{S}_0 \right] dt_0. \quad (4.156)$$

The first term vanishes since the boundary propagator satisfies causality just like the Green's function when sampled at interior points

$$G^{\text{bp}}(\mathbf{x}, t | \mathbf{x}_0, t_0) = 0 \quad \text{if } t \leq t_0 \text{ and } \mathbf{x} \notin \partial\mathcal{R}. \quad (4.157)$$

We are thus left with

$$\begin{aligned} & \frac{\partial \psi(\mathbf{x}, t)}{\partial t} - \nabla_{\mathbf{x}} \cdot (\mathbb{K} \cdot \nabla_{\mathbf{x}} \psi) \\ &= \int_{t_{\text{init}}}^t \left[\oint_{\partial\mathcal{R}} \sigma(\mathbf{x}_0, t_0) \left[\frac{\partial G^{\text{bp}}(\mathbf{x}, t | \mathbf{x}_0, t_0)}{\partial t} - \nabla_{\mathbf{x}} \cdot [\mathbb{K} \cdot \nabla_{\mathbf{x}} G^{\text{bp}}(\mathbf{x}, t | \mathbf{x}_0, t_0)] \right] d\mathcal{S}_0 \right] dt_0. \end{aligned} \quad (4.158)$$

Since the left hand side vanishes via the partial differential equation (4.151a), and since the boundary data, σ , is arbitrary, we are led to the causal boundary value problem for the boundary propagator

$$\frac{\partial G^{\text{bp}}(\mathbf{x}, t | \mathbf{x}_0, t_0)}{\partial t} - \nabla_{\mathbf{x}} \cdot [\mathbb{K} \cdot \nabla_{\mathbf{x}} G^{\text{bp}}(\mathbf{x}, t | \mathbf{x}_0, t_0)] = 0 \quad \mathbf{x} \in \mathcal{R} \quad (4.159a)$$

$$G^{\text{bp}}(\mathbf{x}, t | \mathbf{x}_0, t_0) = 0 \quad \mathbf{x} \notin \partial\mathcal{R}, t \leq t_0 \quad (4.159b)$$

$$G^{\text{bp}}(\mathbf{x}, t | \mathbf{x}_0, t_0) = \delta(t - t_0) \delta^{(2)}(\mathbf{x} - \mathbf{x}_0) \quad \mathbf{x}, \mathbf{x}_0 \in \partial\mathcal{R}. \quad (4.159c)$$

In words, we see that upon firing the Dirac delta source on the boundary at time $t = t_0$ and point $\mathbf{x} = \mathbf{x}_0 \in \partial\mathcal{R}$, the boundary propagator diffuses the Dirac source into the region interior. We thus see that whereas the Dirichlet Green's function, $G(\mathbf{x}, t | \mathbf{x}_0, t_0)$, is zero along the boundary and yet feels the Dirac delta sources within the interior, the boundary propagator, $G^{\text{bp}}(\mathbf{x}, t | \mathbf{x}_0, t_0)$, places the Dirac delta sources just on the boundary and feels no sources within the interior. Just as the causality condition means that the Green's function is a function of $t - t_0$, so too is the boundary propagator. Furthermore, a focus on the boundary propagator rather than the Green's function allows us to dispense with the need to compute the normal gradient of the Green's function at the boundary, with that calculation rather awkward in practice.

Normalization of the boundary propagator

Consider the special case of a uniform constant Dirichlet boundary data, $\sigma = \sigma_{\text{constant}}$ in the solution (4.154). Diffusion will act on this constant boundary value to spread it throughout the region. After sufficient time the solution will reach a steady state whereby $\psi = \sigma_{\text{constant}}$ at every point within the domain. This result means that the boundary Green's function satisfies the normalization condition

$$\lim_{t_{\text{init}} \rightarrow -\infty} \int_{t_{\text{init}}}^t \left[\oint_{\partial\mathcal{R}} G^{\text{bp}}(\mathbf{x}, t | \mathbf{x}_0, t_0) d\mathcal{S}_0 \right] dt_0 = 1, \quad (4.160)$$

where the lower time limit is meant to indicate some arbitrary time sufficiently far in the past so that a steady state has been reached. Again, this normalization condition holds for every point within the domain and for any time. It corresponds to the normalization condition (4.98) satisfied by the boundary Green's function for the Poisson equation.

4.8 Initial value problems and response functions

In this section we focus on initial value problems and study the response functions that help to characterize a dynamical system. For this purpose, consider the first order ordinary differential equation

$$[d/dt + \lambda(t)] \psi(t) = F(t) \quad \text{with } \psi(t \leq t_{\text{init}}) = 0, \quad (4.161)$$

where ψ is some geophysical field, such as the anomalous sea surface temperature, λ is a feedback parameter that is positive for a damped system, and F is a forcing function such as that introduced by atmospheric variability on the surface ocean. We are only concerned with temporal behavior so that all spatial information is ignored. This equation has found widespread use in the climate dynamics community, largely following the work from [Hasselmann \(1976\)](#).

The material in this section could well have been used near the start of this chapter since it only involves ordinary differential equations. However, placing it at the end helps to tie together some of the material sprinkled throughout this chapter, and to connect to current applications of Green's function technology for studies of climate dynamics.

4.8.1 Impulse response function

Consider the system (4.161) with $\lambda > 0$ a time-independent feedback parameter damping the system back to zero, and with the forcing given by a Dirac delta

$$[\mathrm{d}/\mathrm{d}t + \lambda] G(t|t_0) = \alpha \delta(t - t_0) \quad \text{with } G(t|t_0) = 0 \text{ for } t < t_0, \quad (4.162)$$

where $\alpha > 0$ is a constant dimensionless scaling coefficient. We refer to the resulting causal Green's function, $G(t|t_0)$, as the *impulse response function* since it represents the response of the dynamical system to an impulse provided by the Dirac delta.¹²

Initial condition for the impulse response function

To determine the initial condition for the Green's function, integrate equation (4.162) over an interval containing the source time, t_0 , to render

$$\lim_{\epsilon \rightarrow 0} \left[G(t_0 + \epsilon|t_0) - G(t_0 - \epsilon|t_0) + \int_{t_0 - \epsilon}^{t_0 + \epsilon} \lambda G(t|t_0) \, \mathrm{d}t \right] = \alpha. \quad (4.163)$$

Causality means that $G(t_0 - \epsilon|t_0) = 0$ so that

$$\lim_{\epsilon \rightarrow 0} G(t_0 + \epsilon|t_0) + \lim_{\epsilon \rightarrow 0} \int_{t_0}^{t_0 + \epsilon} \lambda G(t|t_0) \, \mathrm{d}t = \alpha. \quad (4.164)$$

We assume that the integral vanishes in the limit of $\epsilon \rightarrow 0$, which is a sensible assumption since the only means to have a nonzero integral is if the Green's function had a singularity similar to a Dirac delta. We are thus led to the initial condition for the Green's function

$$G(t = t_0|t_0) = \alpha. \quad (4.165)$$

Solution for the impulse response function

The causality condition $G(t < t_0|t_0) = 0$ can be satisfied by introducing the Heaviside step function from Section 4.2.5

$$G(t|t_0) = \mathcal{H}(t - t_0) g(t) \quad \text{with} \quad [\mathrm{d}/\mathrm{d}t + \lambda] g = 0 \quad \text{and} \quad g(t = t_0) = \alpha, \quad (4.166)$$

and the solution is readily determined to be the damped exponential

$$G(t|t_0) = \mathcal{H}(t - t_0) \alpha e^{-\lambda(t-t_0)}. \quad (4.167)$$

We verify this function satisfies the initial value problem (4.162) by noting that

$$\mathrm{d}G(t|t_0)/\mathrm{d}t = \alpha \delta(t - t_0) e^{-\lambda(t-t_0)} - \lambda G(t|t_0) = \alpha \delta(t - t_0) - \lambda G(t|t_0) \quad (4.168)$$

as per the discussion leading to equation (4.15). As illustrated in Figure 4.4, the impulse response function (4.167) has a particularly simple interpretation as the damped exponential response of the dynamical system to a Dirac impulse fired at $t = t_0$.

¹²Recall our discussion of impulse in Section 4.2.7.

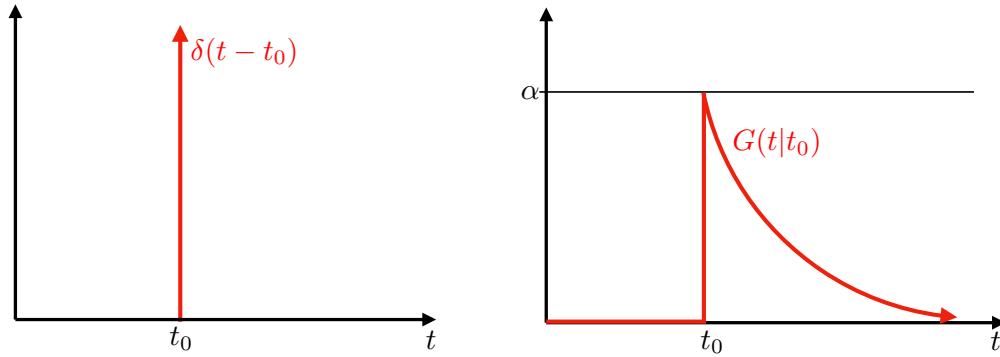


FIGURE 4.4: Left panel: Dirac delta that is fired at time $t = t_0$. Right panel: The impulse response function (4.167) resulting from the Dirac delta impulse as realized for the damped linear system (4.162).

4.8.2 Step response function

Rather than hit the system at a particular moment in time with a Dirac delta, we may choose to impose a force that turns on and remains on after some initial time, as per a Heaviside step function. The *step response function*, $S(t|t_0)$, measures the response of the dynamical system to this step forcing and it satisfies the differential equation

$$[d/dt + \lambda] S(t|t_0) = \alpha \mathcal{H}(t - t_0) \quad \text{with } S(t|t_0) = 0 \text{ for } t < t_0. \quad (4.169)$$

Note that in the steady state at $t \rightarrow \infty$, the step response function asymptotes to the constant

$$\lim_{t \rightarrow \infty} S(t|t_0) = \alpha/\lambda. \quad (4.170)$$

Connection to the impulse response function

The time derivative d/dt_0 acting on the step response function equation (4.169) leads to

$$[d/dt + \lambda] dS(t|t_0)/dt_0 = \alpha d\mathcal{H}(t - t_0)/dt_0. \quad (4.171)$$

The derivative of the Heaviside step function equals to the Dirac delta as per equation (4.19), in which

$$d\mathcal{H}(t - t_0)/dt_0 = -d\mathcal{H}(t - t_0)/dt = -\delta(t - t_0). \quad (4.172)$$

Use of this result in equation (4.171), and comparison to the impulse response function equation (4.162), yields the identity

$$\frac{dS(t|t_0)}{dt_0} = -G(t|t_0). \quad (4.173)$$

This identity holds even when the feedback parameter is a function of time, $\lambda = \lambda(t)$, since the time derivative operator, d/dt_0 , has no effect on $\lambda(t)$.

Initial condition for the step response function

To determine the initial condition for the step response function, integrate equation (4.169) over an interval bounding t_0 and take the limit as that interval vanishes

$$\lim_{\epsilon \rightarrow 0} \left[S(t_0 + \epsilon|t_0) - S(t_0 - \epsilon|t_0) + \int_{t_0-\epsilon}^{t_0+\epsilon} \lambda S(t|t_0) dt \right] = \int_{t_0-\epsilon}^{t_0+\epsilon} \mathcal{H}(t - t_0) dt. \quad (4.174)$$

Causality means that $S(t_0 - \epsilon|t_0) = 0$. Furthermore, the integral of the Heaviside is given by

$$\lim_{\epsilon \rightarrow 0} \int_{t_0-\epsilon}^{t_0+\epsilon} \mathcal{H}(t - t_0) dt = \lim_{\epsilon \rightarrow 0} \int_{t_0}^{t_0+\epsilon} \mathcal{H}(t - t_0) dt = \lim_{\epsilon \rightarrow 0} \epsilon = 0, \quad (4.175)$$

so that

$$\lim_{\epsilon \rightarrow 0} S(t_0 + \epsilon|t_0) = \epsilon \lambda \implies S(t = t_0|t_0) = 0. \quad (4.176)$$

That is, the step response function starts at zero and then grows in time in response to the Heaviside step function forcing.

Solution for the step response function

It is straightforward to show that the causal step response function is given by the saturating exponential

$$S(t|t_0) = \frac{\alpha}{\lambda} \left[1 - e^{-\lambda(t-t_0)} \right] \mathcal{H}(t - t_0). \quad (4.177)$$

Figure 4.5 depicts this function along with the Heaviside step forcing. Furthermore, we verify the connection between $S(t|t_0)$ and $G(t|t_0)$ by computing

$$dS(t|t_0)/dt_0 = -\alpha \mathcal{H}(t - t_0) e^{-\lambda(t-t_0)} - \frac{\alpha}{\lambda} \left[1 - e^{-\lambda(t-t_0)} \right] \delta(t - t_0) \quad (4.178a)$$

$$= -G(t|t_0) - \frac{\alpha}{\lambda} \left[1 - e^{-\lambda(t-t_0)} \right] \delta(t - t_0). \quad (4.178b)$$

The second term on the right hand side vanishes since

$$\int_{t_0-\epsilon}^{t_0+\epsilon} \left[1 - e^{-\lambda(t-t_0)} \right] \delta(t - t_0) dt = 0, \quad (4.179)$$

in which case we have the expected result $dS(t|t_0)/dt_0 = -G(t|t_0)$.

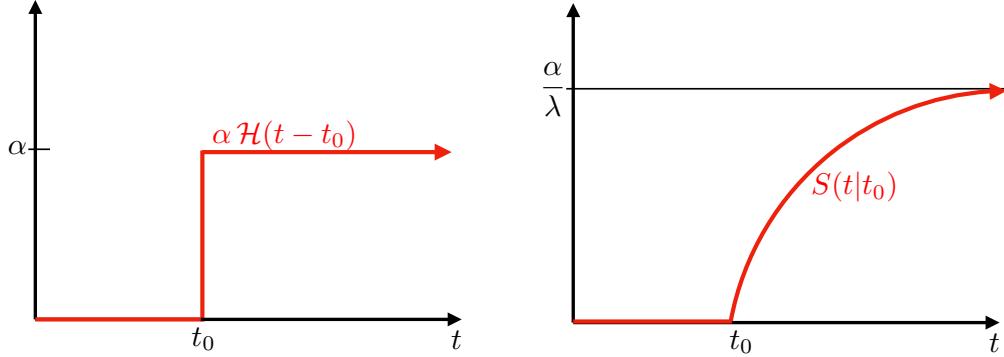


FIGURE 4.5: Left panel: Heaviside step function is fired at time $t = t_0$ and stays on afterward. Right panel: The step response function (4.177) resulting from the Heaviside step forcing impulse as realized for the damped linear system (4.162).

4.8.3 Reciprocity relation

The initial value problem (4.161) is not a self-adjoint system. Hence, we need to develop a reciprocity condition for the impulse response function and its adjoint, following the procedure used for the diffusion equation in Section 4.7.4. Again, the impulse response function satisfies

$$[d/dt + \lambda] G(t|t_1) = \alpha \delta(t - t_1) \quad \text{with } G(t|t_1) = 0 \text{ for } t < t_1 \text{ and } G(t_1|t_1) = \alpha, \quad (4.180)$$

and the adjoint impulse response function satisfies

$$[-d/dt + \lambda] \tilde{G}(t|t_2) = \alpha \delta(t - t_2) \quad \text{with } \tilde{G}(t|t_2) = 0 \text{ for } t > t_2 \text{ and } \tilde{G}(t_2|t_2) = \alpha. \quad (4.181)$$

We here introduced two Dirac delta source times, t_1, t_2 , which both occur after the initial time and before the end time

$$t_{\text{init}} < t_1, t_2 < T. \quad (4.182)$$

As we will see, causality eliminates the final time, T , from the solution for ψ . We retain it merely for bookeeping.

Determining the reciprocity relation between \tilde{G} and G follows by multiplying equation (4.180) by $\tilde{G}(t|t_2)$ and multiplying equation (4.181) by $G(t|t_1)$ and then subtracting

$$\frac{d}{dt} [G(t|t_1) \tilde{G}(t|t_2)] = \alpha [\tilde{G}(t|t_2) \delta(t - t_1) - G(t|t_1) \delta(t - t_2)]. \quad (4.183)$$

Now integrate this equation over the time range $t_{\text{init}} \leq t \leq T$. For the right hand side we assume α to be a constant, which then leads to the difference $\tilde{G}(t_1|t_2) - G(t_2|t_1)$. For the left hand side, use of the causality conditions in equations (4.180) and (4.181) render

$$\int_{t_{\text{init}}}^T \frac{d}{dt} [G(t|t_1) \tilde{G}(t|t_2)] dt = 0, \quad (4.184)$$

thus yielding the reciprocity relation

$$\tilde{G}(t|t_0) = G(t_0|t). \quad (4.185)$$

The feedback parameter, λ , dropped out from the derivation, with the reciprocity relation holding even if λ is time dependent. However, we again needed to assume the coefficient α to be constant.

4.8.4 Response function for general forcing

We now return to the initial value problem (4.161) and determine the response, ψ , to a general forcing function, $F(t)$, that turns on at some initial time $t = t_0 > t_{\text{init}}$. As for our earlier discussions of Green's functions, we express the general response function as an integral over impulse responses. For this purpose, multiply equation (4.161) by $\tilde{G}(t|t_0)$ and the adjoint equation (4.181) by $\psi(t)$, subtract, and then integrate to find

$$\alpha \psi(t_0) = \int_{t_{\text{init}}}^T \tilde{G}(t|t_0) F(t) dt - \int_{t_{\text{init}}}^T \frac{d}{dt} [\tilde{G}(t|t_0) \psi(t)] dt. \quad (4.186)$$

Making use of the causality condition $\tilde{G}(t|t_0) = 0$ for since $t > t_0$ leads to

$$\alpha \psi(t_0) = \tilde{G}(t_{\text{init}}|t_0) \psi(t_{\text{init}}) + \int_{t_{\text{init}}}^{t_0} \tilde{G}(t|t_0) F(t) dt, \quad (4.187)$$

where we retained the possibility of $\psi(t_{\text{init}}) \neq 0$ for a bit of generality. The reciprocity condition (4.185) brings this equation to the form

$$\alpha \psi(t_0) = G(t_0|t_{\text{init}}) \psi(t_{\text{init}}) + \int_{t_{\text{init}}}^{t_0} G(t_0|t) F(t) dt, \quad (4.188)$$

and swapping symbols, $t \leftrightarrow t_0$, yields

$$\alpha \psi(t) = G(t|t_{\text{init}}) \psi(t_{\text{init}}) + \int_{t_{\text{init}}}^t G(t|t_0) F(t_0) dt_0. \quad (4.189)$$

As anticipated, the general response is written as an initial response plus the integral of the general forcing with the impulse response function. Causality ensures that ψ is dependent only on forcing that is active between the initial time, t_{init} , and current time, t . To garner further insights into the general expression (4.189), consider the special case of constant feedback parameter, λ , in which the impulse response function is (4.167) so that

$$\psi(t) = e^{-\lambda(t-t_{\text{init}})} \psi(t_{\text{init}}) + \int_{t_{\text{init}}}^t e^{-\lambda(t-t_0)} F(t_0) dt_0. \quad (4.190)$$

4.8.5 Connection to the boundary propagator

Recall our discussion in Section 4.7.7 of the boundary propagator for the diffusion equation, which solves the causal boundary value problem (4.159a)-(4.159c). Again, the boundary propagator, $G^{\text{bp}}(\mathbf{x}, t | \mathbf{x}_0, t_0)$, measures the response of the system at (\mathbf{x}, t) to a Dirac delta space-time source imposed along the surface boundary. The details of the diffusion process are encoded into the boundary propagator so that the propagator is able to build up the response, ψ , to a general boundary forcing function, σ , as per equation (4.154). The discussion in the current section thus prompts us to consider the boundary propagator as an impulse response function for spatially distributed sources whose influence is mediated by diffusion.

4.8.6 Comments and further study

Many applications of Green's function methods in geophysical fluid mechanics and climate dynamics do not make use of analytical methods to solve for the Green's function. Instead, they make use of numerical estimates based on time stepping passive tracers in ocean and atmospheric models, with source functions approximating Dirac delta sources. Many applications have focused on boundary propagators, which as shown in this section are equivalent to impulse response functions for boundary sources. We return to this point when discussing the passive tracer equation in Section 48.10.

[Hasselmann et al. \(1993\)](#) introduced the impulse response function and step response function to the study of climate model drift. [Marshall et al. \(2014\)](#) and [Zanna et al. \(2019\)](#) presented further studies using this framework. Some of the mathematical formulation of impulse response and step response functions as presented here follow that offered in Exercise 1.52 of [Stakgold \(2000a\)](#).



Curves and surfaces

We encounter curves and surfaces throughout the study of geophysical fluid mechanics, with fluid particle pathlines through space-time and isopycnal/isentropic surfaces providing two examples. Indeed, curves and surfaces are encountered throughout physics. Hence, there is a well developed mathematical framework to describe the geometric properties of these objects. Our goal in this chapter is to introduce some of the basics of curves and surfaces embedded in Euclidean space.

Although the curves and surfaces of geophysical fluid mechanics are commonly moving as part of the fluid flow, we are concerned in this chapter with describing their instantaneous spatial properties. Hence, time does not appear in this chapter. Furthermore, although curves and surfaces can overturn and intersect themselves, we restrict attention to orientable curves and surfaces whose normal direction has a nonzero projection onto the vertical; i.e., they have no overhangs and no wrapping (Figure 5.1). This constraint is satisfied by the surfaces of constant generalized vertical coordinates (e.g., isopycnal surfaces) considered in Chapter 43 and in many other places in this book. It allows us to make use of coordinates known as the *Monge gauge* in condensed matter physics

READER'S GUIDE TO THIS CHAPTER

This chapter requires an understanding of the Cartesian calculus of Chapter 2. The differential geometry presented here is of some use throughout this book, and yet the casual reader need not penetrate the material. Conversely, the interested reader can find far more development of the mathematics, along with physical applications, by studying the physics of fluctuating membranes. Section 10.4 of the condensed matter physics textbook from [Chaikin and Lubensky \(1995\)](#) provides a useful starting point.

5.1	Definitions and notation	110
5.1.1	Definitions	110
5.1.2	Notation	110
5.2	Curves in 2D Euclidian space (planar curves)	111
5.2.1	Differential increments along the curve	111
5.2.2	Length along the curve	112
5.2.3	Curvature of a curve	113
5.3	Surfaces embedded in 3D Euclidean space	114
5.3.1	Area on the surface	115
5.3.2	Curvature of a surface	115
5.3.3	Curves on the surface $z = \eta(x, y)$	117
5.4	Exercises	118

5.1 Definitions and notation

The basic notions of curves and surfaces embedded in Euclidean space are rather intuitive. Nonetheless, it is important to be precise in our usage. For this purpose we here offer some notation and definitions.

5.1.1 Definitions

We assume the notion of a curve and surface embedded in Euclidean space to be self-evident, offering analytical expressions for curves in Section 5.2 and surfaces in Section 5.3. Given such, we here define some related notions used in our study of fluid mechanics.

- **ORIENTABLE:** An *orientable curve* is a curve that allows for normal and tangent directions to specify directions and sides to the curve. Likewise, an *orientable surface* has two sides, allowing one to define positive and negative sides. A Möbius strip is the canonical surface that is not orientable since it only has one side. Likewise, the boundary of a Möbius strip is a non-orientable curve. Although very interesting mathematically, we do not encounter non-orientable objects in this book.
- **PATH or CONTOUR:** A path or contour is a continuous piecewise smooth oriented curve. A *simple path* or *simple contour* does not cross itself. We already encountered contours when considering path integrals in Chapter 2.
- **CIRCUIT:** A circuit is a path that closes, and a *simple circuit* is a circuit that does not cross itself. Finally, a *reducible circuit* is a circuit that can be continuously deformed to a point within the domain without leaving the domain. For example, a circuit within the ocean that encloses an island or continent cannot be deformed to a point since doing so requires the circuit to cross onto land and thus to leave the ocean.

5.1.2 Notation

In this chapter we write the Cartesian position of a point on a surface as

$$\mathbf{S} = x \hat{\mathbf{x}} + y \hat{\mathbf{y}} + \eta(x, y) \hat{\mathbf{z}} \quad \text{position on surface,} \quad (5.1)$$

with the vertical position written as

$$z = \eta(x, y) \quad \text{vertical position on surface.} \quad (5.2)$$

If we are instead referring to a point on a planar curve in the x - z -plane, then we drop the y -dependence to have

$$\mathbf{S} = x \hat{\mathbf{x}} + \eta(x) \hat{\mathbf{z}} \quad \text{position on planar curve.} \quad (5.3)$$

Time dependence is dropped throughout this chapter since we focus on the spatial geometry of curves and surfaces at a particular time instance.

We assume the outward normal direction on the curve or the surface has a nonzero projection into the vertical as shown in Figure 5.1. Indeed, we are only able to write the vertical position as $z = \eta(x, y)$ so long as there are no overturns in the surface, in which case the outward normal direction is

$$\hat{\mathbf{n}} = \frac{\nabla(z - \eta)}{|\nabla(z - \eta)|} = \frac{\hat{\mathbf{z}} - \nabla\eta}{\sqrt{1 + |\nabla\eta|^2}}. \quad (5.4)$$

Figure 5.2 provides an example surface along with the notation.

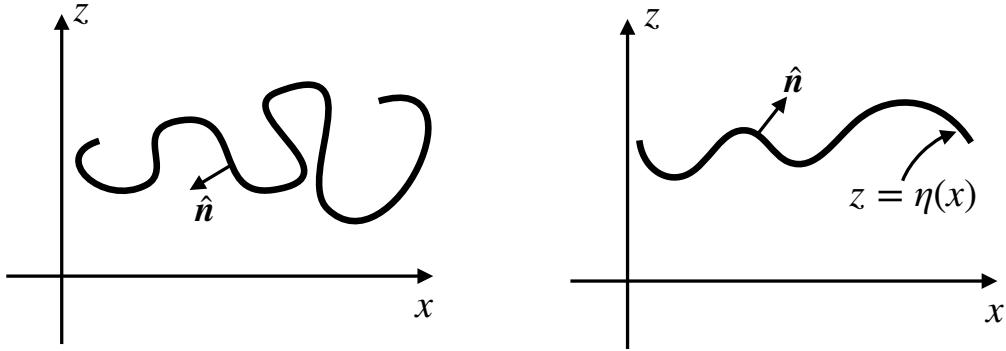


FIGURE 5.1: Two sample curves on the x - z plane. The left panel shows a curve whose outward normal, $\hat{\mathbf{n}}$, encounters points where $\hat{\mathbf{n}} \cdot \hat{\mathbf{z}} = 0$ and where $\hat{\mathbf{n}} \cdot \hat{\mathbf{z}}$ changes sign. This curve, and its generalization to a surface, are not treated in this chapter. The right panel shows a more gently undulating curve where $\hat{\mathbf{n}} \cdot \hat{\mathbf{z}} \neq 0$ everywhere, and thus where $\hat{\mathbf{n}} \cdot \hat{\mathbf{z}}$ is single signed. For curves such as the right panel, we can express the vertical position as a 1-to-1 function of the horizontal position, $z = \eta(x)$. Again, this curve has its natural generalization to a gently undulating surface whereby $z = \eta(x, y)$ provides a unique mapping between horizontal position and vertical. The assumption regarding no overhanging curves and surfaces is consistent with our study of surfaces defined by a constant generalized vertical coordinate (e.g., isopycnals or isentropes) in Chapter 43.

5.2 Curves in 2D Euclidian space (planar curves)

We here describe the geometry of a curve on the x - z -plane (a *planar curve*) as depicted in Figure 5.3. These curves are one-dimensional objects living in a two-dimensional Euclidean space. Extensions to curves on non-Euclidean surfaces, such as the sphere or an isopycnal, are straightforward when those surfaces are embedded in a background Euclidean space as in our studies.

5.2.1 Differential increments along the curve

As a one-dimensional geometric object, an arbitrary curve can be parameterized by a single coordinate, referred to here as φ . Let $\mathbf{S}(\varphi)$ specify the position of a point along the curve. Correspondingly, the differential increment between two infinitesimally close points on the curve is given by

$$\mathbf{S}(\varphi + d\varphi) - \mathbf{S}(\varphi) = d\mathbf{S} = \frac{d\mathbf{S}}{d\varphi} d\varphi \equiv \mathbf{t} d\varphi, \quad (5.5)$$

where

$$\mathbf{t} = \frac{d\mathbf{S}}{d\varphi} \quad (5.6)$$

is tangent to the curve. If $\varphi = s$ is the arc length along the curve, then $\mathbf{t} = \hat{\mathbf{t}}$ is a unit vector

$$\hat{\mathbf{t}} \cdot \hat{\mathbf{t}} = \frac{d\mathbf{S}}{ds} \cdot \frac{d\mathbf{S}}{ds} = 1. \quad (5.7)$$

In some treatments in this book we also write

$$\hat{\mathbf{s}} = \hat{\mathbf{t}} \quad (5.8)$$

to correspond to s for arc length. Recall we made use of the arc length along a curve in Section 2.4 when describing path integration.

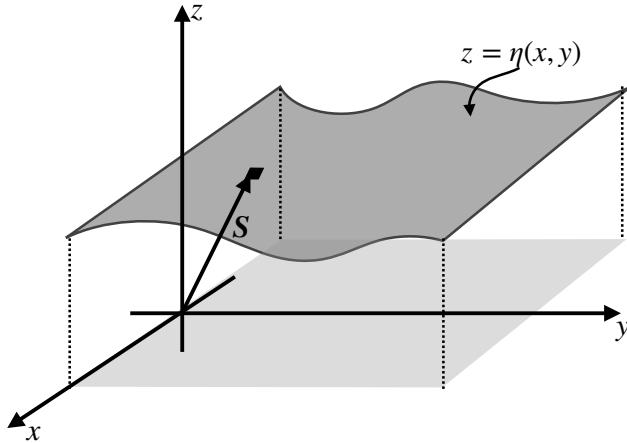


FIGURE 5.2: An example of a surface embedded in Euclidean space. The position of a point on the surface is given by the Cartesian position vector $\mathbf{S} = x \hat{\mathbf{x}} + y \hat{\mathbf{y}} + \eta(x, y) \hat{\mathbf{z}}$. The relation $z = \eta(x, y)$ provides a 1-to-1 mapping between the horizontal position and the vertical position of a point on the surface. Correspondingly, the surface is uniquely specified by finding the envelope of points where $z - \eta(x, y) = 0$. The lightly shaded region represents the projection of the curved surface onto the flat horizontal $x - y$ plane below.

5.2.2 Length along the curve

As in equation (5.3) we can represent the position of a point along the curve using Cartesian coordinates

$$\mathbf{S} = x \hat{\mathbf{x}} + \eta(x) \hat{\mathbf{z}}. \quad (5.9)$$

Hence, letting $\varphi = x$ parameterize the curve leads to the representation of the tangent direction

$$\mathbf{t} = \frac{d\mathbf{S}}{dx} = \hat{\mathbf{x}} + \frac{d\eta}{dx} \hat{\mathbf{z}}, \quad (5.10)$$

which has the magnitude

$$\mathbf{t} \cdot \mathbf{t} = 1 + (d\eta/dx)^2, \quad (5.11)$$

so that the normalized tangent vector is

$$\hat{\mathbf{t}} = \frac{\hat{\mathbf{x}} + (d\eta/dx) \hat{\mathbf{z}}}{\sqrt{1 + (d\eta/dx)^2}}. \quad (5.12)$$

Likewise, the curve's normal vector is given by

$$\hat{\mathbf{n}} = \frac{\nabla(z - \eta)}{|\nabla(z - \eta)|} = \frac{\hat{\mathbf{z}} - (d\eta/dx) \hat{\mathbf{x}}}{\sqrt{1 + (d\eta/dx)^2}}, \quad (5.13)$$

with orthogonality manifest

$$\hat{\mathbf{t}} \cdot \hat{\mathbf{n}} = 0. \quad (5.14)$$

The squared length of an infinitesimal segment along the curve is given by

$$(ds)^2 = d\mathbf{S} \cdot d\mathbf{S} = \left[\frac{d\mathbf{S}}{dx} \cdot \frac{d\mathbf{S}}{dx} \right] dx dx, \quad (5.15)$$

so that the finite length of the curve is determined by the integral

$$L = \int_0^L ds = \int_{x_A}^{x_B} |d\mathbf{S}/dx| dx = \int_{x_A}^{x_B} \sqrt{1 + (d\eta/dx)^2} dx, \quad (5.16)$$

where $x_A \leq x \leq x_B$ is the range over which x runs for the projection of the curve onto the x -axis (see Figure 5.3).

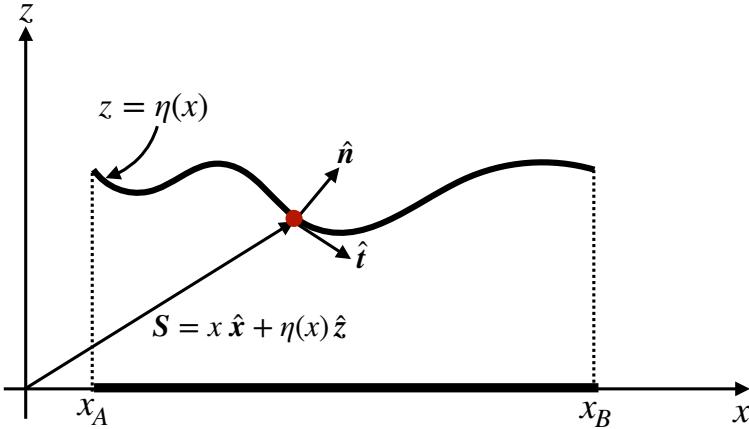


FIGURE 5.3: An orientable path in the x - z plane defined by a planar curve that does not intersect itself. The Cartesian position of a point on the curve is given by $\mathbf{S} = x \hat{\mathbf{x}} + \eta(x) \hat{\mathbf{z}}$, where $z = \eta(x)$ is the vertical position of the point. The projection of the curve onto the horizontal x -axis occupies a range $x_A \leq x \leq x_B$. One way to define the curve is by finding the envelope of points where $z - \eta(x) = 0$, in which case we can readily find the normal direction pointing upward as $\hat{\mathbf{n}} = \nabla(z - \eta)/|\nabla(z - \eta)| = [\hat{\mathbf{z}} - (d\eta/dx) \hat{\mathbf{x}}] [1 + (d\eta/dx)^2]^{-1/2}$, and the normalized tangent direction $\hat{\mathbf{t}} = [\hat{\mathbf{x}} + (d\eta/dx) \hat{\mathbf{z}}] [1 + (d\eta/dx)^2]^{-1/2}$.

5.2.3 Curvature of a curve

Curvature measures the amount that the normal direction changes along the curve. For a planar curve, the curvature at a point equals to the inverse radius of a circle that shares the same tangent plane to the curve at the point (see Figure 5.4). We refer to the radius as the *radius of curvature* and the corresponding circle as the *curvature circle*. To formulate an analytic expression for the radius of curvature at a point on a curve, orient the Cartesian coordinate axes so that the point is at the origin and the tangent plane sits along the x -axis as in Figure 5.4. Consequently, the outward normal, $\hat{\mathbf{n}}$, is parallel to the $\hat{\mathbf{z}}$ direction.

A Taylor series expansion about the origin tells us that the vertical position of a point along the curve and near to the origin can be written

$$\eta(x) = \eta(0) + x \left[\frac{d\eta}{dx} \right]_{x=0} + \frac{x^2}{2} \left[\frac{d^2\eta}{dx^2} \right]_{x=0} + \mathcal{O}(x^3) \quad (5.17a)$$

$$= \frac{x^2}{2} \left[\frac{d^2\eta}{dx^2} \right]_{x=0} + \mathcal{O}(x^3). \quad (5.17b)$$

This result follows since we placed the origin so that $\eta(0) = 0$, and aligned the x -axis so that it is a tangent plane at the origin, in which case $d\eta/dx = 0$ at $x = 0$. Hence, η has a quadratic behavior near the origin.

Now place a circle with center along the z -axis so that it is tangent to the curve at the origin, as depicted in Figure 5.4. What is the radius, R , of the circle that best fits the curve at the origin? To answer this question note that the height of a point on the circle is given by $h(x) = R(1 - \cos \vartheta)$, where $\vartheta = 0$ for a point at the origin and $\vartheta = \pi$ at the diametrically opposite point. For small ϑ this height takes the form

$$h(x) \approx R[1 - 1 + \vartheta^2/2] = x^2/(2R), \quad (5.18)$$

where $\vartheta = x/R$ near the origin. For the height of a point on the curve (equation (5.17b)) to match the height along the circle, to second order accuracy, requires us to set the circle's radius to

$$\frac{1}{R} = \frac{d^2\eta}{dx^2}. \quad (5.19)$$

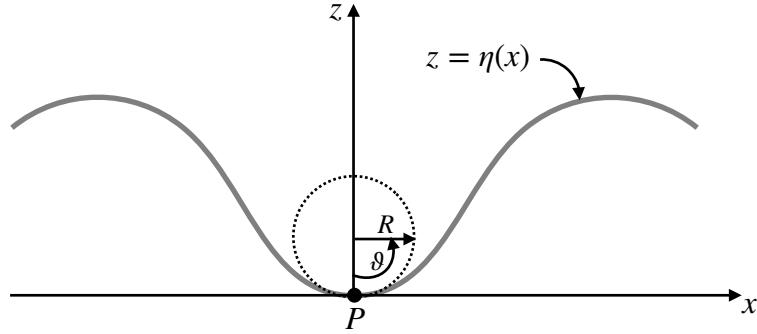


FIGURE 5.4: The radius of curvature at a point on a curve, P , equals to the radius of the curvature circle that shares the same tangent plane as the curve at the point P . When constructing the curvature circle we make use of the angle, ϑ , to measure the height of a point along the circle, $h(x) = R(1 - \cos \vartheta) \approx R\vartheta^2/2 \approx x^2/(2R)$. Setting $R^{-1} = d^2\eta/dx^2$ provides a second order accurate fit of the curvature circle to the curve at the point P .

Equation (5.19) thus provides an expression for the radius of curvature, R , whose inverse is the curvature

$$\text{curvature} = \frac{1}{R}. \quad (5.20)$$

This result supports our expectation that the second derivative measures the curvature. As R gets larger the curvature decreases as the circle approaches a flat plane. In the opposite limit the curvature grows as R decreases. Note that we could have chosen to orient the circle on the opposite side of the tangent (on the convex side), in which case the radius of curvature is negative. That is, $R > 0$ when the normal points towards the concave side (side where the curve rises towards $\hat{\mathbf{n}}$), whereas $R < 0$ when the normal points towards the convex side (side where the curve falls away from $\hat{\mathbf{n}}$).

In closing this section we note that

$$-\nabla \cdot \hat{\mathbf{n}} = \frac{d^2\eta/dx^2}{[1 + (d\eta/dx)^2]^{3/2}}. \quad (5.21)$$

When evaluated at the point of interest along the curve, we set $d\eta/dx = 0$ so that

$$-\nabla \cdot \hat{\mathbf{n}} = \frac{d^2\eta}{dx^2} = \frac{1}{R}. \quad (5.22)$$

This result supports our earlier statement that curvature measures the change in the normal direction along the curve. In fact, the identity

$$-\nabla \cdot \hat{\mathbf{n}} = \frac{1}{R} \quad (5.23)$$

holds for an arbitrary point along the curve since it is a coordinate invariant statement.

5.3 Surfaces embedded in 3D Euclidean space

We now extend the previous discussion to a two-dimensional surface embedded in three-dimensional Euclidean space such as in Figure 5.2. In general, a 2D surface in 3D space can be parameterized by two variables, φ^1 and φ^2 , so that infinitesimal increments along the surface satisfy

$$d\mathbf{S} = \frac{\partial \mathbf{S}}{\partial \varphi^1} d\varphi^1 + \frac{\partial \mathbf{S}}{\partial \varphi^2} d\varphi^2 = \mathbf{t}_1 d\varphi^1 + \mathbf{t}_2 d\varphi^2. \quad (5.24)$$

The vectors \mathbf{t}_1 and \mathbf{t}_2 are tangent to the surface at the point (φ^1, φ^2) , and yet they are not generally orthogonal to one another.

Making use of the Cartesian expression (5.1)

$$\mathbf{S} = x \hat{\mathbf{x}} + y \hat{\mathbf{y}} + \eta(x, y) \hat{\mathbf{z}} \quad (5.25)$$

brings the two tangent directions into the form

$$\mathbf{t}_1 = \frac{\partial \mathbf{S}}{\partial x} = \hat{\mathbf{x}} + \frac{\partial \eta}{\partial x} \hat{\mathbf{z}} \quad (5.26a)$$

$$\mathbf{t}_2 = \frac{\partial \mathbf{S}}{\partial y} = \hat{\mathbf{y}} + \frac{\partial \eta}{\partial y} \hat{\mathbf{z}}. \quad (5.26b)$$

Likewise, the surface normal vector is given by

$$\hat{\mathbf{n}} = \frac{\nabla(z - \eta)}{|\nabla(z - \eta)|} = \frac{\hat{\mathbf{z}} - \nabla\eta}{|\hat{\mathbf{z}} - \nabla\eta|} = \frac{\hat{\mathbf{z}} - \nabla\eta}{\sqrt{1 + |\nabla\eta|^2}}, \quad (5.27)$$

and it is straightforward to show orthogonality with the two tangent vectors

$$\hat{\mathbf{t}}_1 \cdot \hat{\mathbf{n}} = \hat{\mathbf{t}}_2 \cdot \hat{\mathbf{n}} = 0. \quad (5.28)$$

5.3.1 Area on the surface

Recall from Section 1.4.5 that the magnitude of a vector product of two vectors equals to the area of the parallelogram subtended by the vectors. Hence, the area of an infinitesimal surface element with sides $d\varphi_1$ and $d\varphi_2$ is given by

$$d\mathcal{S} = \left| \frac{\partial \mathbf{S}}{\partial \varphi^1} \wedge \frac{\partial \mathbf{S}}{\partial \varphi^2} \right| d\varphi_1 d\varphi_2. \quad (5.29)$$

Making use of Cartesian coordinates brings the area element to

$$d\mathcal{S} = \sqrt{1 + |\nabla\eta|^2} dx dy = \sqrt{1 + |\nabla\eta|^2} dA, \quad (5.30)$$

where

$$dA = dx dy \quad (5.31)$$

is the area of the surface projected onto the horizontal plane. Hence, the area of a finite region is given by the integral

$$\mathcal{S} = \int d\mathcal{S} = \int |\nabla(z - \eta)| dA = \int \sqrt{1 + |\nabla\eta|^2} dx dy, \quad (5.32)$$

where the second and third integrals extend over the region defined by the projection of the surface onto the horizontal (see Figure 5.2).

5.3.2 Curvature of a surface

We now seek an expression for the curvature of a point on the surface. Since the surface has two dimensions, we expect the curvature to be measured by two numbers rather than the single curvature for the curve discussed in Section 5.2.3. The method for developing the curvature is

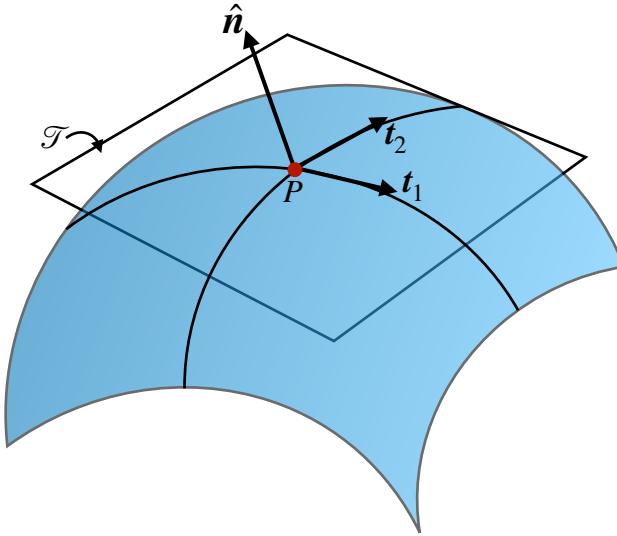


FIGURE 5.5: Depicting the elements needed to construct the curvature of a surface at an arbitrary point, P . The local normal direction is given by \hat{n} , along with the two tangent vectors t_1 and t_2 . The tangent vectors span the space of the tangent plane, \mathcal{T} , shown as a flat surface that is tangent at the chosen point on the surface. In this case the surface falls away from the normal direction, as per a convex surface, so that the two radii of curvature are negative. Other surfaces can be concave, whereby both radii of curvature are positive, or hyperbolic (saddle), whereby one is positive and another negative.

analogous to that used for a curve, yet with a bit more mathematics needed to allow for the extra dimension. Figure 5.5 depicts the situation.

Let $\mathbf{x} = (x_1, x_2) = (x, y)$ be Cartesian coordinates on a tangent plane local to an arbitrary point on the surface, with the origin of the coordinate system taken at the point. Near to the point, we can estimate the vertical distance of a point on the surface from the tangent plane according to the quadratic form

$$\eta \approx \frac{1}{2} x_m \mathbb{C}_{mn} x_n, \quad (5.33)$$

where \mathbb{C} is the second order tensor of second partial derivatives evaluated at the point

$$\mathbb{C} = \begin{bmatrix} \frac{\partial^2 \eta}{\partial x_1^2} & \frac{\partial^2 \eta}{\partial x_1 \partial x_2} \\ \frac{\partial^2 \eta}{\partial x_1 \partial x_2} & \frac{\partial^2 \eta}{\partial x_2^2} \end{bmatrix}. \quad (5.34)$$

As a symmetric tensor, \mathbb{C} is diagonalizable and it has two eigenvalues, R_1^{-1} and R_2^{-1} , along with its associated eigenvectors, \mathbf{e}_1 and \mathbf{e}_2 . The quadratic form (5.33) can thus be written as

$$\eta \approx \frac{1}{2} R_1^{-1} (\mathbf{x} \cdot \mathbf{e}_1)^2 + \frac{1}{2} R_2^{-1} (\mathbf{x} \cdot \mathbf{e}_2)^2. \quad (5.35)$$

R_1 and R_2 are the principle radii of curvature for the surface at the point P . They correspond, respectively, to the radii of the curvature circles in the $\hat{n}\text{-}\mathbf{e}_1$ and $\hat{n}\text{-}\mathbf{e}_2$ planes. If the radius of curvature R_m is positive, then the surface curves towards \hat{n} along the $\hat{n}\text{-}\mathbf{e}_i$ plane, and conversely if R_m is negative. The surface takes the shape of a saddle when the two radii of curvature have opposite signs.

There are two scalar invariants of the tensor \mathbb{C} that commonly arise in applications.

- $\text{Tr}(\mathbb{C}) = R_1^{-1} + R_2^{-1}$, which is twice the mean curvature for the surface. With the normal vector given by equation (5.4), one can show that

$$-\nabla \cdot \hat{\mathbf{n}} = \frac{\nabla^2 \eta}{[1 + (\nabla \eta)^2]^{3/2}}. \quad (5.36)$$

A bit of algebra leads us to conclude that

$$-\nabla \cdot \hat{\mathbf{n}} = \frac{1}{R_1} + \frac{1}{R_2}, \quad (5.37)$$

for any point along the surface, thus generalizing the result (5.22) found for a curve.

- $\det(\mathbb{C}) = 1/(R_1 R_2)$ is known as the *Gaussian curvature*, which is the product of the two curvatures.

5.3.3 Curves on the surface $z = \eta(x, y)$

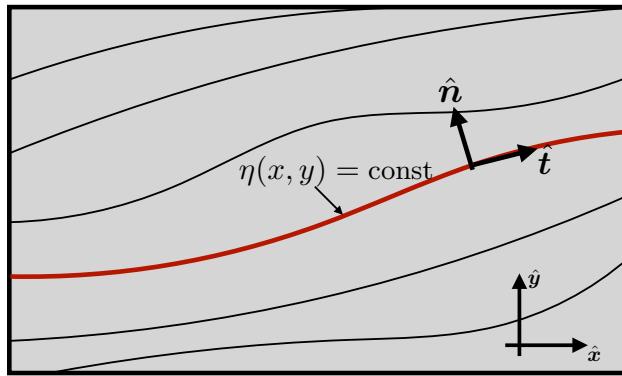


FIGURE 5.6: Geometry depicting a contour along a particular line of constant $z = \eta(x, y)$, such as for a constant elevation path along a mountain or valley. The along-contour direction is $\hat{\mathbf{t}} = d\mathbf{x}/ds$, with s the arc length along the contour. The unit direction pointing to the left of the contour direction is $\hat{\mathbf{n}}$, with $\hat{\mathbf{n}} \cdot \hat{\mathbf{t}} = 0$ and $\hat{\mathbf{t}} \wedge \hat{\mathbf{n}} = \hat{\mathbf{z}}$. Both $\hat{\mathbf{n}}$ and $\hat{\mathbf{t}}$ are horizontal unit vectors.

We now consider a curve, as in Section 5.2, defined along a two dimensional surface, $z = \eta(x, y)$, with the curve defined by lines of constant $z = \eta(x, y)$. For example, if $\eta(x, y)$ is the solid earth topography, then lines of constant η are contours of constant topography. By definition, these contours have no projection into the vertical direction (i.e., they do not go uphill or downhill), and they are determined by

$$d\eta = 0 = \nabla \eta \cdot d\mathbf{x}, \quad (5.38)$$

where $d\mathbf{x} = \hat{\mathbf{x}} dx + \hat{\mathbf{y}} dy$ is the horizontal space increment along the contour. Following the introduction of arc length in Section 5.2.1, we write

$$d\eta = 0 = \nabla \eta \cdot d\mathbf{x} = \nabla \eta \cdot \frac{d\mathbf{x}}{ds} ds = \nabla \eta \cdot \hat{\mathbf{t}} ds, \quad (5.39)$$

where $\hat{\mathbf{t}}$ is a unit vector pointing in the direction of the contour. To build an orthogonal triad of coordinates, we then define an orthogonal unit vector, $\hat{\mathbf{n}}$, that points to the left of the contour direction so that

$$\hat{\mathbf{n}} \cdot \hat{\mathbf{t}} = 0 \quad \text{and} \quad \hat{\mathbf{t}} \wedge \hat{\mathbf{n}} = \hat{\mathbf{z}}, \quad (5.40)$$

as depicted in Figure 5.6.



5.4 Exercises

EXERCISE 5.1: JACOBIAN EVALUATED ALONG A CONTOUR

Consider the vector cross product of two functions, $\psi(x, y)$ and $Q(x, y)$

$$\hat{z} \cdot (\nabla\psi \wedge \nabla Q) = \partial_x\psi \partial_y Q - \partial_y\psi \partial_x Q \equiv J(\psi, Q), \quad (5.41)$$

where the final equality defined the Jacobian operator. Show that when evaluated along a contour of constant Q , the Jacobian is given by

$$J(\psi, Q) = -(\hat{n} \cdot \nabla Q) (\hat{t} \cdot \nabla\psi) \quad (5.42)$$

where \hat{t} is the unit tangent along the contour and \hat{n} is a unit vector pointing to the left of the tangent (e.g., see Figure 5.6).

EXERCISE 5.2: CURVATURE OF A CIRCLE

Make use of equation (5.23), as well as the polar coordinates from Section 8.3, to show that for a circle of radius r we have

$$-\nabla \cdot \hat{n} = \frac{1}{r}, \quad (5.43)$$

where $\hat{n} = -\hat{r}$ points to the left when moving around the circle in a counterclockwise direction. Hence, the radius of curvature for the circle equals to the circle's radius.



General tensors in brief

Vector calculus, as formalized by the Cartesian tensor analysis of Chapters 1 and 2, is sufficient for many areas of geophysical fluid mechanics. However, there are a number of applications where general tensors prove of use to ensure that the physics shines through the maths. We thus go beyond Cartesian tensor analysis to here consider general tensors, thus enabling a more versatile, and precise, mathematical framework for the study of geophysical fluid mechanics. Besides enhancing our tools for geophysical fluid mechanics, general tensor analysis is found throughout physics so that understanding the formalism, if only its rudiments, can greatly help to understand the literature.

Geophysical fluid mechanics applications require only a modest level of new formalism in the transition from Cartesian tensors to general tensors. The key reason is that geophysical fluids are assumed to be embedded within Euclidean space, \mathbb{R}^3 , where Euclidean space is the flat space familiar from Newtonian mechanics. Notably, \mathbb{R}^3 has zero intrinsic curvature so we say that it is a *flat space*. So although we are concerned with fluid motion on curved manifolds (e.g., spherical planets); motion on curved and fluctuating manifolds (e.g., isopycnals); and in describing that motion using non-orthogonal coordinates (e.g., generalized vertical coordinates), the fluid remains embedded within a background Euclidean space. Through this embedding, the local geometry inherits features from the Euclidean space such as how to measure distance between points via the Pythagorean theorem. A further simplification for geophysical fluid mechanics arises since we make use of universal Newtonian time. Hence, time is measured the same by all observers and reference frames. That is, the spatial coordinates used by geophysical fluid mechanics can be a function of time, but the time coordinate is always the same. Each of these features of the space and time used for geophysical fluid mechanics means that our mathematical needs are far simpler than the general relativist who studies fluids moving in strong gravity fields and/or over galactic distances.

Throughout our exploration of general tensor analysis it is useful to keep in mind the fundamental goal is to uncover objective statements about how physical systems operate. In particular, we aim to make use of a formalism that

- supports computations to allow for the comparison of theory with measurements;
- provides a systematic technology to transform between coordinate systems, since there are many occasions where such transformations are useful to develop physical insight and support calculations;
- provides equations that are independent of specific coordinate system.

These goals are realized by general tensor analysis so that the technology reflects the underlying objective nature of physical relationships while allowing for a quantitative realization of those relationships in particular situations.

READER'S GUIDE TO THIS CHAPTER

This brief chapter offers an overview sufficient to appreciate why we need general tensors for this book. We do so by offering a conceptual platform for general tensor analysis, with details of the technology presented in Chapter 7. We focus on tensor analysis on spatial manifolds endowed with a metric, thus touching on the rudiments of *Riemannian differential geometry*. Furthermore, we restrict attention to manifolds embedded in Euclidean space, thus providing a natural extension of the Cartesian tensor analysis from Chapters 1 and 2.

6.1	Coordinate independent physical relations	120
6.1.1	Tensor operations	121
6.1.2	Comments	121
6.2	Points, trajectories, and coordinates	122
6.2.1	Time as a parameter and time as a coordinate	122
6.2.2	The importance of index placement	122
6.3	Example coordinate descriptions	123
6.3.1	Eulerian coordinates	123
6.3.2	Isopycnal coordinates	123
6.3.3	Lagrangian or material coordinates	124
6.3.4	Tracer coordinates	124
6.4	The velocity vector and basis vectors	124
6.4.1	Coordinate representation	125
6.4.2	Basis vectors	125
6.5	Notational conventions	125
6.5.1	Placement of tensor labels	126
6.5.2	Covariant, contravariant, and Einstein summation	126
6.5.3	The boldface notation	126
6.5.4	Space-time approach	127

6.1 Coordinate independent physical relations

Physical relations are independent of subjective choices for their mathematical representations. This principle motivates us to seek mathematical expressions between objects whose meaning transcends a particular coordinate representation. At its most profound level, there is no *a priori* notion of the underlying geometry, with an insistence on such generalities leading to general relativity and the notion of *general covariance*. We are not going that far in this book, given that we accept the notion of a background Euclidean space on which the fluid moves, and we make use of a universal Newtonian time measured by all observers. Nonetheless, we insist that the mathematical expression of a physical relation be independent of the choice of coordinates (*coordinate invariance*), which is ensured when expressed as relations between geometric objects such as points, vectors, and tensors.

Although physics does not care about coordinates, physicists often do. Namely, it is convenient, and sometimes necessary, to work with specific coordinates suited to the symmetry of the physical system. After deriving a physical law in one set of coordinates, it is often of interest to establish the form of the law in another set of coordinates. How does the physical law, typically represented as a differential equation, transform into other coordinates? So long as the equations are written in a proper tensorial form then the equations are form invariant. That is, coordinate invariance means that the equations look the same regardless the choice of coordinates. Consequently, “physics as geometry” has a major practical implication. Namely, we can establish the validity of a physical

relation in any convenient set of coordinates, and then extend that relation to all coordinates so long as we respect the basic rules of tensor analysis.

6.1.1 Tensor operations

Extending a mathematical equation to all coordinates requires the equations to respect certain tensor rules. In brief, all tensor indices are properly matched and each derivative is covariant as specified in Section 7.10. In chapter 7, we provide the details needed to understand general coordinate invariance. In this chapter we outline the procedure. The elegance and power rendered by coordinate invariance is the key reason that tensor analysis is ubiquitous in theoretical physics.

To ensure an equation respects coordinate invariance requires us to understand certain properties of tensors and operations with tensors that produce components of new tensors. We here summarize the specific properties characterizing coordinate invariance (taken after page 153 of *Schutz, 1985*):

1. Manipulations of tensor components are called *permissible tensor operations* if they produce components of new tensors. The following are permissible operations:
 - (a) Multiplication of a tensor by a scalar produces a new tensor of the same type.
 - (b) Addition of components of two tensors of the same type gives components of a new tensor of the same type. In particular, only tensors of the same type can be equal.
 - (c) Multiplication of components of two tensors of arbitrary type gives components of a new tensor whose type is given by the sum of the types for the individual tensors. This operation is called the *outer product* or *tensor product* and is denoted by the operator \otimes . For example, $\mathbf{A} \otimes \mathbf{B}$ is a second order tensor built from the outer product of two vectors, \mathbf{A} and \mathbf{B} .
 - (d) A corollary of the multiplication rule is that if two objects yield a tensor, and one of these objects is a tensor then so too is the other. This is the *quotient rule*.
 - (e) Covariant differentiation (Sections 7.10 and 7.11) increases by one the order of a tensor, with the covariant derivative operator denoted by ∇ .
 - (f) Contraction on a pair of indices of the components of a tensor reduces by one the order of a tensor.
2. If two tensors of the same type have equal components in a given coordinate system, then they have equal components in all coordinate systems. Hence, the tensors are identical.
3. If a mathematical equation consists of tensors combined only by the permissible tensor operations, and if the equation is true in one coordinate system, then it is true in any coordinate system. If the equations involve covariant derivatives, then the equations remain form invariant under changes in coordinates. For the partial differential equations of geophysical fluid mechanics, covariant differentiation is the key to coordinate invariance.

6.1.2 Comments

The remainder of this chapter, as well as Chapter 7, provide the needed details for supporting the above notions of coordinate invariance. Even without penetrating these details, the reader should be able to appreciate why coordinate invariance is so central to physics.

6.2 Points, trajectories, and coordinates

Consider a point in space, \mathcal{P} , at a particular time τ . As time progresses, the point traces out a curve in space-time. We call that curve a *trajectory*. The trajectory could be of a point particle following the fluid flow, thus defining the Lagrangian reference frame (Section 13.2). Or it could trace the path of something else such as a fish, balloon, boat, or airplane. As the trajectory is a one-dimensional curve, it is specified mathematically by a single parameter (see Section 5.2.1). We choose the time measured by an observer on the trajectory for this parameter, in which case the trajectory is written $\mathcal{P}(\tau)$.

A point in the fluid and its trajectory in space-time are geometric objects that exist independently of any coordinate representation. Even so, we find the need to represent points, trajectories, vectors, and other geometric objects using coordinates. For example, coordinates are needed to make quantitative statements about fluid flow in relation to other observers. What is its speed and direction relative to a chosen reference frame? What is the distance from an origin or from another particle? Tensor analysis provides a formalism that enables us to answer such quantitative questions while maintaining a clear view on the underlying physics and geometry.

6.2.1 Time as a parameter and time as a coordinate

In special and general relativity, there is a mixing of space and time that warrants the use of four-dimensional space-time tensor analysis. In contrast, for classical mechanics forming the foundation of geophysical fluid mechanics, time remains numerically the same throughout space. We thus make use of the same universal (or Newtonian) time since the fluid velocity and wave speeds are far smaller than the speed of light.

The time parameter, τ , specifies a point along a trajectory. The coordinate time, t , measures time for all positions throughout space. This distinction between the time parameter and time coordinate is pedantic given that $\tau = t$ (up to a constant offset) in a Newtonian universe. Nonetheless, it is convenient to make the distinction when measuring how fluid properties change since these changes are subject to motion of the observer. For example, changes following a trajectory, found by computing the trajectory time derivative $\partial/\partial\tau$, are generally distinct from changes found by computing the time derivative $\partial/\partial t$, in which the spatial coordinates are held fixed.

When the trajectory is defined by a fluid particle, we refer to $\partial/\partial\tau$ as the material or Lagrangian time derivative. This is also the time derivative computed when working with Newtonian particle mechanics as in Part II of this book. In contrast, if the spatial coordinates are fixed in space, then $\partial/\partial t$ is an Eulerian time derivative. When alternative spatial coordinates are used, some of which can move (see Section 6.2.2), then $\partial/\partial t$ can be a mixture of Lagrangian and Eulerian or perhaps neither.

6.2.2 The importance of index placement

Much of the formalism of general tensor analysis builds from Cartesian representations of vector and matrix analysis, with generalizations that provide objective statements independent of coordinates. One key point of distinction from Cartesian tensors is that the position (up or down) of a tensor label has significance in general tensor analysis. We follow the standard convention by labeling an arbitrary coordinate with indices upstairs in their *contravariant* position, ξ^α . It is furthermore common to express each of the coordinates in an ordered list according to

$$\xi^\alpha = (\xi^0, \xi^1, \xi^2, \xi^3). \quad (6.1)$$

We emphasize that the index is a label for the coordinate; it is not an exponent or power. We use a convention whereby Greek labels run from $\alpha = 0, 1, 2, 3$ with $\alpha = 0$ the time coordinate and $\alpha = a = 1, 2, 3$ the three coordinates used to locate a point in the Euclidean space of Newtonian mechanics. The following shorthand notations are commonly used in this book

$$\xi^\alpha = (\xi^0, \xi^1, \xi^2, \xi^3) = (\xi^0, \xi^a) = (\xi^0, \boldsymbol{\xi}). \quad (6.2)$$

6.3 Example coordinate descriptions

We here offer a few examples of coordinates used for describing geophysical fluid systems. As the time coordinate remains universal in our study, we are here only concerned with the spatial coordinates, $a = 1, 2, 3$.

6.3.1 Eulerian coordinates

The Cartesian coordinates for a point are written

$$\xi^a = (x, y, z) \quad \text{Cartesian}, \quad (6.3)$$

whereas for spherical coordinates we write (see Figure 8.1)

$$\xi^a = (\lambda, \phi, r) \quad \text{spherical} \quad (6.4)$$

and cylindrical polar coordinates (see Figure 8.2)

$$\xi^a = (r, \vartheta, z) \quad \text{cylindrical}. \quad (6.5)$$

These coordinates identify fixed positions in space. We can use these *Eulerian* coordinates to mark the trajectory $\mathcal{P}(\tau)$ as it crosses the spatial point ξ^a at time t . We provide a more complete discussion of Eulerian coordinates in Section 13.3.

As shown in Section 6.4, Cartesian coordinates are notable for having basis vectors maintaining a fixed direction throughout space. This feature lends much simplicity to Cartesian coordinates and its corresponding Cartesian tensor analysis (Chapters 1 and 2). In contrast, the spherical basis vectors are spatially dependent. Likewise, the radial and angular basis vectors for polar cylindrical coordinates are spatially dependent, whereas the vertical direction is fixed. Additionally, the spherical and cylindrical coordinates do not all have the same physical dimensions. Each of these features of spherical and cylindrical coordinates places them outside the purview of Cartesian tensor analysis.

6.3.2 Isopycnal coordinates

In geophysical fluids that are stably stratified in the vertical, it is common to measure the vertical position of a fluid element by specifying its entropy, buoyancy, or potential density depending on the application. We generically write these *isentropic*, *buoyancy*, or *isopycnal* coordinates as

$$\xi^a = (x, y, b) \quad \text{with} \quad b = b(x, y, z, t) \quad \text{isopycnal coordinates}, \quad (6.6)$$

where $b = b(x, y, z, t)$ is a generic symbol for entropy, buoyancy, or potential density. Entropy, buoyancy, and potential density are materially invariant for perfect fluid flow (flow absent irreversible processes such as mixing or heating). Hence, all fluid particle motion occurs on surfaces of constant

b. Under such perfect fluid conditions, isopycnal coordinates are of great use for describing fluid mechanics of stably stratified geophysical flows.

The isopycnal coordinates are generally not orthogonal since the direction normal to a buoyancy surface is not generally vertical. Hence, even if the horizontal coordinates are Cartesian, the use of b to measure the vertical precludes the use of Cartesian tensor analysis. Furthermore, we note the distinct physical dimensions of the three spatial coordinates (x, y, b) , again necessitating the use of general tensor analysis.

6.3.3 Lagrangian or material coordinates

We often conceive of a fluid as a continuum of constant mass fluid elements distinguished by continuum marker coordinates or labels. The initial position for a fluid element offers a common choice for these *material coordinates*. The fluid dynamical equations of motion (i.e., Newton's Law of motion) can be formulated using material coordinates so long as the material coordinate maintains a 1-to-1 relation to points in space. This kinematical framework is termed *Lagrangian* or *material*. The resulting dynamical equations share much in common with Newtonian particle mechanics, though with the added feature of contact forces acting between the fluid elements. We formally describe Lagrangian coordinates in Section 13.3 and encounter Lagrangian descriptions throughout this book.

If we represent material coordinates by the Cartesian positions of fluid elements at an arbitrary initial time, then we can make use of Cartesian tensor analysis. However, it is sometimes useful to make use of alternative markers. One example is the isopycnal coordinate mentioned above, whose value remains invariant under perfect fluid motion. In this manner, we refer to the isopycnal coordinates as "quasi-Lagrangian" since its vertical coordinate follows the vertical position of a perfect fluid parcel whereas its horizontal coordinates are Eulerian.

6.3.4 Tracer coordinates

Consider a triplet of linearly independent tracer concentrations $C^a = C^a(x, y, z, t)$ that spans \mathbb{R}^3 . Hence, at any point in space there is a unique intersection of three constant tracer surfaces, so that we can uniquely determine a point in space by specifying the value for the three tracer concentrations. We can thus use tracer concentrations as the spatial coordinates

$$\xi^a = (C^1, C^2, C^3). \quad (6.7)$$

In some cases there are only two linearly independent tracers, in which case the two may be used in combination with a third spatial coordinate such as depth or pressure. Furthermore, the case of one tracer coordinate formally reduces to the isopycnal coordinate system described above.

6.4 The velocity vector and basis vectors

Consider two points in space that sit along a particular trajectory separated by an infinitesimal time increment, $d\tau$. The velocity vector for this trajectory is defined by

$$\vec{v}(\tau) = \lim_{\Delta\tau \rightarrow 0} \frac{\mathcal{P}(\tau + \Delta\tau/2) - \mathcal{P}(\tau - \Delta\tau/2)}{\Delta\tau} = \frac{d\mathcal{P}(\tau)}{d\tau}. \quad (6.8)$$

The velocity is a vector pointing in the direction determined by the difference between two points on a trajectory, in the limit as the time separation between the points vanishes. Consequently, the

velocity points in a direction tangent to the trajectory. Notably, the above definition for the velocity makes no use of coordinates. Rather, the velocity vector is determined by the geometry of the trajectory and the specification of the trajectory's time parameter. Hence, velocity is fundamentally an arrow with a length and direction and as such, the velocity is a geometric object with this fact emphasized by exposing the arrow.

The definition of velocity as a vector tangent to the trajectory is a general property of all vectors living on a manifold. Namely, a vector at a point on a manifold lives within the tangent plane to the manifold at that point. This observation signals to us the need to be very careful when comparing vectors on a curved manifold such as the sphere or an isopycnal surface. Before comparing the two vectors they must be brought to the same point so they can be compared.

6.4.1 Coordinate representation

We now establish an arbitrary set of spatial coordinates, ξ^a , to represent points in space. These coordinates are used to measure the spatial position of the trajectory according to

$$\mathcal{P}(\tau) = \mathcal{P}[\xi^a(\tau)], \quad (6.9)$$

where $\xi^a(\tau)$ is the coordinate position on the trajectory at time τ . This coordinate representation for the trajectory induces a coordinate representation for the velocity through use of the chain rule

$$\vec{v}(\tau) = \frac{d\mathcal{P}(\tau)}{d\tau} = \frac{\partial \mathcal{P}}{\partial \xi^a} \frac{d\xi^a}{d\tau} \equiv \vec{e}_a v^a. \quad (6.10)$$

The expansion coefficients

$$v^a = \frac{d\xi^a}{d\tau} \quad (6.11)$$

provide a representation for the velocity vector $\vec{v}(\tau)$ within the coordinate system ξ^a .

6.4.2 Basis vectors

For each number v^a there is a corresponding basis vector \vec{e}_a whose value at the point \mathcal{P} is given by

$$\vec{e}_a(\mathcal{P}) = \frac{\partial \mathcal{P}}{\partial \xi^a}. \quad (6.12)$$

Abstracting this expression allows us to identify the basis vector as the partial derivative

$$\vec{e}_a = \frac{\partial}{\partial \xi^a}. \quad (6.13)$$

The basis vectors are generally a function of space and time, with Cartesian coordinates a notable exception whereby they are space-time constants.

6.5 Notational conventions

We here introduce notational conventions that help to simplify many of the manipulations, sometimes referred to as *index gymnastics*, encountered with general tensors.

6.5.1 Placement of tensor labels

As indicated in Section 6.2.2, the placement of tensor labels has specific meaning with general tensor analysis. It is therefore critical to maintain proper usage to ensure *conservation of indices* across an equals sign. As a first example of this usage, notice how the basis vectors in equation (6.12) inherit a lowered tensor label. This placement follows from the partial derivative operator that carries an upper coordinate label in the denominator of the operator.

6.5.2 Covariant, contravariant, and Einstein summation

For general tensors, the Einstein summation convention assumes that labels are summed over their range when a lower *covariant* label matches an upper *contravariant* label. In this way we have an arbitrary vector in space written as

$$\vec{F}(\tau) = \sum_{a=1}^3 \vec{e}_a F^a = \vec{e}_a F^a. \quad (6.14)$$

This rule generalizes that used for Cartesian tensors in Chapter 1. For general tensors, contraction is between a lower and an upper label. We return to such contractions in Section 7.2, where we show that the contraction of a vector and a one-form renders a scalar.

The names *covariant* and *contravariant* originate from their relation to the labels placed on a coordinate basis vectors (Section 6.4.2), and thus how they change under coordinate transformations relative to how basis vectors change (Section 7.1). The covariant tensor label accords with the downstairs label placement for a coordinate basis, whereas the upstairs contravariant label is contrary to the coordinate basis (e.g., see Section 2.26 of [Schutz \(1980\)](#)). A useful mnemonic is “co-low” to signal that the covariant label is downstairs (“low”). The names covariant and contravariant are used infrequently in modern tensor analysis, which instead considers tensors as geometric objects and so is not primarily concerned with the coordinate representations of tensors.

6.5.3 The boldface notation

In Cartesian tensors we have no need to distinguish \vec{F} from the boldface \mathbf{F} . One means to extend this notation to general tensors is to organize the contravariant representation F^a into an ordered list (F^1, F^2, F^3) and to use the boldface notation

$$\mathbf{F} = (F^1, F^2, F^3). \quad (6.15)$$

Likewise, we can organize the basis vectors according to

$$\vec{\mathbf{e}} = (\vec{e}_1, \vec{e}_2, \vec{e}_3) \quad (6.16)$$

and the coordinates as

$$\boldsymbol{\xi} = (\xi^1, \xi^2, \xi^3). \quad (6.17)$$

With this notation the velocity vector representation from equation (6.10) takes on the form

$$\vec{v} = \vec{e}_a v^a = \vec{\mathbf{e}} \cdot \mathbf{v}. \quad (6.18)$$

Likewise, a trajectory can be represented in terms of a chosen set of coordinates according to

$$\mathcal{P} = \vec{e}_a \xi^a = \vec{\mathbf{e}} \cdot \boldsymbol{\xi}. \quad (6.19)$$

Notice that the arrow symbol over the basis vector remains even when using the boldface. This usage is required since the arrow carries information about the vector nature of the object, whereas the boldface is merely a shorthand for an ordered list.

Although the above notation makes good sense, we more generally allow the boldface to be synonymous with the vector arrow notation

$$\mathbf{F} = \vec{F} = \vec{e}_a F^a. \quad (6.20)$$

Though corrupting the convention in equation (6.18), this notation is readily used in the literature and as such will be employed in this book where the meaning is clear and after our tensor brain-muscle has been exercised.

6.5.4 Space-time approach

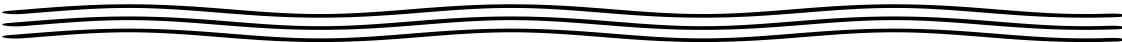
As introduced in Section 6.2.2, we make use of a Greek label when incorporating time to the tensor indices, with $\alpha = 0$ denoting the time coordinate. Time remains universal in the Newtonian world of geophysical fluid mechanics. However, many coordinates are functions of both space and time. Therefore, when measuring time changes in one coordinate system it will differ from time changes in another. Following equation (6.2) used for coordinates, we make use of the following index notation and ordered list for contravariant components of 4-vectors

$$F^\alpha = (F^0, F^1, F^2, F^3) = (F^0, F^a) = (F^0, \mathbf{F}). \quad (6.21)$$

In Cartesian coordinates, the time component of the velocity 4-vector is unity

$$v^\alpha = (1, v^1, v^2, v^3) = (1, v^a) = (1, \mathbf{v}) \quad \text{4-velocity in Cartesian coordinates.} \quad (6.22)$$

We make use of the 4-vector formalism in Section 7.9 when performing the transformation of partial derivatives. In that discussion we note that the time coordinate is not a function of space and yet the space coordinates are generally a function of time, which makes the 4-vector formalism of use to transform partial derivatives. Otherwise, we retain the space plus time approach (3+1) of classical mechanics.



General tensor analysis

In this chapter we generalize the Cartesian algebra of Chapter 1 and Cartesian calculus of Chapter 2 to general tensors. We focus on space tensors so that time typically remains outside of the tensor formalism. The single exception is that we consider a space-time transformation of the partial time derivative operator (Section 7.9), since the time derivative cares about the motion of space coordinates.

Cartesian tensor analysis provides a systematic formulation of vector analysis. The Euclidean metric provided by the Kronecker symbol (i.e., the identity tensor) underlies Cartesian tensors. Here, we extend the formalism of Cartesian tensors by allowing for an arbitrary spatial metric, and in so doing move into the realm of *Riemannian differential geometry*. A more mathematically deductive approach would be to first study calculus on topological manifolds that are not endowed with a metric. This topic is briefly introduced in Chapter 9 for the purpose of providing a mathematical foundation for water mass analysis as discussed in Section 52.12.

READER'S GUIDE TO THIS CHAPTER

This chapter is necessary for understanding the mathematics of generalized vertical coordinates in Chapters 43, 44, and 45. Otherwise, it can be skimmed or returned to later if the need arises. Material here is an updated version of Chapters 20 and 21 from [Griffies \(2004\)](#) along with some streamlining. Other resources include the treatment of tensors for fluid mechanics given by [Aris \(1962\)](#) and the physics treatment by [Thorne and Blandford \(2017\)](#).

7.1	The metric tensor and coordinate transformations	130
7.1.1	Cartesian coordinates in Euclidean space	130
7.1.2	The metric as a symmetric second order tensor	131
7.1.3	Coordinate representation of the metric tensor	131
7.1.4	Transforming the coordinate representation of the metric tensor	132
7.1.5	Basis vectors	132
7.1.6	Finite distance between points	133
7.2	One-forms	133
7.2.1	Coordinate representation of a one-form	133
7.2.2	Basis one-forms and the bi-orthogonality relation	134
7.2.3	Metric as a mapping between vectors and one-forms	134
7.2.4	Transformation of the coordinate representation	135
7.2.5	Arbitrary coordinate representation of inverse metric	135
7.2.6	Functions versus scalars	135
7.3	Scalar product	136
7.4	Worked example: oblique coordinates	136

7.4.1	Turning the crank	137
7.4.2	Comments	138
7.5	The volume element and Jacobian of transformation	138
7.5.1	Jacobian of transformation	139
7.5.2	Relating the Jacobian to the determinant of the metric	139
7.6	The permutation symbol and the determinant	140
7.6.1	Connecting the permutation symbol to the determinant	140
7.6.2	Further identities satisfied by the determinant	140
7.7	The Levi-Civita tensor and the volume element	141
7.7.1	General coordinate representation of the Levi-Civita tensor	141
7.7.2	The Levi-Civita tensor and the volume element	142
7.8	Vector cross product	142
7.9	Coordinate transformation of partial derivatives	142
7.9.1	The space-time transformation matrix	143
7.9.2	Determinant of the transformation matrix	143
7.9.3	Multiplying the transformation matrix and its inverse	144
7.9.4	Transformation of space and time partial derivatives	144
7.10	Covariant derivative of a scalar	145
7.11	Covariant derivative of a vector	145
7.11.1	Derivative of a vector	145
7.11.2	An alternative derivation	146
7.11.3	Christoffel symbols as the metric connection	146
7.12	Covariant derivative of a one-form	146
7.13	Covariant derivative of the metric tensor	147
7.14	Christoffel symbols in terms of the metric	147
7.15	Covariant divergence of a vector	147
7.15.1	Contraction of the Christoffel symbols	148
7.15.2	Exponential of the determinant	148
7.16	Covariant Laplacian of a scalar	148
7.17	Covariant curl of a vector	149
7.18	Gauss's divergence theorem	149
7.19	Stokes' theorem	149

7.1 The metric tensor and coordinate transformations

In the study of fluid mechanics we find the need to measure the distance between two points in space at a particular time instance. Since we assume all points live on a smooth and orientable manifold (e.g., a sphere, an isopycnal), it is sufficient to consider the distance between two infinitesimally close points and use integration to measure finite distances. The measurement of distance requires a metric tensor, which is the subject of this section.

7.1.1 Cartesian coordinates in Euclidean space

Consider a Cartesian coordinate representation for the spatial position of two points, with point \mathcal{P} having space coordinates $\xi^a = x^a$ and the other point \mathcal{Q} an infinitesimal distance away at $x^a + dx^a$. Furthermore, let

$$d\mathbf{x} = dx^a \vec{e}_a \quad (7.1)$$

be the infinitesimal space vector pointing from \mathcal{P} to \mathcal{Q} . Since the space is Euclidean, the squared distance between the two points is based on the Euclidean norm; i.e., the familiar scalar or dot product (Section 1.3)

$$ds^2 = d\mathbf{x} \cdot d\mathbf{x} = \vec{e}_a \cdot \vec{e}_b dx^a dx^b = \delta_{ab} dx^a dx^b. \quad (7.2)$$

In this expression,

$$(ds)^2 \equiv ds^2 \quad (7.3)$$

is the squared infinitesimal arc-length separating the two points. The Kronecker symbol, δ_{ab} , is symmetric

$$\delta_{ab} = \delta_{ba}, \quad (7.4)$$

and vanishes when $a \neq b$ and is unity when $a = b$

$$\delta_{ab} = \begin{cases} 0 & \text{if } a \neq b \\ 1 & \text{if } a = b. \end{cases} \quad (7.5)$$

The Kronecker symbol is a representation of the identity tensor.

7.1.2 The metric as a symmetric second order tensor

As defined by equation (7.2), δ_{ab} forms the Cartesian representation of the *metric tensor* for Euclidean space. The metric is a second order tensor, meaning that its coordinate representation carries two tensor labels. Contracting the metric tensor with two vectors leads to a number, namely the squared distance between the two points. Hence, the metric establishes the means to measure the distance between two points that live on a manifold.

We write this distance-measuring property of the metric tensor in a geometric manner through

$$\text{distance}(\vec{P}, \vec{Q}) = \sqrt{\mathcal{G}(\vec{P}, \vec{Q})}. \quad (7.6)$$

Here, \mathcal{G} is the metric tensor with coordinate representation g_{ab} and \vec{P}, \vec{Q} are infinitesimally close vectors with coordinate representations

$$\vec{P} = \xi^a \vec{e}_a \quad \vec{Q} = \vec{P} + d\xi^a \vec{e}_a. \quad (7.7)$$

Equation (7.6) indicates that the metric tensor takes two vectors as argument and produces a scalar. Furthermore, since

$$\text{distance}(\vec{P}, \vec{Q}) = \text{distance}(\vec{Q}, \vec{P}) \geq 0, \quad (7.8)$$

the metric tensor is a symmetric and positive tensor that produces zero only when $\vec{P} = \vec{Q}$.

7.1.3 Coordinate representation of the metric tensor

Given the geometric expression (7.6) for the metric, we determine its representation in an arbitrary coordinate system by considering the squared distance between the coordinate basis vectors

$$\text{distance}(\vec{e}_a, \vec{e}_b) = \sqrt{\mathcal{G}(\vec{e}_a, \vec{e}_b)}. \quad (7.9)$$

This relation determines the coordinate components of the metric tensor

$$\mathcal{G}(\vec{e}_a, \vec{e}_b) \equiv g_{ab}. \quad (7.10)$$

Furthermore, for a manifold embedded in Euclidean space this relation is written

$$g_{ab} = \vec{e}_a \cdot \vec{e}_b. \quad (7.11)$$

In this manner we see that the basis vectors determine the metric tensor components. Note that if the basis vectors are orthogonal, then the metric tensor components vanish unless $a = b$.

7.1.4 Transforming the coordinate representation of the metric tensor

We find opportunities to represent the metric tensor in various coordinate systems. Here, we consider the transformation from Cartesian coordinates $\xi^a = x^a$ to arbitrary coordinates $\xi^{\bar{a}}$. Use of the chain rule along with some index gymnastics leads to the equivalent expression for the squared infinitesimal length between two points

$$ds^2 = \delta_{ab} d\xi^a d\xi^b = \delta_{ab} \frac{\partial \xi^a}{\partial \xi^{\bar{a}}} \frac{\partial \xi^b}{\partial \xi^{\bar{b}}} d\xi^{\bar{a}} d\xi^{\bar{b}} \equiv \delta_{ab} \Lambda_{\bar{a}}^a \Lambda_{\bar{b}}^b d\xi^{\bar{a}} d\xi^{\bar{b}} \equiv g_{\bar{a}\bar{b}} d\xi^{\bar{a}} d\xi^{\bar{b}}, \quad (7.12)$$

where

$$g_{\bar{a}\bar{b}} = \delta_{ab} \Lambda_{\bar{a}}^a \Lambda_{\bar{b}}^b \quad (7.13)$$

defines the components to the metric tensor as represented by the new set of coordinates $\xi^{\bar{a}}$. We also introduced elements to the *transformation matrix*

$$\Lambda^a_{\bar{a}} = \frac{\partial \xi^a}{\partial \xi^{\bar{a}}}. \quad (7.14)$$

This matrix of partial derivatives has a non-zero entry when a coordinate in one representation changes while moving along the direction of a coordinate in the other representation. As for any partial derivative, the complement coordinates are held fixed when performing the derivative. Although carrying indices, the numbers $\Lambda^a_{\bar{a}}$ are *not* components of a tensor. Instead, they are components of a matrix used to transform tensor representations from one coordinate system to another. Organized as a matrix, we follow a convention whereby the row is denoted by a and the column by \bar{a}

$$\Lambda^a_{\bar{a}} = \begin{bmatrix} (\partial \xi^1 / \partial \xi^{\bar{1}})_{\bar{2},\bar{3}} & (\partial \xi^1 / \partial \xi^{\bar{2}})_{\bar{1},\bar{3}} & (\partial \xi^1 / \partial \xi^{\bar{3}})_{\bar{1},\bar{2}} \\ (\partial \xi^2 / \partial \xi^{\bar{1}})_{\bar{2},\bar{3}} & (\partial \xi^2 / \partial \xi^{\bar{2}})_{\bar{1},\bar{3}} & (\partial \xi^2 / \partial \xi^{\bar{3}})_{\bar{1},\bar{2}} \\ (\partial \xi^3 / \partial \xi^{\bar{1}})_{\bar{2},\bar{3}} & (\partial \xi^3 / \partial \xi^{\bar{2}})_{\bar{1},\bar{3}} & (\partial \xi^3 / \partial \xi^{\bar{3}})_{\bar{1},\bar{2}} \end{bmatrix}. \quad (7.15)$$

Note that the lower index is displaced to the right to delineate which index refers to the column. The transformation matrix is nonsingular for one-to-one invertible coordinate transformations, in which case its determinant, called the *Jacobian of the transformation*, is nonvanishing and single signed. Finally, we sometimes find it useful to write the un-barred coordinates as an ordered list in a column, $\boldsymbol{\xi} = (\xi^1, \xi^2, \xi^3)^T$, in which case the transformation matrix takes on the abbreviated form

$$\Lambda^a_{\bar{a}} = \begin{bmatrix} (\partial \boldsymbol{\xi} / \partial \xi^{\bar{1}})_{\bar{2},\bar{3}} & (\partial \boldsymbol{\xi} / \partial \xi^{\bar{2}})_{\bar{1},\bar{3}} & (\partial \boldsymbol{\xi} / \partial \xi^{\bar{3}})_{\bar{1},\bar{2}} \end{bmatrix}. \quad (7.16)$$

Use of this transformation matrix leads to the arbitrary coordinate representation of the metric tensor

$$g_{\bar{a}\bar{b}} = \vec{e}_{\bar{a}} \cdot \vec{e}_{\bar{b}} = \begin{bmatrix} \frac{\partial \boldsymbol{\xi}}{\partial \xi^{\bar{1}}} \cdot \frac{\partial \boldsymbol{\xi}}{\partial \xi^{\bar{1}}} & \frac{\partial \boldsymbol{\xi}}{\partial \xi^{\bar{1}}} \cdot \frac{\partial \boldsymbol{\xi}}{\partial \xi^{\bar{2}}} & \frac{\partial \boldsymbol{\xi}}{\partial \xi^{\bar{1}}} \cdot \frac{\partial \boldsymbol{\xi}}{\partial \xi^{\bar{3}}} \\ \frac{\partial \boldsymbol{\xi}}{\partial \xi^{\bar{2}}} \cdot \frac{\partial \boldsymbol{\xi}}{\partial \xi^{\bar{1}}} & \frac{\partial \boldsymbol{\xi}}{\partial \xi^{\bar{2}}} \cdot \frac{\partial \boldsymbol{\xi}}{\partial \xi^{\bar{2}}} & \frac{\partial \boldsymbol{\xi}}{\partial \xi^{\bar{2}}} \cdot \frac{\partial \boldsymbol{\xi}}{\partial \xi^{\bar{3}}} \\ \frac{\partial \boldsymbol{\xi}}{\partial \xi^{\bar{3}}} \cdot \frac{\partial \boldsymbol{\xi}}{\partial \xi^{\bar{1}}} & \frac{\partial \boldsymbol{\xi}}{\partial \xi^{\bar{3}}} \cdot \frac{\partial \boldsymbol{\xi}}{\partial \xi^{\bar{2}}} & \frac{\partial \boldsymbol{\xi}}{\partial \xi^{\bar{3}}} \cdot \frac{\partial \boldsymbol{\xi}}{\partial \xi^{\bar{3}}} \end{bmatrix}. \quad (7.17)$$

7.1.5 Basis vectors

We transform the basis vectors from Cartesian into arbitrary coordinates through the transformation

$$\vec{e}_{\bar{a}} = \Lambda^a_{\bar{a}} \vec{e}_a. \quad (7.18)$$

Use of the transformation matrix (7.16) renders the arbitrary coordinate basis vectors

$$\vec{e}_1 = \frac{\partial \xi}{\partial \xi^1} \quad \text{and} \quad \vec{e}_2 = \frac{\partial \xi}{\partial \xi^2} \quad \text{and} \quad \vec{e}_3 = \frac{\partial \xi}{\partial \xi^3}. \quad (7.19)$$

7.1.6 Finite distance between points

Once the metric is determined, the distance along a curve between two finitely separated points is given by integration

$$L = \int \sqrt{ds^2} = \int_{\varphi_1}^{\varphi_2} \left| g_{ab} \frac{d\xi^a}{d\varphi} \frac{d\xi^b}{d\varphi} \right|^{1/2} d\varphi, \quad (7.20)$$

where φ is a parameter specifying the curve (e.g., the arc length as in Section 2.4), with φ_1 and φ_2 specifying the curve's endpoints.

7.2 One-forms

The metric tensor \mathcal{G} is a function of two vectors. When the metric “eats” the two vectors, the result is the scalar distance between the vectors (equation (7.6))

$$\text{distance}(\vec{A}, \vec{B}) = \sqrt{\mathcal{G}(\vec{A}, \vec{B})}. \quad (7.21)$$

What if the metric only eats one vector? The resulting geometric object is known as a one-form

$$\tilde{A} \equiv \mathcal{G}(\vec{A},), \quad (7.22)$$

with the tilde used to distinguish a one-form from a vector.

7.2.1 Coordinate representation of a one-form

We can determine the coordinate representation of a one-form by eating a basis vector

$$\tilde{A}(\vec{e}_b) = \mathcal{G}(\vec{A}, \vec{e}_b) \quad (7.23a)$$

$$= \mathcal{G}(A^a \vec{e}_a, \vec{e}_b) \quad (7.23b)$$

$$= \mathcal{G}(\vec{e}_a, \vec{e}_b) A^a \quad (7.23c)$$

$$= g_{ab} A^a. \quad (7.23d)$$

To reach this result we pulled the coordinate representation A^a outside of the metric tensor since the tensor eats vectors rather than numbers. This equation defines the coordinate representation of the one-form \tilde{A} in terms of its *dual* vector \vec{A} and the metric tensor

$$A_b = g_{ab} A^a. \quad (7.24)$$

We thus see that the metric tensor provides the means to lower an index on the representation of a vector, thus producing the representation of a one-form. We return to this point in Section 7.2.3.

7.2.2 Basis one-forms and the bi-orthogonality relation

Just as for vectors, we find use for a basis of one-forms to specify their coordinate representation. The basis of one-forms, \tilde{e}^a , are defined through the bi-orthogonality relation

$$\mathcal{G}(\tilde{e}^a, \vec{e}_b) = \tilde{e}^a \cdot \vec{e}_b = \delta_b^a, \quad (7.25)$$

where

$$\delta_b^a = g^{ac} g_{cb} \quad (7.26)$$

are components to the Kronecker delta tensor, taking the value of unity when $a = b$ and zero otherwise

$$\delta_b^a = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}. \quad (7.27)$$

It is only for Cartesian coordinates that we have

$$\delta_c^a = g^{ab} \delta_{bc} \quad \text{Cartesian coordinates}, \quad (7.28)$$

which follows since $g^{ab} = \delta^{ab}$ in Cartesian coordinates.

We can obtain an explicit expression for the basis one-forms in arbitrary coordinates by transforming from Cartesian coordinates through use of the inverse transformation

$$\tilde{e}^{\bar{a}} = \Lambda_a^{\bar{a}} \tilde{e}^a, \quad (7.29)$$

which renders

$$\tilde{e}^{\bar{1}} = \hat{x} \frac{\partial \xi^{\bar{1}}}{\partial x} + \hat{y} \frac{\partial \xi^{\bar{1}}}{\partial y} + \hat{z} \frac{\partial \xi^{\bar{1}}}{\partial z} = \nabla \xi^{\bar{1}} \quad (7.30a)$$

$$\tilde{e}^{\bar{2}} = \hat{x} \frac{\partial \xi^{\bar{2}}}{\partial x} + \hat{y} \frac{\partial \xi^{\bar{2}}}{\partial y} + \hat{z} \frac{\partial \xi^{\bar{2}}}{\partial z} = \nabla \xi^{\bar{2}} \quad (7.30b)$$

$$\tilde{e}^{\bar{3}} = \hat{x} \frac{\partial \xi^{\bar{3}}}{\partial x} + \hat{y} \frac{\partial \xi^{\bar{3}}}{\partial y} + \hat{z} \frac{\partial \xi^{\bar{3}}}{\partial z} = \nabla \xi^{\bar{3}}. \quad (7.30c)$$

In Section 7.9.3 we verify that the basis one-forms satisfy the orthogonality relation (7.25) with the basis vectors

$$\tilde{e}^{\bar{a}} \cdot \vec{e}_{\bar{b}} = \delta_{\bar{b}}^{\bar{a}}. \quad (7.31)$$

7.2.3 Metric as a mapping between vectors and one-forms

We can contract the expression (7.24) with components of the inverse metric tensor, g^{ab} , to render

$$g^{ab} A_b = g^{ab} g_{bc} A^c = \delta_c^a A^c = A^a. \quad (7.32)$$

This identity, as well as equation (7.24), show that the metric provides a map between coordinate representations of one-forms and vectors.

To every vector \vec{A} there is a corresponding one-form \tilde{A} . We say that the one-forms and vectors are *dual*, with the mapping between one-forms and vectors rendered by the metric tensor. Abstractly, we say that a vector at a point on a manifold lives on the tangent space at that point, whereas a one-form lives on the cotangent space. The metric tensor attaches the tangent space to the co-tangent space. For Cartesian tensor analysis, the duality between one-forms and vectors is

the duality between row vectors and column vectors. Furthermore, as for Cartesian tensors, we construct an inner product by contracting one-forms and vectors to produce a scalar. Finally, the duality relation given by equation (7.24) offers us the means to raise and lower tensor indices in a manner akin to the transpose operation in linear algebra that produces a row vector from a column vector.

7.2.4 Transformation of the coordinate representation

The transformation matrix (7.14) provides the means to convert any arbitrary coordinate representation of a tensor from one coordinate system to another. For example, consider the coordinate representation of a vector, which is realized by letting the vector eat one of the basis one-forms

$$\vec{F}(\tilde{e}^a) = F^a. \quad (7.33)$$

Now consider another coordinate system with basis one-forms $\tilde{e}^{\bar{a}}$, so that the vector has a representation

$$\vec{F}(\tilde{e}^{\bar{a}}) = F^{\bar{a}}. \quad (7.34)$$

Transforming the basis one-form using the transformation matrix leads to

$$F^{\bar{a}} = \vec{F}(\tilde{e}^{\bar{a}}) = \vec{F}(\Lambda_a^{\bar{a}} \tilde{e}^a) = \Lambda_a^{\bar{a}} \vec{F}(\tilde{e}^a) = \Lambda_a^{\bar{a}} F^a. \quad (7.35)$$

Transformation of an arbitrary one-form representation takes place with the inverse transformation matrix

$$F_{\bar{a}} = \vec{F}(\tilde{e}_{\bar{a}}) = \vec{F}(\Lambda_{\bar{a}}^a \tilde{e}_a) = \Lambda_{\bar{a}}^a \vec{F}(\tilde{e}_a) = \Lambda_{\bar{a}}^a F_a. \quad (7.36)$$

7.2.5 Arbitrary coordinate representation of inverse metric

The inverse metric tensor has an arbitrary coordinate representation given by

$$g^{\bar{a}\bar{b}} = \tilde{e}^{\bar{a}} \cdot \tilde{e}^{\bar{b}} = \begin{bmatrix} \nabla\xi^1 \cdot \nabla\xi^1 & \nabla\xi^1 \cdot \nabla\xi^2 & \nabla\xi^1 \cdot \nabla\xi^3 \\ \nabla\xi^2 \cdot \nabla\xi^1 & \nabla\xi^2 \cdot \nabla\xi^2 & \nabla\xi^2 \cdot \nabla\xi^3 \\ \nabla\xi^3 \cdot \nabla\xi^1 & \nabla\xi^3 \cdot \nabla\xi^2 & \nabla\xi^3 \cdot \nabla\xi^3 \end{bmatrix}. \quad (7.37)$$

Proof that $g^{\bar{a}\bar{b}} g_{\bar{b}\bar{c}} = \delta_{\bar{c}}^{\bar{a}}$ requires use of the chain rule relations derived in Section 7.9.3.

7.2.6 Functions versus scalars

A scalar field carries no tensor labels and as such it does not transform under coordinate transformations. That is, a scalar is coordinate invariant and it is the simplest type of tensor field. Temperature, kinetic energy, and specific entropy are example scalar fields considered in this book. Now consider the first component of a vector field, F^1 , whose value at a point, \mathcal{P} and time, τ , is written $F^1(\mathcal{P}, \tau)$. Now change the coordinates so that the first component to the vector field at the same (\mathcal{P}, τ) is written $F^1(\mathcal{P}, \tau)$. Yet under a coordinate transformation we generally have $F^1(\mathcal{P}, \tau) \neq F^1(\mathcal{P}, \tau)$, so that the individual components to a vector field, though functions of space and time, *do not* individually transform as a scalar field. Instead, they transform as components to a vector field. This very basic example illustrates the distinction between functions of space and time, which includes any component of any tensor field, from scalar fields, which are coordinate invariant.

7.3 Scalar product

In Section 1.3.2 we defined the scalar product between two Cartesian vectors. The natural generalization is given by

$$\vec{P} \cdot \vec{Q} = P^a Q^b \vec{e}_a \cdot \vec{e}_b = P^a Q^b g_{ab} = P^a Q_a = P_b Q^b, \quad (7.38)$$

where the second equality made use of the metric tensor coordinate representation given by equation (7.11). We can conceive of the scalar product in a somewhat more general manner by recalling that a one-form operates on a vector, $\tilde{P}(\vec{Q})$. Conversely, a vector operates on a one-form, $\vec{Q}(\tilde{P})$. Exposing components leads to

$$\tilde{P}(\vec{Q}) = \tilde{P}(Q^a \vec{e}_a) = Q^a \tilde{P}(\vec{e}_a) = Q^a P_a, \quad (7.39)$$

which equals to

$$\vec{Q}(\tilde{P}) = \vec{Q}(P_a \tilde{e}^a) = P_a \vec{Q}(\tilde{e}^a) = P_a Q^a. \quad (7.40)$$

The scalar product is invariant to coordinate changes, as seen through

$$\vec{Q}(\tilde{P}) = \vec{Q}(P_a \tilde{e}^a) = P_a Q^a = \vec{Q}(P_{\bar{a}} \tilde{e}^{\bar{a}}) = P_{\bar{a}} Q^{\bar{a}}. \quad (7.41)$$

The invariance is also revealed by working just with the coordinate representations and introducing the transformation matrix elements

$$P_a Q^a = (\Lambda^{\bar{a}}_a P_{\bar{a}}) (\Lambda^a_{\bar{b}} Q^{\bar{b}}) = \Lambda^{\bar{a}}_a \Lambda^a_{\bar{b}} P_{\bar{a}} Q^{\bar{b}} = \delta^{\bar{a}}_{\bar{b}} P_{\bar{a}} Q^{\bar{b}} = P_{\bar{a}} Q^{\bar{a}}. \quad (7.42)$$

7.4 Worked example: oblique coordinates

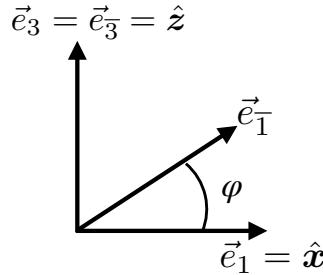


FIGURE 7.1: Oblique basis vectors for the x - z plane where $\vec{e}_{\bar{1}} = \vec{e}_1 \cos \varphi + \vec{e}_3 \sin \varphi$ and $\vec{e}_{\bar{3}} = \vec{e}_3$, with $\vec{e}_1 = \hat{x}$ and $\vec{e}_3 = \hat{z}$. These coordinate basis vectors are related to those used for generalized vertical coordinates shown in the left panel of Figure 43.2.

We pause here to consider the example of *oblique coordinates* for the x - z plane as specified by the basis vectors

$$\vec{e}_{\bar{1}} = \vec{e}_1 \cos \varphi + \vec{e}_3 \sin \varphi = \hat{x} \cos \varphi + \hat{z} \sin \varphi. \quad \text{and} \quad \vec{e}_{\bar{3}} = \vec{e}_3 = \hat{z}. \quad (7.43)$$

The oblique coordinate basis vectors, $\vec{e}_{\bar{a}}$, are orthogonal when the angle $\varphi = 0, \pi$; otherwise they are non-orthogonal. Also note that if $\varphi = \pi/2, 3\pi/2$ then $\vec{e}_{\bar{1}} = \pm \vec{e}_3$, in which case the vectors no longer form a basis for the x - z plane. So in the following we assume $\varphi \in (-\pi/2, \pi/2)$. These coordinates are oriented so that they correspond to the generalized vertical coordinate basis vectors depicted in Figure 43.2. Finally, for the purposes of this section we ignore the 2 direction and just work within the x - z plane. Tensor indices thus carry values 1 and 3 with 2 ignored.

7.4.1 Turning the crank

We here exhibit the results of turning the crank for the oblique coordinates.

Transformation matrix and its inverse

We can construct the transformation matrix through $\vec{e}_{\bar{a}} = \Lambda^a_{\bar{a}} \vec{e}_a$, and use of the coordinate basis definition (7.43)

$$\vec{e}_{\bar{1}} = \Lambda^1_{\bar{1}} \vec{e}_1 + \Lambda^3_{\bar{1}} \vec{e}_3 \implies \Lambda^1_{\bar{1}} = \cos \varphi, \quad \Lambda^3_{\bar{1}} = \sin \varphi \quad (7.44a)$$

$$\vec{e}_{\bar{3}} = \Lambda^1_{\bar{3}} \vec{e}_1 + \Lambda^3_{\bar{3}} \vec{e}_3 \implies \Lambda^1_{\bar{3}} = 0, \quad \Lambda^3_{\bar{3}} = 1, \quad (7.44b)$$

so that the transformation matrix is

$$\Lambda^a_{\bar{a}} = \begin{bmatrix} \Lambda^1_{\bar{1}} & \Lambda^1_{\bar{3}} \\ \Lambda^3_{\bar{1}} & \Lambda^3_{\bar{3}} \end{bmatrix} = \begin{bmatrix} \cos \varphi & 0 \\ \sin \varphi & 1 \end{bmatrix}, \quad (7.45)$$

and a matrix inversion yields the inverse transformation matrix

$$\Lambda^{\bar{a}}_a = \begin{bmatrix} \Lambda^{\bar{1}}_1 & \Lambda^{\bar{1}}_3 \\ \Lambda^{\bar{3}}_1 & \Lambda^{\bar{3}}_3 \end{bmatrix} = \frac{1}{\cos \varphi} \begin{bmatrix} 1 & 0 \\ -\sin \varphi & \cos \varphi \end{bmatrix}. \quad (7.46)$$

Basis one-forms

The basis one-forms using oblique coordinates are determined by

$$\tilde{e}^{\bar{a}} = \Lambda^{\bar{a}}_a \tilde{e}^a = \Lambda^{\bar{a}}_1 \tilde{e}^1 + \Lambda^{\bar{a}}_3 \tilde{e}^3 \quad (7.47)$$

with $\tilde{e}^1 = \hat{x}$ and $\tilde{e}^3 = \hat{z}$ for Cartesian coordinates. Making use of the inverse transformation matrix (7.46) leads to

$$\tilde{e}^{\bar{1}} = \Lambda^{\bar{1}}_1 \tilde{e}^1 + \Lambda^{\bar{1}}_3 \tilde{e}^3 = \frac{\hat{x}}{\cos \varphi} \quad (7.48a)$$

$$\tilde{e}^{\bar{3}} = \Lambda^{\bar{3}}_1 \tilde{e}^1 + \Lambda^{\bar{3}}_3 \tilde{e}^3 = -\hat{x} \tan \varphi + \hat{z}. \quad (7.48b)$$

We can readily verify the bi-orthogonality relation (7.25) whereby

$$\vec{e}_{\bar{a}} \cdot \tilde{e}^{\bar{b}} = \delta_{\bar{a}}^{\bar{b}}. \quad (7.49)$$

Representing a vector

The inverse transformation matrix (7.46) can be used to derive the relation between the Cartesian coordinate representation of an arbitrary vector, $\vec{P} = P^a \vec{e}_a$, and the oblique coordinate representation, $\vec{P} = P^{\bar{a}} \vec{e}_{\bar{a}}$. Doing so leads to the contravariant components written using oblique coordinates

$$P^{\bar{a}} = \Lambda^{\bar{a}}_a P^a = \Lambda^{\bar{a}}_1 P^1 + \Lambda^{\bar{a}}_3 P^3 \implies P^{\bar{1}} = \frac{1}{\cos \varphi} P^1 \text{ and } P^{\bar{3}} = -\tan \varphi P^1 + P^3. \quad (7.50)$$

Likewise, the covariant representation can be found by using the transformation matrix (7.45) to render

$$P_{\bar{a}} = \Lambda^a_{\bar{a}} P_a = \Lambda^1_{\bar{a}} P_1 + \Lambda^3_{\bar{a}} P_3 \implies P_{\bar{1}} = P_1 \cos \varphi + P_3 \sin \varphi \text{ and } P_{\bar{3}} = P_3, \quad (7.51)$$

where $P^a = P_a$ for the Cartesian coordinate representation.

Representing the metric tensor

The covariant representation of the metric tensor is given by

$$g_{\bar{a}\bar{b}} = \delta_{ab} \Lambda^a_{\bar{a}} \Lambda^b_{\bar{b}} = \begin{bmatrix} g_{\bar{1}\bar{1}} & g_{\bar{1}\bar{3}} \\ g_{\bar{3}\bar{1}} & g_{\bar{3}\bar{3}} \end{bmatrix} = \begin{bmatrix} 1 & \sin \varphi \\ \sin \varphi & 1 \end{bmatrix}, \quad (7.52)$$

and its inverse is

$$g^{\bar{a}\bar{b}} = \delta^{ab} \Lambda^{\bar{a}}_a \Lambda^{\bar{b}}_b = \begin{bmatrix} g^{\bar{1}\bar{1}} & g^{\bar{1}\bar{3}} \\ g^{\bar{3}\bar{1}} & g^{\bar{3}\bar{3}} \end{bmatrix} = \frac{1}{(\cos \varphi)^2} \begin{bmatrix} 1 & -\sin \varphi \\ -\sin \varphi & 1 \end{bmatrix}. \quad (7.53)$$

Scalar product

The scalar product is given by

$$\vec{P} \cdot \vec{P} = g_{\bar{a}\bar{b}} P^{\bar{a}} P^{\bar{b}} \quad (7.54a)$$

$$= P_{\bar{b}} P^{\bar{b}} \quad (7.54b)$$

$$= (P_1 \cos \varphi + P_3 \sin \varphi) \frac{1}{\cos \varphi} P^1 + P_3 (P^3 - \tan \varphi P^1) \quad (7.54c)$$

$$= P_1 P^1 + P_3 P^3 \quad (7.54d)$$

$$= \delta_{ab} P^a P^b. \quad (7.54e)$$

7.4.2 Comments

Oblique coordinate offer a pedagogical step towards the more complex case of generalized vertical coordinates studied in Chapter 43. Indeed, much of the tensor algebra needed for generalized vertical coordinates is concisely summarized in the above steps using oblique coordinates.

7.5 The volume element and Jacobian of transformation

Recall from Section 1.5.2 that we derived an expression for the volume of an infinitesimal region of Euclidean space, \mathbb{R}^3 , using Cartesian coordinates

$$dV = dx dy dz (\hat{x} \wedge \hat{y}) \cdot \hat{z} = dx dy dz. \quad (7.55)$$

This volume element is used for integrating over a region of \mathbb{R}^3 when using Cartesian coordinates. Furthermore, its material fluid expression measures the volume of a fluid element. We now generalize this result to arbitrary coordinates.

7.5.1 Jacobian of transformation

From multi-variate calculus, the relation between $d\xi^1 d\xi^2 d\xi^3$ and $d\xi^{\bar{1}} d\xi^{\bar{2}} d\xi^{\bar{3}}$ for two sets of coordinates is given by

$$d\xi^1 d\xi^2 d\xi^3 = \frac{\partial \xi}{\partial \bar{\xi}} d\xi^{\bar{1}} d\xi^{\bar{2}} d\xi^{\bar{3}} \quad (7.56a)$$

$$= \frac{\partial(\xi^1, \xi^2, \xi^3)}{\partial(\xi^{\bar{1}}, \xi^{\bar{2}}, \xi^{\bar{3}})} d\xi^{\bar{1}} d\xi^{\bar{2}} d\xi^{\bar{3}} \quad (7.56b)$$

$$= \left[\frac{\partial \xi}{\partial \xi^{\bar{1}}} \wedge \frac{\partial \xi}{\partial \xi^{\bar{2}}} \right] \cdot \frac{\partial \xi}{\partial \xi^{\bar{3}}} d\xi^{\bar{1}} d\xi^{\bar{2}} d\xi^{\bar{3}} \quad (7.56c)$$

$$= \det(\Lambda^a_{\bar{a}}) d\xi^{\bar{1}} d\xi^{\bar{2}} d\xi^{\bar{3}}, \quad (7.56d)$$

where $\det(\Lambda^a_{\bar{a}})$ is the determinant of the transformation matrix, also known as the *Jacobian of transformation*. The transformation is well defined so long as the Jacobian does not vanish. We maintain labels on the transformation matrix inside the determinant symbol to help indicate the sense for the transformation. This notation also helps maintain proper conservation of tensor indices.

7.5.2 Relating the Jacobian to the determinant of the metric

Recall the expression (7.12) for the transformation of the metric

$$g_{\bar{a}\bar{b}} = \Lambda^a_{\bar{a}} \Lambda^b_{\bar{b}} g_{ab}. \quad (7.57)$$

We can write this expression as a matrix equation

$$\bar{\mathcal{G}} = \Lambda^T \mathcal{G} \Lambda \quad (7.58)$$

where Λ^T is the transposed matrix. Taking determinants of both sides yields

$$\det(\bar{\mathcal{G}}) = \det(\Lambda^T \mathcal{G} \Lambda) = \det(\Lambda^T) \det(\mathcal{G}) \det(\Lambda) = [\det(\Lambda)]^2 \det(\mathcal{G}). \quad (7.59)$$

To reach this result we used the property of determinants that $\det(AB) = \det(A)\det(B)$ for any two matrices, and $\det(\Lambda^T) = \det(\Lambda)$. Consequently,

$$\det(\Lambda^a_{\bar{a}}) = \frac{\sqrt{\det(g_{\bar{a}\bar{b}})}}{\sqrt{\det(g_{ab})}} = \frac{\sqrt{\det(\bar{\mathcal{G}})}}{\sqrt{\det(\mathcal{G})}}. \quad (7.60)$$

We are thus led to the equivalent expressions for the volume element

$$dV \equiv \sqrt{\det(\mathcal{G})} d\xi^1 d\xi^2 d\xi^3 = \sqrt{\det(\bar{\mathcal{G}})} d\xi^{\bar{1}} d\xi^{\bar{2}} d\xi^{\bar{3}}. \quad (7.61)$$

This relation provides us with our desired general coordinate expression for the volume element. For the special case when the unbarred coordinates are Cartesian, $g_{ab} = \delta_{ab}$ so that $\det(\mathcal{G}) = 1$ and

$$\det(\Lambda^a_{\bar{a}}) = \sqrt{\det(g_{\bar{a}\bar{b}})} \quad \text{unbarred coordinates are Cartesian.} \quad (7.62)$$

This is a rather useful expression for our purposes, since we can always use Cartesian as the unbarred coordinates given that geophysical fluids move in a background Euclidean space.

7.6 The permutation symbol and the determinant

As discussed in Section 1.4.1, the Cartesian components of the Levi-Civita tensor are given by the permutation symbol, ϵ_{abc} . To help determine the general coordinate representation of the Levi-Civita tensor, we here develop some identities satisfied by the determinant of the transformation matrix.

7.6.1 Connecting the permutation symbol to the determinant

Consider a two-dimensional space with a transformation matrix $\Lambda^a_{\bar{a}}$ between two sets of coordinates. The determinant of the transformation is given by

$$\det(\Lambda^a_{\bar{a}}) = \Lambda^1_{\bar{1}} \Lambda^2_{\bar{2}} - \Lambda^1_{\bar{2}} \Lambda^2_{\bar{1}}. \quad (7.63)$$

Introducing the permutation symbol ϵ_{ab} allows us to write this expression in a more tidy manner

$$\det(\Lambda^a_{\bar{a}}) = \epsilon_{ab} \Lambda^a_{\bar{1}} \Lambda^b_{\bar{2}} \quad (7.64)$$

with

$$\epsilon_{12} = 1 \quad \text{and} \quad \epsilon_{21} = -1. \quad (7.65)$$

The permutation symbol is defined to have numerically the same values whether the labels are raised or lowered: $\epsilon^{ab} = \epsilon_{ab}$.

We can generalize the above to any number of dimensions, each of which adds one more label to the permutation symbol and one more number added to the permutation string. We already encountered the three dimensional version in Section 1.4.1 when discussing the vector cross product, in which case the permutation symbol is

$$\epsilon_{123} = 1 \quad (7.66a)$$

$$\epsilon_{abc} = \begin{cases} 0 & \text{if any two labels are the same,} \\ 1 & \text{if } a, b, c \text{ is an even permutation of 1, 2, 3,} \\ -1 & \text{if } a, b, c \text{ is an odd permutation of 1, 2, 3.} \end{cases} \quad (7.66b)$$

Likewise, the determinant of the transformation matrix takes the form

$$\det(\Lambda^a_{\bar{a}}) = \frac{\partial \xi}{\partial \bar{\xi}} = \frac{\partial(\xi^1, \xi^2, \xi^3)}{\partial(\xi^{\bar{1}}, \xi^{\bar{2}}, \xi^{\bar{3}})} = \epsilon_{abc} \Lambda^a_{\bar{1}} \Lambda^b_{\bar{2}} \Lambda^c_{\bar{3}}. \quad (7.67)$$

7.6.2 Further identities satisfied by the determinant

The following identity in two dimensions can be readily verified through enumeration

$$\epsilon_{ab} \Lambda^a_{\bar{a}} \Lambda^b_{\bar{b}} = \epsilon_{\bar{a}\bar{b}} \det(\Lambda^a_{\bar{a}}), \quad (7.68)$$

which follows directly from the definition of the determinant and can be explicitly verified so long as we assume the permutation symbol $\epsilon_{\bar{a}\bar{b}}$ is numerically identical to ϵ_{ab} . Now contract both sides of this relation with $\epsilon^{\bar{a}\bar{b}}$ to isolate the determinant

$$\frac{1}{2} \epsilon^{\bar{a}\bar{b}} \epsilon_{ab} \Lambda^a_{\bar{a}} \Lambda^b_{\bar{b}} = \det(\Lambda^a_{\bar{a}}), \quad (7.69)$$

where we used

$$\epsilon^{\bar{a}\bar{b}} \epsilon_{ab} = \epsilon^{\bar{1}\bar{2}} \epsilon_{12} + \epsilon^{\bar{2}\bar{1}} \epsilon_{21} = 2. \quad (7.70)$$

The three dimensional version takes the form

$$\epsilon_{abc} \Lambda^a_{\bar{a}} \Lambda^b_{\bar{b}} \Lambda^c_{\bar{c}} = \epsilon_{\bar{a}\bar{b}\bar{c}} \det(\Lambda^a_{\bar{a}}), \quad (7.71)$$

so that

$$\frac{1}{3!} \epsilon^{\bar{a}\bar{b}\bar{c}} \epsilon_{abc} \Lambda^a_{\bar{a}} \Lambda^b_{\bar{b}} \Lambda^c_{\bar{c}} = \det(\Lambda^a_{\bar{a}}). \quad (7.72)$$

7.7 The Levi-Civita tensor and the volume element

The metric tensor introduced in Section 7.1 provides a means to measure distance between two points. The Levi-Civita tensor allows us to compute volumes (or areas for two dimensional manifolds). We make particular use of this tensor to compute the volume element used for integration. This section generalizes the Cartesian coordinate discussion provided in Section 1.5.3.

7.7.1 General coordinate representation of the Levi-Civita tensor

The relations (7.68) and (7.71) indicate that the permutation symbol *does not* transform as the components to a second order covariant tensor, unless the determinant of the transformation is unity. Unit determinants occur for special transformations, such as rotations (i.e., Cartesian to Cartesian coordinate transformation as in Chapter 1) and the identity transformation. Indeed, we have already noted that the permutation symbol has the same representation regardless the coordinate choice. As we now show, the permutation symbol is the Cartesian coordinate representation of the Levi-Civita tensor.

The above relations for the determinant motivate us to introduce the general coordinate form of the *Levi-Civita tensor*

$$\varepsilon_{abc} = \sqrt{\det(\mathcal{G})} \epsilon_{abc}. \quad (7.73)$$

We highlight the distinct symbols in this definition, with ε the Levi-Civita tensor and ϵ the permutation symbol. By construction, the components to the Levi-Civita tensor transform as

$$\Lambda^a_{\bar{a}} \Lambda^b_{\bar{b}} \Lambda^c_{\bar{c}} \varepsilon_{abc} = \Lambda^a_{\bar{a}} \Lambda^b_{\bar{b}} \Lambda^c_{\bar{c}} \sqrt{\det(\mathcal{G})} \epsilon_{abc} \quad (7.74a)$$

$$= \sqrt{\det(\mathcal{G})} \epsilon_{\bar{a}\bar{b}\bar{c}} \det(\Lambda^a_{\bar{a}}) \quad (7.74b)$$

$$= \sqrt{\det(\mathcal{G})} \epsilon_{\bar{a}\bar{b}\bar{c}} \quad (7.74c)$$

$$= \varepsilon_{\bar{a}\bar{b}\bar{c}}, \quad (7.74d)$$

where equations (7.60) and (7.68) were used. Therefore, ε_{abc} transforms as components to a third order covariant tensor. Likewise,

$$\varepsilon^{abc} = \frac{\epsilon^{abc}}{\sqrt{\det(\mathcal{G})}} \quad (7.75)$$

transforms as the components to a third order contravariant tensor. These transformation rules allow us to identify ε as a tensor rather than just a combination of numbers.

7.7.2 The Levi-Civita tensor and the volume element

As a third order tensor, the Levi-Civita tensor takes three vectors as its argument. In particular, for three infinitesimal vectors we have

$$\varepsilon(\vec{e}_1 d\xi^1, \vec{e}_2 d\xi^2, \vec{e}_3 d\xi^3) = d\xi^1 d\xi^2 d\xi^3 \varepsilon(\vec{e}_1, \vec{e}_2, \vec{e}_3) \quad (7.76a)$$

$$= d\xi^1 d\xi^2 d\xi^3 \varepsilon_{123} \quad (7.76b)$$

$$= d\xi^1 d\xi^2 d\xi^3 \sqrt{\det(\mathcal{G})} \epsilon_{123} \quad (7.76c)$$

$$= dV, \quad (7.76d)$$

where we used equation (7.61) for the final step. This result means that geometrically, the Levi-Civita tensor measures the volume defined by three vectors

$$\varepsilon(\vec{A}, \vec{B}, \vec{C}) = \text{volume}(\vec{A}, \vec{B}, \vec{C}). \quad (7.77)$$

This interpretation accords with the Cartesian coordinate discussion of the Levi-Civita tensor in Section 1.5.3.

7.8 Vector cross product

The vector cross product of two Cartesian basis vectors yields the third, so that

$$\hat{x} \wedge \hat{y} = \hat{z} \quad \text{and cyclic permutations.} \quad (7.78)$$

The coordinate invariant generalization of this relation is given by

$$\vec{e}_a \wedge \vec{e}_b \equiv \varepsilon_{abc} \tilde{e}^c. \quad (7.79)$$

That is, the vector cross product of two vectors leads to a one-form. We are thus led to the general coordinate expression for the vector cross product of two arbitrary vectors

$$\vec{P} \wedge \vec{Q} = P^a Q^b \vec{e}_a \wedge \vec{e}_b \quad (7.80a)$$

$$= P^a Q^b \varepsilon_{abc} \tilde{e}^c. \quad (7.80b)$$

7.9 Coordinate transformation of partial derivatives

Throughout this book, the background space is Euclidean and time is universal. We are thus concerned with space tensors rather than the space-time tensors of special and general relativity. Nonetheless, our description of points in space generally makes use of curved surfaces that are time dependent. Curved surfaces motivate the use of general tensors. Time dependence motivates a space-time formulation (Section 6.5.4) for the purpose of transforming the partial time derivative operator. In this section we establish some properties of the space-time transformation matrix and then exhibit the transformation of space and time partial time derivatives. We have further use of a space-time formulation in Sections 13.6 and 13.7, also for considering the transformation of partial derivatives.

7.9.1 The space-time transformation matrix

As discussed in Section 7.1.4, transformations between coordinate representations are enabled by the transformation matrix built from partial derivatives of the coordinate transformations. The transformation matrix with a universal Newtonian time plus spatial coordinates (that are functions of space and time) takes on the form

$$\Lambda^{\alpha}_{\bar{\alpha}} = \frac{\partial \xi^{\alpha}}{\partial \xi^{\bar{\alpha}}} = \begin{bmatrix} \frac{\partial \xi^0}{\partial \xi^0} & \frac{\partial \xi^0}{\partial \xi^1} & \frac{\partial \xi^0}{\partial \xi^2} & \frac{\partial \xi^0}{\partial \xi^3} \\ \frac{\partial \xi^0}{\partial \xi^1} & \frac{\partial \xi^1}{\partial \xi^1} & \frac{\partial \xi^1}{\partial \xi^2} & \frac{\partial \xi^1}{\partial \xi^3} \\ \frac{\partial \xi^1}{\partial \xi^0} & \frac{\partial \xi^1}{\partial \xi^1} & \frac{\partial \xi^1}{\partial \xi^2} & \frac{\partial \xi^1}{\partial \xi^3} \\ \frac{\partial \xi^0}{\partial \xi^2} & \frac{\partial \xi^2}{\partial \xi^1} & \frac{\partial \xi^2}{\partial \xi^2} & \frac{\partial \xi^2}{\partial \xi^3} \\ \frac{\partial \xi^2}{\partial \xi^0} & \frac{\partial \xi^2}{\partial \xi^1} & \frac{\partial \xi^2}{\partial \xi^2} & \frac{\partial \xi^2}{\partial \xi^3} \\ \frac{\partial \xi^0}{\partial \xi^3} & \frac{\partial \xi^3}{\partial \xi^1} & \frac{\partial \xi^3}{\partial \xi^2} & \frac{\partial \xi^3}{\partial \xi^3} \\ \frac{\partial \xi^3}{\partial \xi^0} & \frac{\partial \xi^3}{\partial \xi^1} & \frac{\partial \xi^3}{\partial \xi^2} & \frac{\partial \xi^3}{\partial \xi^3} \\ \frac{\partial \xi^0}{\partial \xi^1} & \frac{\partial \xi^1}{\partial \xi^1} & \frac{\partial \xi^2}{\partial \xi^2} & \frac{\partial \xi^3}{\partial \xi^3} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ \frac{\partial \xi^1}{\partial \xi^0} & \frac{\partial \xi^1}{\partial \xi^1} & \frac{\partial \xi^1}{\partial \xi^2} & \frac{\partial \xi^1}{\partial \xi^3} \\ \frac{\partial \xi^2}{\partial \xi^0} & \frac{\partial \xi^2}{\partial \xi^1} & \frac{\partial \xi^2}{\partial \xi^2} & \frac{\partial \xi^2}{\partial \xi^3} \\ \frac{\partial \xi^3}{\partial \xi^0} & \frac{\partial \xi^3}{\partial \xi^1} & \frac{\partial \xi^3}{\partial \xi^2} & \frac{\partial \xi^3}{\partial \xi^3} \\ \frac{\partial \xi^0}{\partial \xi^0} & \frac{\partial \xi^1}{\partial \xi^1} & \frac{\partial \xi^2}{\partial \xi^2} & \frac{\partial \xi^3}{\partial \xi^3} \\ \frac{\partial \xi^1}{\partial \xi^0} & \frac{\partial \xi^1}{\partial \xi^1} & \frac{\partial \xi^1}{\partial \xi^2} & \frac{\partial \xi^1}{\partial \xi^3} \\ \frac{\partial \xi^2}{\partial \xi^0} & \frac{\partial \xi^2}{\partial \xi^1} & \frac{\partial \xi^2}{\partial \xi^2} & \frac{\partial \xi^2}{\partial \xi^3} \\ \frac{\partial \xi^3}{\partial \xi^0} & \frac{\partial \xi^3}{\partial \xi^1} & \frac{\partial \xi^3}{\partial \xi^2} & \frac{\partial \xi^3}{\partial \xi^3} \end{bmatrix}. \quad (7.81)$$

The final equality made use of our assumption that $\xi^0 = \xi^0$ since the time coordinate remains universal. Hence, when computing $\partial \xi^0 / \partial \xi^{\bar{\alpha}}$ we keep ξ^0 fixed so that the derivative vanishes as in the specific case of

$$\left[\frac{\partial \xi^0}{\partial \xi^{\bar{\alpha}}} \right]_{\xi^0, \xi^1, \xi^2, \xi^3} = \left[\frac{\partial \xi^0}{\partial \xi^{\bar{\alpha}}} \right]_{\xi^0, \xi^1, \xi^2, \xi^3} = 0. \quad (7.82)$$

Zero elements in the first row of the transformation matrix (7.81) reveals that time is not a function of space

$$\frac{\partial \xi^0}{\partial \xi^{\bar{\alpha}}} = 0. \quad (7.83)$$

In contrast, nonzero elements in the first column indicate that our description of space is a function of time

$$\frac{\partial \xi^a}{\partial \xi^0} \neq 0. \quad (7.84)$$

We see the same overall structure in the inverse space-time transformation matrix

$$\Lambda^{\bar{\alpha}}_{\alpha} = \frac{\partial \xi^{\bar{\alpha}}}{\partial \xi^{\alpha}} = \begin{bmatrix} \frac{\partial \xi^0}{\partial \xi^0} & \frac{\partial \xi^0}{\partial \xi^1} & \frac{\partial \xi^0}{\partial \xi^2} & \frac{\partial \xi^0}{\partial \xi^3} \\ \frac{\partial \xi^0}{\partial \xi^1} & \frac{\partial \xi^1}{\partial \xi^1} & \frac{\partial \xi^1}{\partial \xi^2} & \frac{\partial \xi^1}{\partial \xi^3} \\ \frac{\partial \xi^1}{\partial \xi^0} & \frac{\partial \xi^1}{\partial \xi^1} & \frac{\partial \xi^1}{\partial \xi^2} & \frac{\partial \xi^1}{\partial \xi^3} \\ \frac{\partial \xi^0}{\partial \xi^2} & \frac{\partial \xi^2}{\partial \xi^1} & \frac{\partial \xi^2}{\partial \xi^2} & \frac{\partial \xi^2}{\partial \xi^3} \\ \frac{\partial \xi^2}{\partial \xi^0} & \frac{\partial \xi^2}{\partial \xi^1} & \frac{\partial \xi^2}{\partial \xi^2} & \frac{\partial \xi^2}{\partial \xi^3} \\ \frac{\partial \xi^0}{\partial \xi^3} & \frac{\partial \xi^3}{\partial \xi^1} & \frac{\partial \xi^3}{\partial \xi^2} & \frac{\partial \xi^3}{\partial \xi^3} \\ \frac{\partial \xi^3}{\partial \xi^0} & \frac{\partial \xi^3}{\partial \xi^1} & \frac{\partial \xi^3}{\partial \xi^2} & \frac{\partial \xi^3}{\partial \xi^3} \\ \frac{\partial \xi^0}{\partial \xi^1} & \frac{\partial \xi^1}{\partial \xi^1} & \frac{\partial \xi^2}{\partial \xi^2} & \frac{\partial \xi^3}{\partial \xi^3} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ \frac{\partial \xi^1}{\partial \xi^0} & \frac{\partial \xi^1}{\partial \xi^1} & \frac{\partial \xi^1}{\partial \xi^2} & \frac{\partial \xi^1}{\partial \xi^3} \\ \frac{\partial \xi^2}{\partial \xi^0} & \frac{\partial \xi^2}{\partial \xi^1} & \frac{\partial \xi^2}{\partial \xi^2} & \frac{\partial \xi^2}{\partial \xi^3} \\ \frac{\partial \xi^3}{\partial \xi^0} & \frac{\partial \xi^3}{\partial \xi^1} & \frac{\partial \xi^3}{\partial \xi^2} & \frac{\partial \xi^3}{\partial \xi^3} \\ \frac{\partial \xi^0}{\partial \xi^0} & \frac{\partial \xi^1}{\partial \xi^1} & \frac{\partial \xi^2}{\partial \xi^2} & \frac{\partial \xi^3}{\partial \xi^3} \\ \frac{\partial \xi^1}{\partial \xi^0} & \frac{\partial \xi^1}{\partial \xi^1} & \frac{\partial \xi^1}{\partial \xi^2} & \frac{\partial \xi^1}{\partial \xi^3} \\ \frac{\partial \xi^2}{\partial \xi^0} & \frac{\partial \xi^2}{\partial \xi^1} & \frac{\partial \xi^2}{\partial \xi^2} & \frac{\partial \xi^2}{\partial \xi^3} \\ \frac{\partial \xi^3}{\partial \xi^0} & \frac{\partial \xi^3}{\partial \xi^1} & \frac{\partial \xi^3}{\partial \xi^2} & \frac{\partial \xi^3}{\partial \xi^3} \end{bmatrix}. \quad (7.85)$$

7.9.2 Determinant of the transformation matrix

The determinant of the space-time transformation and its inverse remains identical to the determinant of their purely space portions

$$\det(\Lambda^{\alpha}_{\bar{\alpha}}) = \det(\Lambda^a_{\bar{a}}) \quad \text{and} \quad \det(\Lambda^{\bar{\alpha}}_{\alpha}) = \det(\Lambda^{\bar{a}}_a), \quad (7.86)$$

which follows since the first row in both transformations has only a single non-zero value, $\Lambda^0_0 = 1$ and $\Lambda^{\bar{0}}_0 = 1$. Hence, the relations developed in Sections 7.5 and 7.6 for the volume element and Jacobian of transformation remain unchanged when adding the universal time coordinate.

7.9.3 Multiplying the transformation matrix and its inverse

We here verify that the transformation matrix (7.81) indeed has its inverse given by (7.85). For this purpose we must prove that

$$\delta_{\beta}^{\alpha} = \Lambda_{\beta}^{\alpha} \Lambda_{\beta}^{\bar{\beta}} \quad \text{and} \quad \delta_{\bar{\beta}}^{\bar{\alpha}} = \Lambda_{\bar{\beta}}^{\bar{\alpha}} \Lambda_{\bar{\beta}}^{\beta}, \quad (7.87)$$

where δ_{β}^{α} and $\delta_{\bar{\beta}}^{\bar{\alpha}}$ are components to the identity tensor. The proof relies on writing the space-time coordinate transformation as a composite function

$$\xi^{\bar{\alpha}} = \xi^{\bar{\alpha}}(\xi^{\alpha}) = \xi^{\bar{\alpha}}[\xi^{\alpha}(\xi^{\bar{\beta}})]. \quad (7.88)$$

Taking partial derivatives and using the chain rule renders

$$\delta_{\bar{\beta}}^{\bar{\alpha}} = \frac{\partial \xi^{\bar{\alpha}}}{\partial \xi^{\bar{\beta}}} = \frac{\partial \xi^{\bar{\alpha}}}{\partial \xi^{\alpha}} \frac{\partial \xi^{\alpha}}{\partial \xi^{\bar{\beta}}} = \Lambda_{\alpha}^{\bar{\alpha}} \Lambda_{\beta}^{\alpha} \quad \text{and} \quad \delta_{\beta}^{\alpha} = \frac{\partial \xi^{\alpha}}{\partial \xi^{\beta}} = \frac{\partial \xi^{\alpha}}{\partial \xi^{\bar{\alpha}}} \frac{\partial \xi^{\bar{\alpha}}}{\partial \xi^{\beta}} = \Lambda_{\bar{\alpha}}^{\alpha} \Lambda_{\beta}^{\bar{\alpha}}. \quad (7.89)$$

Furthermore, the space subcomponents decouple from time, which can be seen by considering a few representative cases

$$1 = \delta_0^0 = \Lambda_{\bar{\alpha}}^0 \Lambda_{\alpha}^{\bar{\alpha}} = \Lambda_{\bar{0}}^0 \Lambda_{\alpha}^{\bar{0}} \quad (7.90a)$$

$$1 = \delta_1^1 = \Lambda_{\bar{\alpha}}^1 \Lambda_{\alpha}^{\bar{\alpha}} = \Lambda_{\bar{1}}^1 \Lambda_{\alpha}^{\bar{1}} \quad (7.90b)$$

$$0 = \delta_1^0 = \Lambda_{\bar{\alpha}}^0 \Lambda_{\alpha}^{\bar{\alpha}} = \Lambda_{\bar{0}}^0 \Lambda_{\alpha}^{\bar{0}} \quad (7.90c)$$

$$0 = \delta_2^1 = \Lambda_{\bar{\alpha}}^1 \Lambda_{\alpha}^{\bar{\alpha}} = \Lambda_{\bar{1}}^1 \Lambda_{\alpha}^{\bar{1}}. \quad (7.90d)$$

Consequently, the spatial components satisfy

$$\delta_b^a = \Lambda_{\bar{b}}^a \Lambda_{\bar{b}}^{\bar{b}} \quad \text{and} \quad \delta_{\bar{b}}^{\bar{a}} = \Lambda_{\bar{b}}^{\bar{a}} \Lambda_{\bar{b}}^b, \quad (7.91)$$

which allows for a splitting of the spatial components from the time component.

7.9.4 Transformation of space and time partial derivatives

Application of the chain rule leads to the transformation of the partial derivative operator

$$\partial_{\bar{\alpha}} = \frac{\partial}{\partial \xi^{\bar{\alpha}}} = \frac{\partial \xi^{\alpha}}{\partial \xi^{\bar{\alpha}}} \frac{\partial}{\partial \xi^{\alpha}} = \Lambda_{\bar{\alpha}}^{\alpha} \partial_{\alpha}. \quad (7.92)$$

Extracting the time and space components from the transformation matrix (7.81) yields

$$\partial_{\bar{0}} = \Lambda_{\bar{0}}^{\alpha} \partial_{\alpha} = \partial_0 + \Lambda_{\bar{0}}^a \partial_a \quad (7.93a)$$

$$\partial_{\bar{a}} = \Lambda_{\bar{a}}^{\alpha} \partial_{\alpha} = \Lambda_{\bar{a}}^a \partial_a. \quad (7.93b)$$

Notably, the time derivative operator in one coordinate system transforms into both space and time derivative operators in the new coordinate system. We expect this result since time changes in one coordinate system are computed with its spatial coordinates held fixed, but these coordinates are generally moving with respect to the other coordinates. In contrast, the spatial components to the partial derivative operator transform among just the other spatial components; there is no mixing with the time derivative operator. This property of the spatial derivative operator follows from the use of universal Newtonian time. It allows us to focus on space tensors in the following sections.

7.10 Covariant derivative of a scalar

We now return to a focus on space tensor analysis by considering the contraction of spatial components to the partial derivative operator with the basis of one-forms. This contraction renders the geometric expression of the gradient operator acting on a scalar

$$\text{grad}(\psi) = \nabla\psi = \tilde{e}^a \partial_a \psi = \tilde{e}^{\bar{a}} \partial_{\bar{a}} \psi. \quad (7.94)$$

From this expression we define the *covariant derivative operator*

$$\nabla = \tilde{e}^a \partial_a \quad (7.95)$$

so that we refer to equation (7.94) as either the gradient acting on a scalar or the covariant derivative acting on a scalar.

7.11 Covariant derivative of a vector

The covariant derivative operator can act on a vector, in which case we consider $\nabla \vec{F}$. To perform calculations requires us to unpack the manifestly covariant expression $\nabla \vec{F}$ by introducing a coordinate representation

$$\nabla \vec{F} = (\tilde{e}^b \partial_b) (F^a \vec{e}_a). \quad (7.96)$$

7.11.1 Derivative of a vector

The chain rule leads to the expression for the partial derivative operator acting on a vector field

$$\partial_b \vec{F} = \partial_b (\vec{e}_a F^a) \quad \text{coordinate representation of the vector } \vec{F} \quad (7.97a)$$

$$= (\partial_b F^a) \vec{e}_a + F^a \partial_b \vec{e}_a \quad \text{chain rule} \quad (7.97b)$$

$$= (\partial_b F^a) \vec{e}_a + F^a \Gamma_{ba}^c \vec{e}_c \quad \text{define Christoffel symbols} \quad (7.97c)$$

$$= (\partial_b F^a + F^c \Gamma_{bc}^a) \vec{e}_a \quad \text{reorganize} \quad (7.97d)$$

$$= (\nabla_b F^a) \vec{e}_a \quad \text{define covariant derivative acting on vector component.} \quad (7.97e)$$

In the third equality we introduced the *Christoffel symbols*

$$\partial_b \vec{e}_a = \Gamma_{ba}^c \vec{e}_c. \quad (7.98)$$

The Christoffel symbols carry information about the partial derivatives of the basis vectors. They vanish in Cartesian coordinates yet are generally nonzero. In the final equality we introduced components to the covariant derivative acting on the vector components

$$\nabla_b F^a = \partial_b F^a + \Gamma_{bc}^a F^c. \quad (7.99)$$

Contracting $\partial_b \vec{F}$ with the basis one-form \tilde{e}^b leads to the coordinate invariant expression for the covariant derivative of a vector field

$$\nabla \vec{F} = (\tilde{e}^b \partial_b) \vec{F} = (\tilde{e}^b \nabla_b F^a) \vec{e}_a. \quad (7.100)$$

7.11.2 An alternative derivation

A heuristic explanation of these ideas follows by applying the elementary calculus notions to a vector field \vec{F} as represented by arbitrary coordinates ξ^a , in which case

$$\partial_b \vec{F} = \lim_{\Delta \rightarrow 0} \frac{\vec{F}(\vec{P} + \Delta \vec{e}_b) - \vec{F}(\vec{P})}{\Delta}, \quad (7.101)$$

where $\vec{P} = \vec{e}_a \xi^a$ is the position vector for an arbitrary point and \vec{e}_b specifies the direction for computing the partial derivative. The basis vectors \vec{e}_a are spatially independent for Cartesian coordinates, so that the derivative of a vector is computed merely by taking the derivative of each Cartesian component

$$\partial_b \vec{F} = (\partial_b F^a) \vec{e}_a \quad \text{Cartesian coordinates.} \quad (7.102)$$

However, for general coordinates both the vector components and the basis vectors are spatially dependent, in which case

$$\vec{F}(\vec{P} + \Delta \vec{e}_b) - \vec{F}(\vec{P}) = [F^a + \Delta \partial_b F^b] [\vec{e}_a + \Delta \partial_b \vec{e}_a] - F^a \vec{e}_a \quad (7.103a)$$

$$= \Delta \partial_b (F^a \vec{e}_a) + \mathcal{O}(\Delta^2). \quad (7.103b)$$

This is the same result as found in the first step of the chain rule used in equation (7.97a). Following through that derivation then leads to the same coordinate expression for the covariant derivative acting on a vector field.

7.11.3 Christoffel symbols as the metric connection

Recall from elementary calculus that the derivative of a function is computed by comparing the function at two points in space, dividing by the distance between those points, and taking the limit as the points get infinitesimally close. This operation is well defined for scalar fields on arbitrary manifolds. However, it is problematic for vectors since the vectors live on distinct tangent spaces and so cannot be directly compared. For example, how do we compare two vectors at distinct points on a sphere? To do so we must provide a method to move one vector to the position of the other before comparing. As seen through the above discussion of covariant derivative of a vector, the Christoffel symbols provide the means to move vectors. Namely, they connect the two vectors by carrying information about how the basis vectors change in space. It is for this reason that some refer to the Christoffel symbols as the *metric connection*.

There is a distinct coordinate dependence of Christoffel symbols. For example, the Christoffel symbols all vanish in Euclidean space when using Cartesian coordinates whereas they are nonzero with other coordinates. As discussed in Section 6.1, a tensor that vanishes in one coordinate system remains zero for all coordinate systems. We thus conclude that the Christoffel symbols are *not* components to a tensor. Rather, they carry information regarding the partial derivatives of the coordinate basis vectors and as such they are fundamentally tied to a chosen coordinate system.

7.12 Covariant derivative of a one-form

The product of a one-form and a vector is a scalar. As seen in Section 7.10, the covariant derivative of a scalar field is given by the gradient operator

$$\nabla(\tilde{E} \cdot \vec{F}) = \tilde{e}^b \partial_b (E_a F^a). \quad (7.104)$$

Expanding the partial derivative yields

$$\partial_b(E_a F^a) = F^a \partial_b E_a + E_a \partial_b F^a \quad (7.105a)$$

$$= F^a \partial_b E_a + E_a (\nabla_b F^a - \Gamma_{bc}^a F^c) \quad (7.105b)$$

$$= F^a (\partial_b E_a - \Gamma_{ba}^c E_c) + E_a \nabla_b F^a \quad (7.105c)$$

$$\equiv F^a \nabla_b E_a + E_a \nabla_b F^a. \quad (7.105d)$$

The last equality defines the covariant derivative when acting on the components to a one form

$$\nabla_b E_a = \partial_b E_a - \Gamma_{ba}^c E_c, \quad (7.106)$$

which leads to the coordinate invariant expression for the covariant derivative of a one-form

$$\nabla \tilde{E} = (\tilde{e}^b \partial_b) \tilde{E} = (\tilde{e}^b \nabla_b E_a) \tilde{e}^a. \quad (7.107)$$

7.13 Covariant derivative of the metric tensor

When written in Cartesian coordinates, the covariant derivative of components to the metric tensor for Euclidean space vanishes,

$$\nabla g_{ab} = \nabla \delta_{ab} = 0, \quad (7.108)$$

because the Cartesian representation of the metric is the unit tensor δ_{ab} so that all Christoffel symbols vanish. Previous results establish the tensorial nature of the covariant derivative. Hence, $\nabla g_{ab} = 0$ is a valid result for *all* coordinates. This result is often called the *metricity* condition. It represents a self-consistency condition required for the manifolds considered in geophysical fluid mechanics.

7.14 Christoffel symbols in terms of the metric

We can develop an expression for the covariant derivative when acting on the components to a second order tensor. When applied to the metric tensor, its vanishing covariant derivative (equation (7.108)) then leads to the identity

$$0 = \nabla_c g_{ab} = \partial_c g_{ab} - \Gamma_{ca}^d g_{db} - \Gamma_{cb}^d g_{ad}. \quad (7.109)$$

We can solve this equation for the Christoffel symbols

$$\Gamma_{ab}^c = \frac{1}{2} g^{cd} (\partial_b g_{da} + \partial_a g_{db} - \partial_d g_{ab}). \quad (7.110)$$

This expression exhibits the symmetry property of the lower two indices on the Christoffel symbols

$$\Gamma_{ab}^c = \Gamma_{ba}^c. \quad (7.111)$$

7.15 Covariant divergence of a vector

The covariant divergence of the components to a vector results in a scalar

$$\nabla_a F^a = \partial_a F^a + \Gamma_{ab}^a F^b. \quad (7.112)$$

We now bring this expression into a form more convenient for practical calculations.

7.15.1 Contraction of the Christoffel symbols

Expression (7.110) for the Christoffel symbols yields the contraction

$$\Gamma_{ab}^a = \frac{1}{2} g^{ad} (\partial_b g_{da} + \partial_a g_{db} - \partial_d g_{ab}) = \frac{1}{2} g^{ad} \partial_b g_{ad} \quad (7.113)$$

where symmetry of the metric tensor and its inverse was used.

7.15.2 Exponential of the determinant

For any symmetric positive definite matrix such as the metric tensor we can write

$$\det(A) = e^{\ln \det(A)} \quad \text{simple identity} \quad (7.114a)$$

$$= e^{\ln(\Pi_i \Lambda_i)} \quad \text{determinant related to product of eigenvalues} \quad (7.114b)$$

$$= e^{\Sigma_i \ln \Lambda_i} \quad \text{simple identity} \quad (7.114c)$$

$$= e^{\text{Tr}(\ln A)} \quad \text{sum of eigenvalues related to trace of matrix.} \quad (7.114d)$$

Each of these identities is trivial to verify using a set of coordinates in which the matrix is diagonal. For any symmetric and positive definite matrix, such a set of coordinates always exists, in which case

$$\partial_c \ln \det(A) = \partial_c [\text{Tr}(\ln A)] = \text{Tr}(\partial_c \ln A) = \text{Tr}(A^{-1} \partial_c A). \quad (7.115)$$

With A now set equal to the metric tensor \mathcal{G} with components g_{ab} , this result yields

$$\partial_c \ln \det(\mathcal{G}) = g^{ab} \partial_c g_{ab} \quad (7.116)$$

which in turn yields for the contracted Christoffel symbol

$$\Gamma_{ac}^a = \partial_c \ln \sqrt{\det(\mathcal{G})}. \quad (7.117)$$

This result brings the covariant divergence of a vector to the form

$$\nabla_a F^a = \partial_a F^a + F^a \partial_a \ln \sqrt{\det(\mathcal{G})} = \frac{1}{\sqrt{\mathcal{G}}} \partial_a [\sqrt{\mathcal{G}} F^a]. \quad (7.118)$$

This is a very convenient result since it only requires partial derivatives in the chosen coordinate system, with all of the coordinate dependent properties summarized by $\sqrt{\det(\mathcal{G})}$.

7.16 Covariant Laplacian of a scalar

Making use of equation (7.118) with

$$F^a = g^{ab} \partial_b \psi \quad (7.119)$$

leads to the covariant Laplacian of a scalar field

$$\nabla_a (g^{ab} \partial_b \psi) = \frac{1}{\sqrt{\mathcal{G}}} \partial_a [\sqrt{\mathcal{G}} g^{ab} \partial_b \psi]. \quad (7.120)$$

This expression is fundamental to the evolution of scalar fields under the impacts from diffusion (Chapter 48).

7.17 Covariant curl of a vector

The Levi-Civita tensor

$$\varepsilon_{abc} = \sqrt{\det(\mathcal{G})} \epsilon_{abc} \quad (7.121)$$

from Section 7.7 is useful for generalizing the curl operation from Cartesian coordinates in Euclidean space to arbitrary coordinates on a curved manifold. Consequently, we define the curl according to the coordinate invariant expression

$$\text{curl}(\vec{F}) = \vec{e}_a \varepsilon^{abc} (\nabla_b F_c) = \vec{e}^a \varepsilon_{abc} (\nabla^b F^c). \quad (7.122)$$

This expression simplifies by making use of equation (7.106) for the covariant derivative $\nabla_b F_c = \partial_b F_c - \Gamma_{cb}^a F_a$. Conveniently, the contraction $\varepsilon^{abc} \Gamma_{cb}^a$ vanishes identically since $\varepsilon^{abc} = -\varepsilon^{acb}$ whereas $\Gamma_{cb}^a = \Gamma_{bc}^a$. Hence, we are left with a general expression for the covariant curl that involves just the partial derivatives

$$\text{curl}(\vec{F}) = \vec{e}_a \varepsilon^{abc} (\partial_b F_c) = \vec{e}_a \varepsilon^{abc} [\partial_b (g_{cd} F^d)]. \quad (7.123)$$

The second equality made use of the identity $F_c = g_{cd} F^d$.

7.18 Gauss's divergence theorem

The integral theorems from Cartesian vector analysis transform in a straightforward manner to arbitrary coordinates in arbitrary smooth and oriented spaces. An easy way to prove the theorems is to invoke the ideas of general coordinate invariance from Section 6.1, in which the integral theorems are written in a tensorially proper manner with partial derivatives changed to covariant derivatives. The divergence theorem offers a particularly simple example. For this purpose, we make use of the volume element (7.61)

$$dV = \sqrt{\det(\mathcal{G})} d\xi^1 d\xi^2 d\xi^3, \quad (7.124)$$

multiplied by the covariant divergence (7.118). Hence, the volume integral of the divergence is given by

$$\int_{\mathcal{R}} (\nabla_a F^a) dV = \int_{\mathcal{R}} \partial_a [\sqrt{\det(\mathcal{G})} F^a] d\xi^1 d\xi^2 d\xi^3 = \oint_{\partial\mathcal{R}} F^a \hat{n}_a d\mathcal{S}, \quad (7.125)$$

where \hat{n} is the outward normal one-form for the boundary, $\partial\mathcal{R}$, \hat{n}_a are its covariant components, and $d\mathcal{S}$ is the invariant area element on the boundary.

7.19 Stokes' theorem

The Cartesian form of Stokes' Theorem from Section 2.6 is generalized in a manner similar to the divergence theorem

$$\oint_{\partial\mathcal{S}} \vec{F} \cdot d\vec{x} = \int_{\mathcal{S}} \text{curl}(\vec{F}) \cdot \hat{n} d\mathcal{S}, \quad (7.126)$$

where $d\vec{x}$ is the vector line element along the path and $\partial\mathcal{S}$ is the closed path defining the boundary to a simply connected two-dimensional surface \mathcal{S} . For the circulation on the left hand side we have

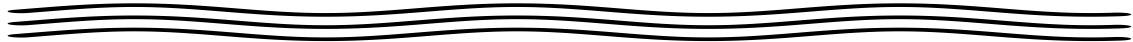
$$\vec{F} \cdot d\vec{x} = F^a \vec{e}_a \cdot \vec{e}_b dx^b = F_b dx^b = F_{\bar{b}} d\xi^{\bar{b}}. \quad (7.127)$$

For the curl on the right hand side we have

$$\text{curl}(\vec{F}) \cdot \hat{n} = \varepsilon^{abc} (\partial_b F_c) \vec{e}_a \cdot \hat{n} = \varepsilon^{abc} (\partial_b F_c) \hat{n}_a = \varepsilon^{\bar{a}\bar{b}\bar{c}} (\partial_{\bar{b}} F_{\bar{c}}) \hat{n}_{\bar{a}}, \quad (7.128)$$

thus leading to the expression of Stokes' theorem in arbitrary coordinates

$$\oint_{\partial\mathcal{S}} F_{\bar{b}} d\xi^{\bar{b}} = \int_{\mathcal{S}} \varepsilon^{\bar{a}\bar{b}\bar{c}} (\partial_{\bar{b}} F_{\bar{c}}) \hat{n}_{\bar{a}} d\mathcal{S}. \quad (7.129)$$



Orthogonal coordinates

READER'S GUIDE TO THIS CHAPTER

This chapter provides a compendium of mathematical results for Cartesian, spherical, and cylindrical coordinates, thus providing explicit examples of the general tensor analysis machinery from Chapter 7. These coordinates are all time independent and orthogonal. Hence, they are simpler than the non-orthogonal generalized vertical coordinates detailed in Chapter 43.

8.1	Cartesian coordinates	151
8.1.1	The basics	152
8.1.2	Summary	153
8.2	Spherical coordinates	153
8.2.1	Cartesian and spherical transformation	154
8.2.2	Basis vectors	155
8.2.3	Basis one-forms	155
8.2.4	Position and velocity	156
8.2.5	Metric tensor	156
8.2.6	Components of a vector field	157
8.2.7	Differential operators	157
8.2.8	Summary	158
8.3	Cylindrical-polar coordinates	158
8.3.1	Cartesian and cylindrical-polar transformation	160
8.3.2	Basis vectors	160
8.3.3	Basis one-forms	160
8.3.4	Position and velocity	161
8.3.5	Metric tensor	161
8.3.6	Volume element and Levi-Civita tensor	161
8.3.7	Vector cross product of basis vectors	161
8.3.8	Components of a vector field	162
8.3.9	Differential operators	162
8.3.10	Summary	163
8.4	General orthogonal coordinates	163

8.1 Cartesian coordinates

Whenever developing a general tensor relation it is useful to check its validity by considering Cartesian coordinates, in which case we can make use of familiar rules from vector calculus. We

here summarize some results from our discussion of Cartesian tensors in Chapters 1 and 2.

8.1.1 The basics

We start by expressing the trajectory of a point through space in the following equivalent forms

$$\mathcal{P}(\tau) = \vec{e}_1 x(\tau) + \vec{e}_2 y(\tau) + \vec{e}_3 z(\tau) \quad (8.1a)$$

$$= \hat{\mathbf{x}} x(\tau) + \hat{\mathbf{y}} y(\tau) + \hat{\mathbf{z}} z(\tau) \quad (8.1b)$$

$$= \vec{x}(\tau) \quad (8.1c)$$

$$= \mathbf{x}(\tau), \quad (8.1d)$$

with the basis vectors written

$$\vec{e}_1 = \hat{\mathbf{x}} \quad \text{and} \quad \vec{e}_2 = \hat{\mathbf{y}} \quad \text{and} \quad \vec{e}_3 = \hat{\mathbf{z}}. \quad (8.2)$$

The boldface notation is used for the position vector in the final equality of equation (8.1d), with the boldface commonly used throughout this book. Notably, the orthogonal unit vectors for Cartesian coordinates are normalized so that

$$\vec{e}_1 \cdot \vec{e}_1 = \vec{e}_2 \cdot \vec{e}_2 = \vec{e}_3 \cdot \vec{e}_3 = 1. \quad (8.3)$$

Furthermore, the basis vectors are identical to the basis one-forms

$$\vec{e}_1 = \tilde{e}^1 = \hat{\mathbf{x}} \quad \text{and} \quad \vec{e}_2 = \tilde{e}^2 = \hat{\mathbf{y}} \quad \text{and} \quad \vec{e}_3 = \tilde{e}^3 = \hat{\mathbf{z}}. \quad (8.4)$$

Since the Cartesian basis vectors are independent of both space and time, we compute the coordinate representation of the velocity vector through taking the time derivative as

$$\vec{v}(\tau) = \mathbf{v}(\tau) = \frac{d\mathcal{P}}{d\tau} = \frac{d\mathbf{x}}{d\tau}, \quad (8.5)$$

which takes on the expanded expressions

$$\vec{v}(\tau) = \vec{e}_1 \frac{dx(\tau)}{d\tau} + \vec{e}_2 \frac{dy(\tau)}{d\tau} + \vec{e}_3 \frac{dz(\tau)}{d\tau} \quad (8.6a)$$

$$= \hat{\mathbf{x}} v^1(\tau) + \hat{\mathbf{y}} v^2(\tau) + \hat{\mathbf{z}} v^3(\tau). \quad (8.6b)$$

$$= \hat{\mathbf{x}} u(\tau) + \hat{\mathbf{y}} v(\tau) + \hat{\mathbf{z}} w(\tau). \quad (8.6c)$$

8.1.2 Summary

In Cartesian coordinates, mathematical operators and integral theorems take their familiar form from vector calculus. We here list the key ones in forms that are encountered throughout this book.

$$\mathbf{x} = (x^1, x^2, x^3) = (x, y, z) \quad \text{Cartesian coordinates} \quad (8.7)$$

$$\mathbf{F} = \hat{\mathbf{x}}F^1 + \hat{\mathbf{y}}F^2 + \hat{\mathbf{z}}F^3 = \hat{\mathbf{x}}F_1 + \hat{\mathbf{y}}F_2 + \hat{\mathbf{z}}F_3 \quad \text{covariant} = \text{contravariant} \quad (8.8)$$

$$\frac{\partial}{\partial x^a} = \partial_a \quad \text{or} \quad (\partial_x, \partial_y, \partial_z) \quad \text{partial derivative operator} \quad (8.9)$$

$$\nabla = \hat{\mathbf{x}} \partial_x + \hat{\mathbf{y}} \partial_y + \hat{\mathbf{z}} \partial_z \quad \text{gradient operator} \quad (8.10)$$

$$\nabla_z = \hat{\mathbf{x}} \partial_x + \hat{\mathbf{y}} \partial_y \quad \text{horizontal gradient operator} \quad (8.11)$$

$$\nabla \cdot \mathbf{F} = \partial_x F_x + \partial_y F_y + \partial_z F_z \quad \text{divergence of a vector} \quad (8.12)$$

$$\nabla_z \cdot \mathbf{F} = \partial_x F_x + \partial_y F_y \quad \text{horizontal divergence of a vector} \quad (8.13)$$

$$(\nabla \wedge \mathbf{F})_a = \epsilon_{abc} \partial_b F_c \quad \text{components to curl of a vector} \quad (8.14)$$

$$\nabla \cdot \nabla \psi = \nabla^2 \psi = (\partial_{xx} + \partial_{yy} + \partial_{zz}) \psi \quad \text{Laplacian of a scalar} \quad (8.15)$$

$$\int_{\mathcal{R}} \nabla \cdot \mathbf{F} dV = \oint_{\partial \mathcal{R}} \mathbf{F} \cdot \hat{\mathbf{n}} d\mathcal{S} \quad \text{Gauss's divergence theorem} \quad (8.16)$$

$$\oint_{\partial \mathcal{S}} \mathbf{F} \cdot d\mathbf{x} = \int_{\mathcal{S}} (\nabla \wedge \mathbf{F}) \cdot \hat{\mathbf{n}} d\mathcal{S}. \quad \text{Stokes' theorem.} \quad (8.17)$$

8.2 Spherical coordinates

We now consider spherical coordinates defined by Figure 8.1 and related to Cartesian coordinates through the transformation

$$x = r \cos \phi \cos \lambda \quad (8.18a)$$

$$y = r \cos \phi \sin \lambda \quad (8.18b)$$

$$z = r \sin \phi. \quad (8.18c)$$

The radial coordinate

$$r = |\mathbf{x}| = \sqrt{\mathbf{x} \cdot \mathbf{x}} = \sqrt{x^2 + y^2 + z^2} \quad (8.19)$$

measures the distance from the center of the sphere to the position of the point. The spherical angle coordinates

$$0 \leq \lambda \leq 2\pi \quad \text{longitude} \quad (8.20)$$

$$-\pi/2 \leq \phi \leq \pi/2 \quad \text{latitude} \quad (8.21)$$

specify the longitude, measuring the radians of the position east of the prime meridian, and latitude, measuring the radians north or south from the equator. To streamline notation in the following, we introduce the unbarred and barred labels for the Cartesian and spherical coordinates, respectively

$$(x, y, z) = (\xi^1, \xi^2, \xi^3) \equiv \xi^a \quad \text{and} \quad (\lambda, \phi, r) = (\xi^{\bar{1}}, \xi^{\bar{2}}, \xi^{\bar{3}}) \equiv \xi^{\bar{a}}. \quad (8.22)$$

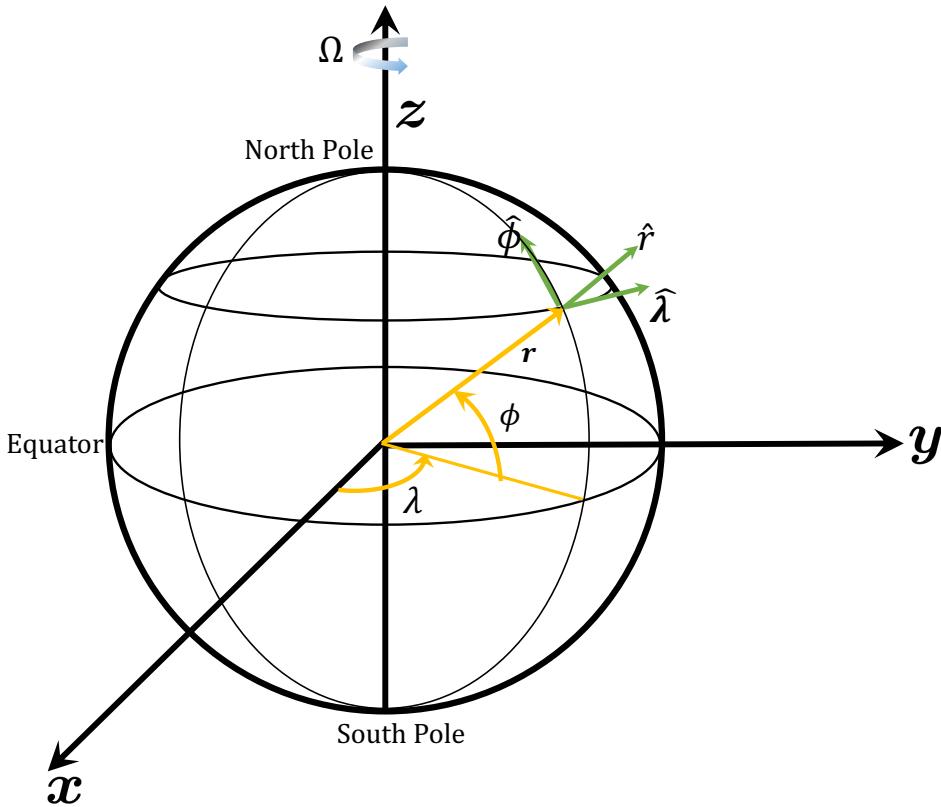


FIGURE 8.1: This schematic illustrates the geometry and notation for motion around a rotating sphere of radius R . For geophysical applications, the sphere rotates counter-clockwise when looking down from the north polar axis and it has an angular speed Ω . The planetary Cartesian triad of orthonormal basis vectors, $(\hat{x}, \hat{y}, \hat{z})$ points along the orthogonal axes and rotates with the sphere. The planetary spherical triad (also rotating with the sphere) of orthonormal basis vectors, $(\hat{\lambda}, \hat{\phi}, \hat{r})$, makes use of the longitudinal unit vector $\hat{\lambda}$, which points in the longitudinal direction (positive eastward), the latitudinal unit vector $\hat{\phi}$, which points in the latitudinal direction (positive northward) and the radial unit vector \hat{r} , which point in the radial direction (positive away from the center).

8.2.1 Cartesian and spherical transformation

Following the general discussion in Section 7.1.4, we consider the infinitesimal distance along one of the Cartesian coordinate axes, $d\xi^a$. The chain rule allows us to relate this distance to those along the axes of the spherical coordinate system

$$d\xi^a = \frac{\partial \xi^a}{\partial \xi^{\bar{a}}} d\xi^{\bar{a}} = \Lambda_{\bar{a}}^a d\xi^{\bar{a}}. \quad (8.23)$$

The partial derivatives $\partial \xi^a / \partial \xi^{\bar{a}}$ form components to the transformation matrix that transforms between coordinate representations. For the coordinate transformation (8.18a)-(8.18c), the transformation matrix is given by

$$\Lambda_{\bar{a}}^a = \begin{bmatrix} \partial \xi^1 / \partial \xi^{\bar{1}} & \partial \xi^1 / \partial \xi^{\bar{2}} & \partial \xi^1 / \partial \xi^{\bar{3}} \\ \partial \xi^2 / \partial \xi^{\bar{1}} & \partial \xi^2 / \partial \xi^{\bar{2}} & \partial \xi^2 / \partial \xi^{\bar{3}} \\ \partial \xi^3 / \partial \xi^{\bar{1}} & \partial \xi^3 / \partial \xi^{\bar{2}} & \partial \xi^3 / \partial \xi^{\bar{3}} \end{bmatrix} = \begin{bmatrix} -r \cos \phi \sin \lambda & -r \sin \phi \cos \lambda & \cos \phi \cos \lambda \\ r \cos \phi \cos \lambda & -r \sin \phi \sin \lambda & \cos \phi \sin \lambda \\ 0 & r \cos \phi & \sin \phi \end{bmatrix}. \quad (8.24)$$

The determinant of the transformation (Jacobian) is given by

$$\det(\Lambda_{\bar{a}}^a) = r^2 \cos \phi. \quad (8.25)$$

The Jacobian vanishes at the north and south poles ($\phi = \pm\pi/2$), where the transformation is singular. Methods familiar from linear algebra render the inverse transformation matrix

$$\Lambda_{\bar{a}}^a = \frac{1}{r^2 \cos \phi} \begin{bmatrix} -r \sin \lambda & r \cos \lambda & 0 \\ -r \cos \phi \sin \phi \cos \lambda & -r \cos \phi \sin \phi \sin \lambda & r \cos^2 \phi \\ r^2 \cos^2 \phi \cos \lambda & r^2 \cos^2 \phi \sin \lambda & r^2 \cos \phi \sin \phi \end{bmatrix}. \quad (8.26)$$

8.2.2 Basis vectors

The spherical coordinate basis vectors, $\vec{e}_{\bar{a}}$, are related to the Cartesian coordinate basis vectors, \vec{e}_a , through the transformation

$$\vec{e}_{\bar{a}} = \Lambda_{\bar{a}}^a \vec{e}_a. \quad (8.27)$$

The transformation matrix (8.24) leads to

$$\vec{e}_\lambda = r \cos \phi (-\hat{x} \sin \lambda + \hat{y} \cos \lambda) \quad (8.28a)$$

$$\vec{e}_\phi = r (-\hat{x} \sin \phi \cos \lambda - \hat{y} \sin \phi \sin \lambda + \hat{z} \cos \phi) \quad (8.28b)$$

$$\vec{e}_r = \hat{x} \cos \phi \cos \lambda + \hat{y} \cos \phi \sin \lambda + \hat{z} \sin \phi. \quad (8.28c)$$

We can introduce the orthonormal unit vectors through

$$\vec{e}_\lambda = r \cos \phi \hat{\lambda} \quad \text{and} \quad \vec{e}_\phi = r \hat{\phi} \quad \text{and} \quad \vec{e}_r = \hat{r}, \quad (8.29)$$

so that

$$\hat{\lambda} = -\hat{x} \sin \lambda + \hat{y} \cos \lambda \quad (8.30a)$$

$$\hat{\phi} = -\hat{x} \cos \lambda \sin \phi - \hat{y} \sin \lambda \sin \phi + \hat{z} \cos \phi \quad (8.30b)$$

$$\hat{r} = \hat{x} \cos \lambda \cos \phi + \hat{y} \sin \lambda \cos \phi + \hat{z} \sin \phi \quad (8.30c)$$

along with the inverse relations

$$\hat{x} = -\hat{\lambda} \sin \lambda - \hat{\phi} \cos \lambda \sin \phi + \hat{r} \cos \lambda \cos \phi \quad (8.31a)$$

$$\hat{y} = \hat{\lambda} \cos \lambda - \hat{\phi} \sin \lambda \sin \phi + \hat{r} \sin \lambda \cos \phi \quad (8.31b)$$

$$\hat{z} = \hat{\phi} \cos \phi + \hat{r} \sin \phi. \quad (8.31c)$$

8.2.3 Basis one-forms

Since spherical coordinates are orthogonal, we can readily derive the one-form basis through inverting the vector basis

$$\tilde{e}^\lambda = (r \cos \phi)^{-1} \hat{\lambda} \quad \text{and} \quad \tilde{e}^\phi = r^{-1} \hat{\phi} \quad \text{and} \quad \tilde{e}^r = \hat{r}, \quad (8.32)$$

which satisfy the bi-orthogonality relation with the basis vectors (Section 7.2.2)

$$\tilde{e}^{\bar{b}} \cdot \vec{e}_{\bar{a}} = \delta^{\bar{b}}_{\bar{a}}. \quad (8.33)$$

8.2.4 Position and velocity

In spherical coordinates, the position of a point is fully specified by the radial position

$$\mathcal{P}(\tau) = r \vec{e}_r = r \hat{\mathbf{r}}. \quad (8.34)$$

The velocity requires all three spherical coordinates since the radial basis vector is a function of the angular positions, which are in turn functions of time. Use of the chain rule renders

$$\vec{v}(\tau) = \mathbf{v} \quad (8.35a)$$

$$= \frac{d\mathcal{P}}{d\tau} \quad (8.35b)$$

$$= \vec{e}_r \frac{dr}{d\tau} + r \frac{d\vec{e}_r}{d\tau} \quad (8.35c)$$

$$= \vec{e}_r \frac{dr}{d\tau} + r \frac{\partial \vec{e}_r}{\partial \lambda} \frac{d\lambda}{d\tau} + r \frac{\partial \vec{e}_r}{\partial \phi} \frac{d\phi}{d\tau} \quad (8.35d)$$

$$\equiv \vec{e}_r \frac{dr}{d\tau} + \vec{e}_\lambda \frac{d\lambda}{d\tau} + \vec{e}_\phi \frac{d\phi}{d\tau} \quad (8.35e)$$

$$= \vec{e}_r v^r + \vec{e}_\lambda v^\lambda + \vec{e}_\phi v^\phi. \quad (8.35f)$$

To reach this result we made use of the identities satisfied by the spherical basis vectors

$$\vec{e}_\lambda = r \frac{\partial \vec{e}_r}{\partial \lambda} \quad \text{and} \quad \vec{e}_\phi = r \frac{\partial \vec{e}_r}{\partial \phi}. \quad (8.36)$$

8.2.5 Metric tensor

The metric tensor for spherical coordinates takes on the diagonal form

$$g_{\bar{a}\bar{b}} = \vec{e}_{\bar{a}} \cdot \vec{e}_{\bar{b}} = \begin{bmatrix} (r \cos \phi)^2 & 0 & 0 \\ 0 & r^2 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad (8.37)$$

as does the spherical representation of the inverse metric tensor

$$g^{\bar{a}\bar{b}} = \vec{e}^{\bar{a}} \cdot \vec{e}^{\bar{b}} = \begin{bmatrix} (r \cos \phi)^{-2} & 0 & 0 \\ 0 & r^{-2} & 0 \\ 0 & 0 & 1 \end{bmatrix}. \quad (8.38)$$

A diagonal metric tensor is a property of orthogonal coordinates.

Volume element and Levi-Civita tensor

The square root of the determinant of the metric tensor written in spherical coordinates (from equation (8.37)) is given by

$$\sqrt{\det(\mathcal{G})} = r^2 \cos \phi \quad (8.39)$$

so that the volume element is

$$dV = r^2 \cos \phi dr d\lambda d\phi. \quad (8.40)$$

The covariant Levi-Civita tensor has the spherical representation

$$\varepsilon_{\bar{a}\bar{b}\bar{c}} = (r^2 \cos \phi) \epsilon_{\bar{a}\bar{b}\bar{c}}, \quad (8.41)$$

where $\epsilon_{\bar{a}\bar{b}\bar{c}}$ are components to the permutation symbol (i.e., the Cartesian components to the Levi-Civita tensor) from Section 1.4.1.

Vector cross product of basis vectors

As a check on the formalism for cross products, let us verify the relation (7.79) for the vector cross product of two basis vectors using spherical coordinates

$$\vec{e}_{\bar{a}} \wedge \vec{e}_{\bar{b}} = \epsilon_{\bar{a}\bar{b}\bar{c}} \tilde{e}^{\bar{c}} \implies \vec{e}_{\bar{a}} \wedge \vec{e}_{\bar{b}} = (r^2 \cos \phi) \epsilon_{\bar{a}\bar{b}\bar{c}} \tilde{e}^{\bar{c}}. \quad (8.42)$$

Making use of the spherical coordinate basis vectors and one-forms renders

$$\vec{e}_r \wedge \vec{e}_{\lambda} = (r \cos \phi) (\hat{r} \wedge \hat{\lambda}) = (r \cos \phi) \hat{\phi} = (r^2 \cos \phi) \tilde{e}^{\phi} = \epsilon_{r\lambda\phi} \tilde{e}^{\phi} \quad (8.43a)$$

$$\vec{e}_{\lambda} \wedge \vec{e}_{\phi} = (r^2 \cos \phi) (\hat{\lambda} \wedge \hat{\phi}) = (r^2 \cos \phi) \hat{r} = (r^2 \cos \phi) \tilde{e}^r = \epsilon_{\lambda\phi r} \tilde{e}^r \quad (8.43b)$$

$$\vec{e}_{\phi} \wedge \vec{e}_r = r (\hat{\phi} \wedge \hat{r}) = r \hat{\lambda} = (r^2 \cos \phi) \tilde{e}^{\lambda} = \epsilon_{\phi r \lambda} \tilde{e}^{\lambda}. \quad (8.43c)$$

To reach these results we made use of the cross products for the spherical coordinate unit vectors

$$\hat{r} \wedge \hat{\lambda} = \hat{\phi} \quad \text{and} \quad \hat{\lambda} \wedge \hat{\phi} = \hat{r} \quad \text{and} \quad \hat{\phi} \wedge \hat{r} = \hat{\lambda}. \quad (8.44)$$

8.2.6 Components of a vector field

A vector field \vec{F} has Cartesian components, F^a , related to spherical components, $F^{\bar{a}}$, via the transformation matrix, $F^{\bar{a}} = \Lambda_a^{\bar{a}} F^a$. This transformation leads to

$$F^{\bar{1}} = (r \cos \phi)^{-1} [-F^x \sin \lambda + F^y \cos \lambda] \quad (8.45a)$$

$$F^{\bar{2}} = r^{-1} [-F^x \sin \phi \cos \lambda - F^y \sin \phi \sin \lambda + F^z \cos \phi] \quad (8.45b)$$

$$F^{\bar{3}} = F^x \cos \phi \cos \lambda + F^y \cos \phi \sin \lambda + F^z \sin \phi. \quad (8.45c)$$

Making use of the spherical unit vector (8.30a)-(8.30c) leads to the more tidy relations

$$(r \cos \phi) F^{\bar{1}} = \hat{\lambda} \cdot \mathbf{F} \quad (8.46a)$$

$$r F^{\bar{2}} = \hat{\phi} \cdot \mathbf{F} \quad (8.46b)$$

$$F^{\bar{3}} = \hat{r} \cdot \mathbf{F}, \quad (8.46c)$$

where \mathbf{F} is the Cartesian representation.

8.2.7 Differential operators

In spherical coordinates the gradient operator $\nabla = \tilde{e}^a \partial_a$ takes on the form

$$\nabla = \hat{\lambda} (r \cos \phi)^{-1} \partial_{\lambda} + \hat{\phi} r^{-1} \partial_{\phi} + \hat{r} \partial_r \quad (8.47)$$

and the covariant divergence of a vector field is given by

$$\nabla_{\bar{a}} F^{\bar{a}} = (r^2 \cos \phi)^{-1} \partial_{\bar{a}} [r^2 \cos \phi F^{\bar{a}}] \quad (8.48a)$$

$$= (r^2 \cos \phi)^{-1} \left(\partial_{\lambda} [r^2 \cos \phi F^{\bar{1}}] + \partial_{\phi} [r^2 \cos \phi F^{\bar{2}}] + \partial_r [r^2 \cos \phi F^{\bar{3}}] \right) \quad (8.48b)$$

$$= \frac{1}{r \cos \phi} \frac{\partial (\hat{\lambda} \cdot \mathbf{F})}{\partial \lambda} + \frac{1}{r \cos \phi} \frac{\partial (\hat{\phi} \cdot \mathbf{F} \cos \phi)}{\partial \phi} + \frac{1}{r^2} \frac{\partial (\hat{r} \cdot \mathbf{F} r^2)}{\partial r}. \quad (8.48c)$$

The covariant Laplacian of a scalar, $\nabla^2\psi = \nabla \cdot \nabla\psi$, is given by

$$\nabla^2\psi = \nabla \cdot [\hat{\lambda}(r \cos \phi)^{-1} \partial_\lambda \psi + \hat{\phi} r^{-1} \partial_\phi \psi + \hat{r} \partial_r \psi] \quad (8.49a)$$

$$= \frac{1}{(r \cos \phi)^2} \frac{\partial^2 \psi}{\partial \lambda^2} + \frac{1}{r^2 \cos \phi} \frac{\partial}{\partial \phi} \left[\cos \phi \frac{\partial \psi}{\partial \phi} \right] + \frac{1}{r^2} \frac{\partial}{\partial r} \left[r^2 \frac{\partial \psi}{\partial r} \right]. \quad (8.49b)$$

The covariant curl (Section 7.17) takes the form

$$(\text{curl } \vec{F})^{\bar{1}} = (r^2 \cos \phi)^{-1} [\partial_\phi F^{\bar{3}} - \partial_r (r^2 F^{\bar{2}})] \quad (8.50a)$$

$$(\text{curl } \vec{F})^{\bar{2}} = (r^2 \cos \phi)^{-1} [\partial_r (r^2 \cos^2 \phi F^{\bar{1}}) - \partial_\lambda F^{\bar{3}}] \quad (8.50b)$$

$$(\text{curl } \vec{F})^{\bar{3}} = (r^2 \cos \phi)^{-1} [\partial_\lambda (r^2 F^{\bar{2}}) - \partial_\phi (r^2 \cos^2 \phi F^{\bar{1}})], \quad (8.50c)$$

which can be written in the more conventional form (e.g., equation (2.33) of [Vallis \(2017\)](#))

$$r \cos \phi (\text{curl } \vec{F})^{\bar{1}} = \frac{1}{r} \left[\frac{\partial(\hat{r} \cdot \mathbf{F})}{\partial \phi} - \frac{\partial(r \hat{\phi} \cdot \mathbf{F})}{\partial r} \right] \quad (8.51a)$$

$$r (\text{curl } \vec{F})^{\bar{2}} = \frac{1}{r} \left[\frac{\partial(r \hat{\lambda} \cdot \mathbf{F})}{\partial r} - \frac{1}{\cos \phi} \frac{\partial(\hat{r} \cdot \mathbf{F})}{\partial \lambda} \right] \quad (8.51b)$$

$$(\text{curl } \vec{F})^{\bar{3}} = \frac{1}{r \cos \phi} \left[\frac{\partial(\hat{\phi} \cdot \mathbf{F})}{\partial \lambda} - \frac{\partial(\cos \phi \hat{\lambda} \cdot \mathbf{F})}{\partial \phi} \right]. \quad (8.51c)$$

8.2.8 Summary

We here summarize the spherical coordinate version of some common mathematical operators.

$$(\lambda, \phi, r) = (x^{\bar{1}}, x^{\bar{2}}, x^{\bar{3}}) \quad \text{spherical coordinates} \quad (8.52)$$

$$(r \cos \phi) F^{\bar{1}} = \hat{\lambda} \cdot \mathbf{F} \quad r F^{\bar{2}} = \hat{\phi} \cdot \mathbf{F} \quad F^{\bar{3}} = \hat{r} \cdot \mathbf{F} \quad \text{vector components} \quad (8.53)$$

$$\nabla = \hat{\lambda} (r \cos \phi)^{-1} \partial_\lambda + \hat{\phi} r^{-1} \partial_\phi + \hat{r} \partial_r \quad \text{gradient operator} \quad (8.54)$$

$$\nabla_{\bar{a}} F^{\bar{a}} = \frac{1}{r \cos \phi} \frac{\partial(\hat{\lambda} \cdot \mathbf{F})}{\partial \lambda} + \frac{1}{r \cos \phi} \frac{\partial(\hat{\phi} \cdot \mathbf{F} \cos \phi)}{\partial \phi} + \frac{1}{r^2} \frac{\partial(\hat{r} \cdot \mathbf{F} r^2)}{\partial r} \quad \text{divergence of vector} \quad (8.55)$$

$$\nabla^2 \psi = \frac{1}{(r \cos \phi)^2} \frac{\partial^2 \psi}{\partial \lambda^2} + \frac{1}{r^2 \cos \phi} \frac{\partial}{\partial \phi} \left[\cos \phi \frac{\partial \psi}{\partial \phi} \right] + \frac{1}{r^2} \frac{\partial}{\partial r} \left[r^2 \frac{\partial \psi}{\partial r} \right] \quad \text{Laplacian of scalar} \quad (8.56)$$

$$(\nabla \wedge \mathbf{F})_{\bar{a}} = \varepsilon_{\bar{a}\bar{b}\bar{c}} \partial_{\bar{b}} F^{\bar{c}} \quad \text{see equations (8.51a) – (8.51c)} \quad \text{curl of vector.} \quad (8.57)$$

8.3 Cylindrical-polar coordinates

Many physical systems exhibit circular symmetry in two-dimensions or cylindrical symmetry in three-dimensions. The primary example encountered in this book is the laboratory motion of liquid in a rotating circular tank. In the following, we emulate the discussion presented for the spherical coordinates in Section 8.2, here focusing on cylindrical-polar coordinates as shown in Figure 8.2. Our task here is somewhat simpler than for the spherical coordinates since the vertical/axial position, z , remains unchanged from its Cartesian value. In a slight corruption of notation, we use the symbol r for the radial distance from the vertical axis in cylindrical-polar coordinates (Figure 8.2), which is distinct from the radial distance, r , used to measure the distance from the origin in spherical coordinates (Figure 8.1).

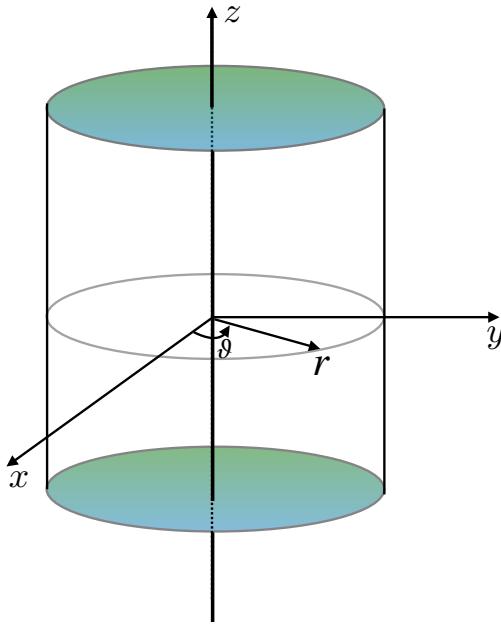


FIGURE 8.2: This schematic illustrates the geometry and notation for cylindrical-polar coordinates. The Cartesian triad of orthonormal basis vectors, $(\hat{x}, \hat{y}, \hat{z})$ points along the orthogonal axes. The cylindrical-polar triad of orthonormal basis vectors, $(\hat{r}, \hat{\theta}, \hat{z})$, makes use of the radial unit vector \hat{r} , which points outward from the vertical axis, the angular unit vector $\hat{\theta}$, which points in the counter-clockwise direction around the circle, and the vertical unit vector \hat{z} . Note that the radial unit vector used for cylindrical-polar coordinates is distinct from that radial vector used in spherical coordinates shown in Figure 8.1.

The coordinate transformation between Cartesian coordinates and cylindrical-polar coordinates is given by

$$x = r \cos \vartheta \equiv \xi^1 \cos \xi^2 \quad (8.58a)$$

$$y = r \sin \vartheta \equiv \xi^1 \sin \xi^2 \quad (8.58b)$$

$$z = \xi^3. \quad (8.58c)$$

The radial coordinate for cylindrical-polar coordinates

$$r = \sqrt{x^2 + y^2} \quad (8.59)$$

measures the distance from the vertical z -axis, and the angular coordinate $0 \leq \vartheta \leq 2\pi$ measures the angle counter-clockwise from the positive x -axis. We introduce the unbarred and barred labels for the Cartesian and cylindrical polar coordinates

$$(x, y, z) = (\xi^1, \xi^2, \xi^3) \equiv \xi^a \quad \text{and} \quad (r, \vartheta, z) = (\xi^1, \xi^2, \xi^3) \equiv \xi^{\bar{a}}. \quad (8.60)$$

Although the vertical position z remains the same in both coordinates, and it is orthogonal to the other coordinates, we find it useful to introduce a distinct symbols, ξ^3 and $\xi^{\bar{3}}$, to specify what other coordinates are held fixed when performing derivative operations.

8.3.1 Cartesian and cylindrical-polar transformation

The coordinate transformation (8.58a)-(8.58c) leads to the transformation matrix

$$\Lambda_{\bar{a}}^a = \begin{bmatrix} \partial\xi^1/\partial\xi^{\bar{1}} & \partial\xi^1/\partial\xi^{\bar{2}} & \partial\xi^1/\partial\xi^{\bar{3}} \\ \partial\xi^2/\partial\xi^{\bar{1}} & \partial\xi^2/\partial\xi^{\bar{2}} & \partial\xi^2/\partial\xi^{\bar{3}} \\ \partial\xi^3/\partial\xi^{\bar{1}} & \partial\xi^3/\partial\xi^{\bar{2}} & \partial\xi^3/\partial\xi^{\bar{3}} \end{bmatrix} = \begin{bmatrix} \cos\vartheta & -r\sin\vartheta & 0 \\ \sin\vartheta & r\cos\vartheta & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad (8.61)$$

and the inverse transformation is given by

$$\Lambda_{\bar{a}}^{\bar{a}} = \frac{1}{r} \begin{bmatrix} r\cos\vartheta & r\sin\vartheta & 0 \\ -\sin\vartheta & \cos\vartheta & 0 \\ 0 & 0 & r \end{bmatrix}. \quad (8.62)$$

The determinant of the transformation (Jacobian) is given by

$$\det(\Lambda_{\bar{a}}^a) = r, \quad (8.63)$$

which vanishes along the vertical axis, which is where the transformation is singular.

8.3.2 Basis vectors

The cylindrical-polar coordinate basis vectors, $\vec{e}_{\bar{a}}$, are related to the Cartesian coordinate basis vectors, \vec{e}_a , through the transformation $\vec{e}_{\bar{a}} = \Lambda_{\bar{a}}^a \vec{e}_a$. The transformation matrix (8.61) leads to

$$\vec{e}_r = \hat{x}\cos\vartheta + \hat{y}\sin\vartheta \quad (8.64a)$$

$$\vec{e}_{\vartheta} = r(-\hat{x}\sin\vartheta + \hat{y}\cos\vartheta) \quad (8.64b)$$

$$\vec{e}_{\bar{z}} = \hat{z}. \quad (8.64c)$$

We sometimes make use of the following orthonormal unit vectors ($\hat{r}, \hat{\vartheta}, \hat{z}$)

$$\vec{e}_r = \hat{r} \quad \text{and} \quad \vec{e}_{\vartheta} = r\hat{\vartheta} \quad \text{and} \quad \vec{e}_{\bar{z}} = \hat{z} \quad (8.65)$$

along with the inverse relations

$$\hat{x} = \hat{r}\cos\vartheta - \hat{\vartheta}\sin\vartheta \quad (8.66a)$$

$$\hat{y} = \hat{r}\sin\vartheta + \hat{\vartheta}\cos\vartheta \quad (8.66b)$$

$$\hat{z} = \hat{z}. \quad (8.66c)$$

8.3.3 Basis one-forms

Since cylindrical-polar coordinates are orthogonal, we can readily derive the one-form basis through inverting the vector basis

$$\tilde{e}^r = \hat{r} \quad \text{and} \quad \tilde{e}^{\vartheta} = r^{-1}\hat{\vartheta} \quad \text{and} \quad \tilde{e}^{\bar{z}} = \hat{z}, \quad (8.67)$$

which satisfy the orthogonality relation (Section 7.2.2)

$$\tilde{e}^{\bar{b}} \cdot \vec{e}_{\bar{a}} = \delta_{\bar{a}}^{\bar{b}}. \quad (8.68)$$

8.3.4 Position and velocity

In cylindrical-polar coordinates, the position of a point is specified by the radial position plus the vertical position

$$\mathcal{P}(\tau) = r \vec{e}_r + z \vec{e}_{\bar{z}}. \quad (8.69)$$

The velocity requires all three coordinates since the radial basis vector is a function of the angular positions, which are in turn functions of time. Use of the chain rule renders

$$\vec{v}(\tau) = \frac{d\mathcal{P}}{d\tau} \quad (8.70a)$$

$$= \vec{e}_r \frac{dr}{d\tau} + r \frac{d\vec{e}_r}{d\tau} + \vec{e}_{\bar{z}} \frac{dz}{d\tau} \quad (8.70b)$$

$$= \vec{e}_r \frac{dr}{d\tau} + r \frac{\partial \vec{e}_r}{\partial \vartheta} \frac{d\vartheta}{d\tau} + \vec{e}_{\bar{z}} \frac{dz}{d\tau} \quad (8.70c)$$

$$= \vec{e}_r \frac{dr}{d\tau} + \vec{e}_{\vartheta} \frac{d\vartheta}{d\tau} + \vec{e}_{\bar{z}} \frac{dz}{d\tau} \quad (8.70d)$$

$$= \vec{e}_r v^r + \vec{e}_{\vartheta} v^{\vartheta} + \vec{e}_{\bar{z}} v^{\bar{z}}. \quad (8.70e)$$

To reach this result we made use of the identity

$$\vec{e}_{\vartheta} = r \frac{\partial \vec{e}_r}{\partial \vartheta} = r \hat{\vartheta}. \quad (8.71)$$

8.3.5 Metric tensor

Cylindrical-polar coordinates are orthogonal with the metric tensor and its inverse represented by the diagonal matrices

$$g_{\bar{a}\bar{b}} = \vec{e}_{\bar{a}} \cdot \vec{e}_{\bar{b}} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & r^2 & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad \text{and} \quad g^{\bar{a}\bar{b}} = \vec{e}^{\bar{a}} \cdot \vec{e}^{\bar{b}} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & r^{-2} & 0 \\ 0 & 0 & 1 \end{bmatrix}. \quad (8.72)$$

8.3.6 Volume element and Levi-Civita tensor

The square root of the determinant of the metric tensor written in cylindrical-polar coordinates (from equation (8.72)) is given by

$$\sqrt{\det(\bar{G})} = r \quad (8.73)$$

so that the volume element is

$$dV = r dr d\vartheta dz. \quad (8.74)$$

The covariant Levi-Civita tensor has the cylindrical-polar representation

$$\varepsilon_{\bar{a}\bar{b}\bar{c}} = r \epsilon_{\bar{a}\bar{b}\bar{c}}. \quad (8.75)$$

8.3.7 Vector cross product of basis vectors

As a check on the formalism for vector cross products, let us verify the relation (7.79) for the cross product of two basis vectors using cylindrical-polar coordinates

$$\vec{e}_{\bar{a}} \wedge \vec{e}_{\bar{b}} = \varepsilon_{\bar{a}\bar{b}\bar{c}} \vec{e}^{\bar{c}} \implies \vec{e}_{\bar{a}} \wedge \vec{e}_{\bar{b}} = r \epsilon_{\bar{a}\bar{b}\bar{c}} \vec{e}^{\bar{c}}. \quad (8.76)$$

Making use of the cylindrical-polar coordinate basis vectors and one-forms renders

$$\vec{e}_r \wedge \vec{e}_\vartheta = r (\hat{\mathbf{r}} \wedge \hat{\boldsymbol{\vartheta}}) = r \tilde{e}^z = \varepsilon_{r\vartheta z} \tilde{e}^z \quad (8.77a)$$

$$\vec{e}_\vartheta \wedge \vec{e}_{\bar{z}} = r (\hat{\boldsymbol{\vartheta}} \wedge \hat{\mathbf{z}}) = r \hat{\mathbf{r}} = r \tilde{e}^r = \varepsilon_{\vartheta \bar{z} r} \tilde{e}^r \quad (8.77b)$$

$$\vec{e}_{\bar{z}} \wedge \vec{e}_r = \hat{\mathbf{z}} \wedge \hat{\mathbf{r}} = \hat{\boldsymbol{\vartheta}} = r \tilde{e}^\vartheta = \varepsilon_{\bar{z} r \vartheta} \tilde{e}^\vartheta. \quad (8.77c)$$

To reach these results we made use of the cross products for the unit vectors

$$\hat{\mathbf{r}} \wedge \hat{\boldsymbol{\vartheta}} = \hat{\mathbf{z}} \quad \text{and} \quad \hat{\boldsymbol{\vartheta}} \wedge \hat{\mathbf{z}} = \hat{\mathbf{r}} \quad \text{and} \quad \hat{\mathbf{z}} \wedge \hat{\mathbf{r}} = \hat{\boldsymbol{\vartheta}}. \quad (8.78)$$

8.3.8 Components of a vector field

A vector field \vec{F} has Cartesian components, F^a , related to cylindrical-polar components, $F^{\bar{a}}$, via the transformation matrix, $F^{\bar{a}} = \Lambda_a^{\bar{a}} F^a$. This transformation leads to

$$F^{\bar{1}} = F^x \cos \vartheta + F^y \sin \vartheta \quad (8.79a)$$

$$F^{\bar{2}} = r^{-1} [-F^x \sin \vartheta + F^y \cos \vartheta] \quad (8.79b)$$

$$F^{\bar{3}} = F^z. \quad (8.79c)$$

Introducing the cylindrical-polar unit vectors (8.65) leads to the more tidy expressions

$$F^{\bar{1}} = \hat{\mathbf{r}} \cdot \mathbf{F} \quad (8.80a)$$

$$r F^{\bar{2}} = \hat{\boldsymbol{\vartheta}} \cdot \mathbf{F} \quad (8.80b)$$

$$F^{\bar{3}} = \hat{\mathbf{z}} \cdot \mathbf{F}, \quad (8.80c)$$

where \mathbf{F} is the Cartesian representation.

8.3.9 Differential operators

In cylindrical-polar coordinates, the gradient operator $\nabla = \tilde{e}^a \partial_a$ takes on the form

$$\nabla = \hat{\mathbf{r}} \frac{\partial}{\partial r} + \frac{\hat{\boldsymbol{\vartheta}}}{r} \frac{\partial}{\partial \vartheta} + \hat{\mathbf{z}} \frac{\partial}{\partial z} \quad (8.81)$$

and the covariant divergence of a vector field is given by

$$\nabla_{\bar{a}} F^{\bar{a}} = r^{-1} \partial_{\bar{a}} (r F^{\bar{a}}) \quad (8.82a)$$

$$= r^{-1} \left(\partial_r [r F^{\bar{1}}] + \partial_\vartheta [r F^{\bar{2}}] + \partial_z [r F^{\bar{3}}] \right) \quad (8.82b)$$

$$= \frac{1}{r} \frac{\partial (r \hat{\mathbf{r}} \cdot \mathbf{F})}{\partial r} + \frac{1}{r} \frac{\partial (\hat{\boldsymbol{\vartheta}} \cdot \mathbf{F})}{\partial \vartheta} + \frac{\partial (\hat{\mathbf{z}} \cdot \mathbf{F})}{\partial z}. \quad (8.82c)$$

The covariant Laplacian of a scalar, $\nabla^2 \psi = \nabla \cdot \nabla \psi$, is given by

$$\nabla^2 \psi = \nabla \cdot \left[\hat{\mathbf{r}} \frac{\partial \psi}{\partial r} + \frac{\hat{\boldsymbol{\vartheta}}}{r} \frac{\partial \psi}{\partial \vartheta} + \hat{\mathbf{z}} \frac{\partial \psi}{\partial z} \right] \quad (8.83a)$$

$$= \frac{1}{r} \frac{\partial}{\partial r} \left[r \frac{\partial \psi}{\partial r} \right] + \frac{1}{r^2} \frac{\partial^2 \psi}{\partial \vartheta^2} + \frac{\partial^2 \psi}{\partial z^2}. \quad (8.83b)$$

The covariant curl (Section 7.17) takes the form

$$(\text{curl} \vec{F})^{\bar{1}} = r^{-1} [\partial_{\vartheta} F^{\bar{3}} - \partial_{\bar{z}} (r^2 F^{\bar{2}})] \quad (8.84a)$$

$$(\text{curl} \vec{F})^{\bar{2}} = r^{-1} [\partial_{\bar{z}} F^{\bar{1}} - \partial_r F^{\bar{3}}] \quad (8.84b)$$

$$(\text{curl} \vec{F})^{\bar{3}} = r^{-1} [\partial_r (r^2 F^{\bar{2}}) - \partial_{\vartheta} F^{\bar{1}}], \quad (8.84c)$$

which can be written more conventionally as

$$(\text{curl} \vec{F})^{\bar{1}} = \frac{1}{r} \frac{\partial (\hat{z} \cdot \mathbf{F})}{\partial \vartheta} - \frac{\partial (\hat{\vartheta} \cdot \mathbf{F})}{\partial z} \quad (8.85a)$$

$$r (\text{curl} \vec{F})^{\bar{2}} = \frac{\partial (\hat{r} \cdot \mathbf{F})}{\partial z} - \frac{\partial (\hat{z} \cdot \mathbf{F})}{\partial r} \quad (8.85b)$$

$$(\text{curl} \vec{F})^{\bar{3}} = \frac{1}{r} \frac{\partial (r \hat{\vartheta} \cdot \mathbf{F})}{\partial r} - \frac{1}{r} \frac{\partial (\hat{r} \cdot \mathbf{F})}{\partial \vartheta}. \quad (8.85c)$$

8.3.10 Summary

We here summarize the cylindrical coordinate version of some common mathematical operators.

$$(r, \vartheta, z) = (x^{\bar{1}}, x^{\bar{2}}, x^{\bar{3}}) \quad \text{cylindrical coordinates} \quad (8.86)$$

$$F^{\bar{1}} = \hat{r} \cdot \mathbf{F} \quad r F^{\bar{2}} = \hat{\vartheta} \cdot \mathbf{F} \quad F^{\bar{3}} = \hat{z} \cdot \mathbf{F} \quad \text{vector components} \quad (8.87)$$

$$\nabla = \hat{r} \frac{\partial}{\partial r} + \frac{\hat{\vartheta}}{r} \frac{\partial}{\partial \vartheta} + \hat{z} \frac{\partial}{\partial z} \quad \text{gradient} \quad (8.88)$$

$$\nabla_{\bar{a}} F^{\bar{a}} = \frac{1}{r} \frac{\partial (r \hat{r} \cdot \mathbf{F})}{\partial r} + \frac{1}{r} \frac{\partial (\hat{\vartheta} \cdot \mathbf{F})}{\partial \vartheta} + \frac{\partial (\hat{z} \cdot \mathbf{F})}{\partial z}. \quad \text{divergence} \quad (8.89)$$

$$\nabla^2 \psi = \frac{1}{r} \frac{\partial}{\partial r} \left[r \frac{\partial \psi}{\partial r} \right] + \frac{1}{r^2} \frac{\partial^2 \psi}{\partial \vartheta^2} + \frac{\partial^2 \psi}{\partial z^2} \quad \text{Laplacian of scalar} \quad (8.90)$$

$$(\nabla \wedge \mathbf{F})_{\bar{a}} = \varepsilon_{\bar{a}\bar{b}\bar{c}} \partial_{\bar{b}} F^{\bar{c}} \quad \text{see equations (8.84a) -- (8.84c).} \quad \text{curl of a vector} \quad (8.91)$$

8.4 General orthogonal coordinates

We can generalize the spherical and cylindrical coordinates by considering a nonsingular and orthogonal set of coordinates defined such that the metric tensor takes on the diagonal form

$$g_{\bar{a}\bar{b}} = \vec{e}_{\bar{a}} \cdot \vec{e}_{\bar{b}} = \begin{bmatrix} h_{\bar{1}} & 0 & 0 \\ 0 & h_{\bar{2}} & 0 \\ 0 & 0 & h_{\bar{3}} \end{bmatrix}, \quad (8.92)$$

where $h_{\bar{a}} > 0$ are “stretching” functions. The corresponding volume element is expressed as

$$dV = h_{\bar{1}} h_{\bar{2}} h_{\bar{3}} d\xi^{\bar{1}} d\xi^{\bar{2}} d\xi^{\bar{3}}. \quad (8.93)$$

These *generalized orthogonal curvilinear coordinates* have a corresponding orthogonal set of basis vectors

$$\vec{e}_{\bar{a}} = h_{\bar{a}} \hat{\mathbf{e}}_{(\bar{a})} \quad \text{no implied sum.} \quad (8.94)$$

The objects $\hat{\mathbf{e}}_{(\bar{a})}$ are the dimensionless unit directions. The corresponding one-form basis is given by

$$\tilde{e}^{\bar{a}} = (h_{\bar{a}})^{-1} \hat{\mathbf{e}}_{(\bar{a})}. \quad (8.95)$$

The index on the unit directions is enclosed in parentheses to advertise that it is not tensorial; i.e., the unit directions do not transform as tensors. Rather, the functions $h_{\bar{a}}$ carry the tensorial properties of the basis vectors $\vec{e}_{\bar{a}}$. Results for the trajectory and velocity are straightforward generalizations of the spherical results in Section 8.2 and cylindrical-polar results from Section 8.3. A detailed presentation of generalized orthogonal coordinates is found in Section 21.11 of [Griffies \(2004\)](#), with these coordinates commonly used for ocean models.



Differential forms

In this chapter we study *differential forms*, which are built from anti-symmetrized products of the 1-forms introduced in Section 7.2. Differential forms are intimately related to integration and they provide the means to integrate over a manifold even if the manifold has no metric. The lack of a metric contrasts to the *Riemannian differential geometry*, which is central to the tensor analysis of Chapters 1, 6, and 7.

A *manifold* is a collection of points that locally appears to be Euclidean. For example, the surface of a sphere is a non-Euclidean manifold that locally appears Euclidean. Indeed, at each point of the sphere we can define a tangent plane that is a flat and where vectors live. Generalizations of these ideas are available that allow us to extend the notion of a manifold and tangent plane to arbitrary dimensioned spaces that are smooth and orientable. As we will see, the presence of a local Euclidean structure allows for a parsimonious development of calculus on the manifold even if the manifold is not endowed with a metric structure. This formalism is of great use to study coordinate-free fluid mechanics, thermodynamics, and water mass analysis. Just as for our study of tensor analysis, we are mostly concerned with differential forms in space, which conforms to our study of classical mechanics where time is universal and thus has the same value regardless the coordinate reference frame. The more general space-time differential forms, which is the topic of special and general relativity, is not our concern.

READER'S GUIDE TO THIS CHAPTER

Our treatment of differential forms is extremely terse and relatively superficial. Even so, we hope it proves sufficient to whet the appetite by providing tools for rudimentary analysis. This chapter is not central to most matters of this book, with the exception of providing the mathematical basis for our study of water mass configuration space in Section 52.12.

There is a rich literature in physics making use of differential forms, with [Flanders \(1989\)](#) a standard reference that features applications to thermodynamics, fluid mechanics, and Hamiltonian dynamics. Other treatments can be found in the general relatively text by [Misner et al. \(1973\)](#), the field theory monograph from [Göckeler and Schücker \(1989\)](#), and the mathematical physics texts by [Schutz \(1980\)](#), [Nakahara \(2003\)](#), and [Frankel \(2004\)](#).

9.1	Loose threads	166
9.2	Introducing differential forms	166
9.2.1	Wedge products	167
9.2.2	Orientability	167
9.3	The exterior derivative	167
9.3.1	Exterior derivatives	167

9.3.2	Anti-symmetry of the exterior derivative	168
9.3.3	Exterior derivative examples	168
9.3.4	Illustrating $d^2 = 0$	169
9.3.5	Exact and inexact differential forms	169
9.4	Stokes-Cartan theorem	170
9.4.1	The divergence theorem	170
9.4.2	Stokes' theorem	170
9.5	Arbitrary coordinates and dimensions	171
9.5.1	Relating Cartesian and spherical coordinates	171
9.5.2	Relating Cartesian and generalized vertical coordinates	172

9.1 Loose threads

- Nothing ATT.

9.2 Introducing differential forms

As introduced in Section 7.2, a one-form is a function of a vector, and higher order forms are a function of more than one vector. Differential forms are based on both extending and specializing the study of forms. In a nutshell, differential forms are anti-symmetrized products of 1-forms. It turns out that we are already familiar with differential forms as they are the mathematical objects that appear inside of an integral. For example, a line integral along a curve,

$$\int_C (A \, dx + B \, dy + C \, dz), \quad (9.1)$$

has an integrand defining a differential 1-form

$$\mathfrak{A} \equiv A \, dx + B \, dy + C \, dz, \quad (9.2)$$

where A, B, C are arbitrary smooth functions and dx, dy, dz are differential increments of a Cartesian coordinate basis for Euclidean space. Likewise, a surface integral,

$$\int_S [P \, dy \, dz + Q \, dz \, dx + R \, dx \, dy], \quad (9.3)$$

contains a differential 2-form

$$\mathfrak{B} \equiv P \, dy \wedge dz + Q \, dz \wedge dx + R \, dx \wedge dy, \quad (9.4)$$

and the volume integral,

$$\int_V H \, dx \, dy \, dz, \quad (9.5)$$

is built from the 3-form

$$\mathfrak{C} \equiv H \, dx \wedge dy \wedge dz. \quad (9.6)$$

Differential forms of order p are referred to as differential p -forms, and spaces with higher dimensions, $N > 3$, allow for higher order differential forms.

9.2.1 Wedge products

We construct differential forms with $p > 1$ by anti-symmetrizing tensor products of lower order differential forms. As part this anti-symmetrization we are led to the *wedge product*.¹ For example, the wedge product of two differential 1-forms satisfies

$$d\varphi \wedge d\zeta = -d\zeta \wedge d\varphi \implies d\varphi \wedge d\varphi = 0, \quad (9.7)$$

where φ and ζ are arbitrary functions.

Observe that in vector analysis we need a surface normal vector, such as \hat{x} , to orient a surface such as through $\hat{x} dy dz$ or $-\hat{x} dy dz$. In contrast, anti-symmetry provides the differential area 2-form, $dy \wedge dz$, with the ability to orient the surface element. Hence, anti-symmetry of the wedge product makes normal vectors unnecessary. Thinking about the right hand rule of Figure 1.3, the wedge product incorporates the wrapping of the first and second fingers (the oriented swirl), and in so doing it captures the orientation sense of either clockwise or counter-clockwise. However, the wedge product jettisons the thumb since we now see that orientation only requires information within the surface not any information off the surface.

9.2.2 Orientability

Any surface, or more generally a manifold with N -dimensions, can be oriented if there exists an N -form that does not vanish anywhere. This N -form can then be used to determine the volume of the manifold. For example, the differential 3-form, $dx \wedge dy \wedge dz$, serves as the oriented volume element for three-dimensional Euclidean space, and $dx \wedge dy$ is the oriented surface element for a horizontal plane (two-dimensional Euclidean space). Note that a differential N -form is the highest order differential form available in N -dimensional space. The reason is that differential p -forms with $p > N$ all vanish due to anti-symmetry. It follows that all p -forms with $p = N$ are directly proportional to one another.

Another way to understand orientability is to consider a patch of a manifold that can be described by two distinct coordinate systems (e.g., Cartesian and spherical for a portion of the sphere). The manifold is locally orientable if the Jacobian of transformation between the two coordinates is one-signed. The manifold is globally orientable if all such patches maintain a single-signed Jacobian. The Möbius strip is the canonical example of a non-orientable manifold. We do not encounter non-orientable manifolds in this book.

9.3 The exterior derivative

The algebra of differential forms is known as *exterior algebra*, which largely follows from the many consequences of anti-symmetry of the wedge product. Likewise, the calculus of differential forms is known as *exterior calculus*. We here build up some of the basic results of exterior algebra and exterior calculus by studying how the exterior derivative, d , acts on differential forms.

9.3.1 Exterior derivatives

The differential increment operator, d , is a fundamental part of any Riemann integral since it provides the infinitesimal increment needed to perform the integral. Its presence in differential forms supports our use of *differential* in the name. In the study of differential forms, d is referred

¹The wedge symbol, \wedge , is also used in this book for a vector cross product for two dimensional or three dimensional space. Both share the anti-symmetry property.

to as the *exterior derivative*, and it satisfies familiar rules of derivatives. Additionally, when acting on an *exact* differential form (see Section 9.3.4 below), the exterior derivative satisfies the property

$$d d = d^2 = 0. \quad (9.8)$$

For example,

$$d(d\zeta) = 0, \quad (9.9)$$

where ζ is any function. We encounter the many implications of this property in the following.

9.3.2 Anti-symmetry of the exterior derivative

Consider the wedge product of an arbitrary p -form, φ , and a q -form, ζ . The exterior derivative of this wedge product is given by

$$d(\varphi \wedge \zeta) = d\varphi \wedge \zeta + (-1)^p \varphi \wedge d\zeta, \quad (9.10)$$

thus reflecting the anti-symmetry properties of d when acting across the wedge product.

9.3.3 Exterior derivative examples

Taking the exterior derivative of a differential form produces a differential form of one order higher. For example, the exterior derivative of a 0-form (a function) produces a 1-form

$$dA = \partial_x A dx + \partial_y A dy + \partial_z A dz. \quad (9.11)$$

Exterior derivative of a differential 1-form

When acting on differential p -forms, the exterior derivative includes an anti-symmetrization step to produce a $p + 1$ -form. For example, the exterior derivative of a differential 1-form is given by the 2-form

$$d\mathfrak{A} = d(A dx) + d(B dy) + d(C dz) \quad (9.12a)$$

$$= dA \wedge dx + A d^2x + dB \wedge dy + B d^2y + dC \wedge dz + C d^2z \quad (9.12b)$$

$$= dA \wedge dx + dB \wedge dy + dC \wedge dz \quad (9.12c)$$

$$= (\partial_y C - \partial_z B) dy \wedge dz + (\partial_z A - \partial_x C) dz \wedge dx + (\partial_x B - \partial_y A) dx \wedge dy, \quad (9.12d)$$

where we dropped the d^2 terms due to equation (9.8) and anti-symmetry of a wedge product. The final expression reveals the connection to the vector curl operation whereby

$$\nabla \wedge (A \hat{x} + B \hat{y} + C \hat{z}) = (\partial_y C - \partial_z B) \hat{x} + (\partial_z A - \partial_x C) \hat{y} + (\partial_x B - \partial_y A) \hat{z}. \quad (9.13)$$

Exterior derivative of a differential 2-form

The exterior derivative of a differential 2-form is given by the 3-form

$$d\mathfrak{B} = dP dy \wedge dz + dQ dz \wedge dx + dR dx \wedge dy \quad (9.14a)$$

$$= (\partial_x P + \partial_y Q + \partial_z R) dx \wedge dy \wedge dz, \quad (9.14b)$$

which reveals the connection to the vector divergence operator.

Exterior derivative of a differential 3-form

The exterior derivative of a differential 3-form vanishes in 3-space, which we see by

$$d\mathfrak{C} = d(H dx \wedge dy \wedge dz) \quad (9.15a)$$

$$= (\partial_x H dx + \partial_y H dy + \partial_z H dz) \wedge (dx \wedge dy \wedge dz) \quad (9.15b)$$

$$= 0, \quad (9.15c)$$

which follows from the associativity property of the wedge products, as well as anti-symmetry that renders

$$dx \wedge dx = dy \wedge dy = dz \wedge dz = 0. \quad (9.16)$$

9.3.4 Illustrating $d^2 = 0$

As an abstract expression, $d^2 = 0$ as written in equation (9.8) is a bit mysterious. But as shown in this subsection, its manifestation in vector calculus merely reflects commutivity of mixed partial derivatives. Before starting, we note that $d^2x = d^2y = d^2z = 0$, which we used in the examples of Section 9.3.3. These identities result from assuming constant differential increments for each coordinate. That is, as a function, $f(x) = x$ has a constant derivative and thus it has a zero second derivative.

Showing that $d^2A = 0$

The differential 1-form, dA , from equation (9.11) has an exterior derivative given by

$$d^2A = d(\partial_x A dx + \partial_y A dy + \partial_z A dz) \quad (9.17a)$$

$$= (\partial_{yx} A - \partial_{xy} A) dx \wedge dy + (\partial_{zy} A - \partial_{yz} A) dy \wedge dz + (\partial_{xz} A - \partial_{zx} A) dz \wedge dx \quad (9.17b)$$

$$= 0, \quad (9.17c)$$

which follows from equivalence of the mixed partial derivatives.

Showing that $d^2\mathfrak{A} = 0$

The differential 2-form, $d\mathfrak{A}$, from equation (9.12d), has an exterior derivative given by

$$d^2\mathfrak{A} = d[(\partial_y C - \partial_z B) dy \wedge dz] + d[(\partial_z A - \partial_x C) dz \wedge dx] + d[(\partial_x B - \partial_y A) dx \wedge dy] \quad (9.18a)$$

$$= (\partial_{yx} C - \partial_{zx} B) dx \wedge dy \wedge dz + (\partial_{zy} A - \partial_{xy} C) dy \wedge dz \wedge dx + (\partial_{xz} B - \partial_{yz} A) dz \wedge dx \wedge dy \quad (9.18b)$$

$$= (\partial_{zy} A - \partial_{yz} A + \partial_{xz} B - \partial_{zx} B + \partial_{yx} C - \partial_{xy} C) dx \wedge dy \wedge dz \quad (9.18c)$$

$$= 0, \quad (9.18d)$$

where we again made use of the equality of mixed partial derivatives.

9.3.5 Exact and inexact differential forms

Consider the 1-form

$$\omega = \varphi dx + \zeta dy + \psi dz, \quad (9.19)$$

whose exterior derivative is

$$d\omega = (\partial_x \zeta - \partial_y \varphi) dx \wedge dy + (\partial_y \psi - \partial_z \zeta) dy \wedge dz + (\partial_z \varphi - \partial_x \psi) dz \wedge dx. \quad (9.20)$$

We say that ω is an *exact* differential form if the following relations hold

$$\partial_x \zeta = \partial_y \varphi \quad \text{and} \quad \partial_y \psi = \partial_z \zeta \quad \text{and} \quad \partial_z \varphi = \partial_x \psi, \quad (9.21)$$

in which case $d\omega = 0$. With $\omega = dA$ as in equation (9.11), then ω is an exact differential 1-form, as shown by equation (9.17c). However, there are occasions when we cannot write ω as dA , in which case ω is an *inexact* differential form.² Just as for 1-forms, there can be inexact differential p -forms with $p > 1$.

9.4 Stokes-Cartan theorem

The exterior calculus of differential forms provides an elegant unification of the variety of integral theorems from vector calculus studied in Chapters 2 and 7. We refer to the unified integral theorem as the *Stokes-Cartan theorem*, which takes the form

$$\int_V d\omega = \int_{\partial V} \omega. \quad (9.22)$$

If the space is three-dimensional, then V is a volume and ∂V is the surface bounding the volume. If we are instead integrating over a two-dimensional space, then V is a 2-surface whereas ∂V is the one-dimensional curve bounding the surface. Finally, if we are integrating over a curve, then ∂V are the endpoints to the curve. Furthermore, the relation (9.22) holds for a differential form, ω , of arbitrary order. Equation (9.22) is a remarkable unification of the variety of integral theorems encountered in standard tensor analysis. We are not in a position to prove this result, but instead settle for illustrating how it includes the divergence theorem and Stokes' theorem.

9.4.1 The divergence theorem

To connect equation (9.22) to the divergence theorem, let V be equal to a closed volume in 3-space and ω equal to the differential 2-form given by equation (9.4), $\omega = \mathfrak{B}$, in which case $d\omega$ is given by equation (9.14b). The generalized Stokes' theorem (9.22) thus specializes to

$$\int_V d\omega = \int_V (\partial_x P + \partial_y Q + \partial_z R) dx \wedge dy \wedge dz = \oint_{\partial V} [P dy \wedge dz + Q dz \wedge dx + R dx \wedge dy]. \quad (9.23)$$

This equation is an expression of the divergence theorem, whereby the volume integral of the divergence of a vector field equals to the vector field integrated over the oriented area of the surface bounding the volume. We emphasize the absence of a surface normal vector, as the wedge products are sufficient to orient the surface integrals.

9.4.2 Stokes' theorem

Now set ω equal to the one form, $\omega = A dx + B dy + C dz$ as given by equation (9.2), so that $d\omega$ is given by equation (9.12d). We also compute the integral over a 2-surface, S , with a one-dimensional boundary ∂S . The Stokes-Cartan theorem (9.22) thus takes the form

$$\int_S d\omega = \int_S [(\partial_y C - \partial_z B) dy \wedge dz + (\partial_z A - \partial_x C) dz \wedge dx + (\partial_x B - \partial_y A) dx \wedge dy] \quad (9.24a)$$

$$= \oint_{\partial S} (A dx + B dy + C dz), \quad (9.24b)$$

²We encountered a similar definition for exact and inexact differential forms in Section 2.8.

which is the standard expression of Stokes' theorem that is sometimes referred to as *Green's theorem*, where we assumed the right hand rule to orient the closed line integral.

9.5 Arbitrary coordinates and dimensions

Throughout this chapter we have made use of Cartesian coordinates and assumed space to be three-dimensional. However, one of the powerful features of differential forms is that all formula hold regardless the coordinates or the number of dimensions. We make use of that generality when working with a generalized water mass configuration space in Section 52.12. In this section we illustrate a few of the steps needed to move between coordinate representations within the context of differential forms.

9.5.1 Relating Cartesian and spherical coordinates

In Section 8.2 we relate Cartesian coordinates to spherical coordinates, which is expressed by the coordinate transformation

$$x = r \cos \phi \cos \lambda \quad (9.25a)$$

$$y = r \cos \phi \sin \lambda \quad (9.25b)$$

$$z = r \sin \phi, \quad (9.25c)$$

as illustrated by Figure 8.1. We now derive the expression for the oriented volume element using spherical coordinates

$$dx \wedge dy \wedge dz = dx(r, \lambda, \phi) \wedge dy(r, \lambda, \phi) \wedge dz(r, \lambda, \phi), \quad (9.26)$$

with the right hand side exposing the functional relations between the Cartesian coordinates, (x, y, z) , and the spherical coordinates, (r, λ, ϕ) . Performing the exterior derivatives for the Cartesian coordinates renders the expressions

$$dx = dr \cos \phi \cos \lambda - r d\phi \sin \phi \cos \lambda - r d\lambda \cos \phi \sin \lambda \quad (9.27a)$$

$$dy = dr \cos \phi \sin \lambda - r d\phi \sin \phi \sin \lambda + r d\lambda \cos \phi \cos \lambda \quad (9.27b)$$

$$dz = dr \sin \phi + r d\phi \cos \phi, \quad (9.27c)$$

which then leads to the horizontal area element 2-form

$$dx \wedge dy = r \cos^2 \phi dr \wedge d\lambda + r^2 \cos \phi \sin \phi d\lambda \wedge d\phi \quad (9.28)$$

as well as the volume element 3-form

$$dx \wedge dy \wedge dz = r^2 \cos \phi dr \wedge d\lambda \wedge d\phi. \quad (9.29)$$

The final expression reveals the Jacobian of transformation,

$$\frac{\partial(x, y, z)}{\partial(r, \lambda, \phi)} = r^2 \cos \phi, \quad (9.30)$$

which was also derived in Section 8.2.1.

9.5.2 Relating Cartesian and generalized vertical coordinates

Now consider differential volume form represented using the generalized vertical coordinates of Chapter 43

$$x = x \quad (9.31a)$$

$$y = y \quad (9.31b)$$

$$\sigma = \sigma(t, x, y, z), \quad (9.31c)$$

Just as for the spherical coordinates discussed in Section 9.5.1, we write

$$dx \wedge dy \wedge dz = dx \wedge dy \wedge dz(x, y, \sigma), \quad (9.32)$$

with the right hand side exposing the functional relations between the Cartesian coordinates, (x, y, z) , and the generalized vertical coordinates, (x, y, σ) . Performing the spatial exterior derivative leads to the volume element 3-form

$$dx \wedge dy \wedge dz = dx \wedge dy \wedge (\partial z / \partial \sigma) d\sigma, \quad (9.33)$$

thus revealing the Jacobian of transformation,

$$\frac{\partial(x, y, z)}{\partial(x, y, \sigma)} = \partial z / \partial \sigma. \quad (9.34)$$



Part II

Geophysical particle mechanics

The earth's rotation and gravitation are two defining features of atmospheric and oceanic fluid mechanics. Geophysical fluids exhibit motions whose speed is small relative to that of the rotating planet. We thus say that geophysical fluids are in near solid-body motion. Hence, it is very convenient to make use of a rotating (non-inertial) terrestrial reference frame for a physical description.

Rotating and gravitating motion around a sphere introduces many new physical ideas beyond the non-rotating case. We are thus motivated to develop our understanding of geophysical fluids in a stepwise manner to ensure concepts and tools are mastered at each stage. For that purpose, we here study the Newtonian mechanics of a point particle moving around a rotating and gravitating sphere as viewed in the terrestrial (rotating) reference frame. This study of *geophysical particle mechanics* provides a particle mechanics preview of concepts later encountered in geophysical fluid mechanics, such as trajectories, linear momentum, angular momentum, body forces, non-inertial accelerations (Coriolis and centrifugal), symmetries, planetary Cartesian coordinates and planetary spherical coordinates.

Particle mechanics

In this chapter we study the kinematics and dynamics of a point particle of fixed mass and zero electric charge moving around a rotating and gravitating sphere. The motion of the gravitating sphere is prescribed with a fixed kinetic energy and fixed angular momentum around its axis of rotation. Hence, the only interesting mechanics is that of the moving particle. We examine that motion in a spherical geometry as observed from a non-inertial rotating reference frame. This perspective is motivated by our concern with particle motion that is in near solid-body motion with the planet, with such motion relevant to the study of geophysical fluids. Notably, a particle at rest on the rotating sphere has both kinetic energy and angular momentum. Motion relevant to geophysical fluids modifies these solid body values.

The only inertial force acting on the particle arises from the gravitational field of the sphere (ignoring friction and other forces). Two non-inertial accelerations (planetary centrifugal and Coriolis) appear when viewing the motion from the rotating terrestrial reference frame. We make use of both the planetary Cartesian and planetary spherical coordinates to represent the position, velocity, and acceleration vectors, working through both the physics and the maths in support of later descriptions of geophysical fluid motions.

READER'S GUIDE TO THIS CHAPTER

This chapter makes use of basic features of both Cartesian and general tensor algebra as presented in Chapters 1, 6, 7, and Section 8.2. We offer the salient features of tensor technology in this chapter where needed, thus providing a reasonably self-contained presentation. This chapter is notable for working through some relatively tedious algebra as part of developing the spherical coordinate representation of acceleration. This work is relevant also for geophysical fluids, although we rarely encounter the full spherical coordinate equations in later chapters. Even so, the reader is encouraged to read through the details, at least once, as they are part of all numerical models of atmosphere and ocean circulation.

10.1 A survey of Newtonian particle mechanics	176
10.1.1 Mechanics of a point particle	177
10.1.2 Galilean invariance of inertial reference frames	177
10.1.3 Inertial and non-inertial reference frames	178
10.1.4 Angular momentum	178
10.1.5 Mechanical work and kinetic energy	179
10.1.6 Conservative forces	179
10.1.7 Friction as a non-conservative force	179
10.1.8 Comments	180

10.2	The rotating earth	180
10.3	Reference frames and non-inertial accelerations	182
10.4	Rotation of a vector	182
10.4.1	Brief derivation	183
10.4.2	Detailed derivation for planar motion	183
10.4.3	Changes to Cartesian basis vectors under rotation	184
10.4.4	Changes to polar coordinate unit vectors under rotation	184
10.5	Some tensor algebra	184
10.5.1	Why we need general tensors	185
10.5.2	The coordinate representation of a vector	185
10.5.3	Transformation of a coordinate representation	186
10.6	The velocity vector	187
10.6.1	Coordinate velocity	187
10.6.2	Changes to the basis vectors	187
10.7	Inertial acceleration and its decomposition	187
10.8	Representing the position vector	188
10.9	Representing the velocity vector	189
10.9.1	Planetary Cartesian coordinate representation	189
10.9.2	Planetary spherical coordinate representation	190
10.9.3	Transforming from Cartesian to spherical	191
10.9.4	Axial angular momentum	191
10.10	Planetary Cartesian representation of acceleration	191
10.10.1	Planetary Cartesian representation	192
10.10.2	Summary of acceleration representation	193
10.10.3	Further study	194
10.11	Spherical representation of acceleration	194
10.11.1	Transforming from Cartesian to spherical	195
10.11.2	Decomposing the acceleration	195
10.11.3	Spherical coordinate acceleration	196
10.11.4	Metric acceleration	196
10.11.5	Centrifugal acceleration	196
10.11.6	Coriolis acceleration	197
10.11.7	Coriolis acceleration for large-scale motions	197
10.12	Newtonian gravity	198
10.12.1	Newtonian gravity from Poisson's equation	198
10.12.2	Gravitational field for a spherical earth	199
10.12.3	Approximate gravitational acceleration	199
10.12.4	Effective gravitational force from the geopotential	200
10.12.5	Further study	201
10.13	Newton's law of motion	201
10.13.1	Cartesian coordinate representation	201
10.13.2	Spherical coordinate representation	202
10.13.3	Geopotential coordinate representation	203
10.13.4	Comments	203
10.13.5	Further study	204
10.14	Exercises	204

10.1 A survey of Newtonian particle mechanics

Before diving into the mechanics of a particle moving around a rotating sphere, we here summarize foundational elements of Newtonian particle mechanics. Doing so exposes some basic concepts such as space, time, mass, and force. Notably, within the bounds of classical physics these concepts

remain intuitive; i.e., based on common experience, with this perspective sufficient for geophysical fluid motions.

10.1.1 Mechanics of a point particle

Let $\mathbf{X}(t)$ be the position in space of a point particle at a particular time instance, t . The spatial position is measured relative to an arbitrary origin, with the vector \mathbf{X} pointing from the origin to the particle. As the particle moves in time it traces out a *trajectory* through space-time. The velocity of the particle at any time instance is measured by the time derivative

$$\mathbf{V} = \frac{d\mathbf{X}}{dt} = \dot{\mathbf{X}}, \quad (10.1)$$

where the second equality introduced a common “dot” notation for time derivatives that will sometimes be used in this book. Geometrically, the velocity at each point in space-time is tangent to the trajectory.

The *linear momentum* of the particle equals to the mass of the particle, m , times its velocity

$$\mathbf{P} = m \dot{\mathbf{X}}. \quad (10.2)$$

The momentum of the particle changes when it interacts with other objects and/or experiences forces. The vector sum of these forces is written \mathbf{F} , and Newton’s second law of motion states that there exists *inertial reference frames* where motion of the particle is described by the differential equation

$$\frac{d\mathbf{P}}{dt} = \mathbf{F}. \quad (10.3)$$

If the particle mass is fixed, then this equation becomes a second order differential equation for the particle position or trajectory

$$\mathbf{A} = \frac{d\mathbf{V}}{dt} = \frac{d^2\mathbf{X}}{dt^2} = \mathbf{F}/m. \quad (10.4)$$

We refer to $\mathbf{A} = \dot{\mathbf{V}}$ as the *acceleration*.

Many conclusions in mechanics are expressed in terms of conservation laws (Chapter 11), which provide relations or conditions whereby mechanical properties of a physical system remain unchanged in time. Newton’s second law provides our first conservation law. Namely, in the absence of a force, the linear momentum of a physical system remains unchanged: $d\mathbf{P}/dt = 0$. Depending on the nature of the forces, this conservation law might hold for one, two, or all three of the vector components to the linear momentum.

10.1.2 Galilean invariance of inertial reference frames

If the forces acting on the particle are not directly dependent on the particle velocity, then the inertial frame equation of motion (10.4) is unchanged if shifting the velocity by a constant. Such velocity-independent forces (such as Newtonian gravity) are commonly found in reversible (non-dissipative) motion of charge-free particles. This arbitrariness in the velocity reflects a *symmetry* respected by the equation of motion, where symmetry refers to an operation that can be performed on the system without changing any physics. We give the symmetry the name *Galilean invariance*, with Galilean invariance a property of all inertial reference frames in Newtonian mechanics.

These considerations indicate that there is no experiment that can distinguish between two arbitrary inertial reference frames, so long as the experiments are described by Newton’s equation of motion (10.4) and the force is independent of the velocity. The reason we cannot make a

distinction is that the equation of motion is indistinguishable in the two inertial frames. As a corollary, two inertial reference frames can at most be moving relative to one another by a constant velocity. Otherwise, at least one of the reference frames must be accelerating, which in turn would mean that it is not an inertial frame.

This discussion offers an example of the mathematical transformation theory introduced in Part I of this book. Mathematically, the Galilean transformation is written

$$\bar{t} = t \quad \text{and} \quad \bar{\mathbf{X}} = \mathbf{X} + \mathbf{U}t, \quad (10.5)$$

where the barred position vector is measured in the moving reference frame. Time remains unchanged, as per the normal situation with Newtonian universal time. In contrast, the position of the particle in the new frame equals to that in the original reference frame plus a contribution from the constant velocity, \mathbf{U} . We may sometimes refer to the barred reference frame as a *Galilean boosted* frame. The particle velocity in the moving (boosted) reference frame is given by

$$\bar{\mathbf{V}} = \frac{d\bar{\mathbf{X}}}{dt} = \frac{d\mathbf{X}}{dt} + \frac{d(\mathbf{U}t)}{dt} = \mathbf{V} + \mathbf{U}, \quad (10.6)$$

where we set $d\mathbf{U}/dt = 0$ since \mathbf{U} has a fixed magnitude and direction (as per our assumption that it is a constant vector). As expected, the velocity is shifted by the constant reference frame velocity \mathbf{U} . Furthermore, as advertised, the acceleration in the two reference frames is identical

$$\bar{\mathbf{A}} = \frac{d^2\bar{\mathbf{X}}}{dt^2} = \frac{d\mathbf{V}}{dt} = \mathbf{A}. \quad (10.7)$$

10.1.3 Inertial and non-inertial reference frames

As seen above, inertial reference frames (also called *absolute* reference frames) have a special status in Newtonian mechanics since it is in these reference frames that Newton's second law holds as per equations (10.3) or (10.4). In practice, inertial reference frames are an idealization never met exactly. For experiments taking place within many laboratories, the earth or *laboratory reference frame* provide a good approximation to an inertial reference frame. However, when the scale of the physical system under study increases, then it can become untenable to ignore the earth's rotation. In such cases the rotating earth reference frame is no longer a good approximation to an inertial frame, with this situation holding for most geophysical fluid motions.

10.1.4 Angular momentum

The angular momentum of a particle defined relative to a chosen origin is given by the vector cross product

$$\mathbf{L} = \mathbf{X} \wedge \mathbf{P}. \quad (10.8)$$

Making use of Newton's second law (10.3) leads to the equation of motion for the angular momentum

$$\frac{d\mathbf{L}}{dt} = \mathbf{X} \wedge \mathbf{F}, \quad (10.9)$$

where we set $\dot{\mathbf{X}} \wedge \mathbf{P} = \dot{\mathbf{X}} \wedge m\dot{\mathbf{X}} = 0$. The cross product, $\mathbf{X} \wedge \mathbf{F}$, is the *torque* acting on the system relative to the origin. Equation (10.9) leads to our second conservation law: a particle has a constant angular momentum when experiencing zero torques, with this statement dependent on the choice of origin for the angular momentum and the corresponding torques.

10.1.5 Mechanical work and kinetic energy

When a force is applied to a particle as it moves along its trajectory, the force does *mechanical work* on the particle. The work is computed by the line integral along the trajectory

$$W = \int_{\mathbf{x}_1}^{\mathbf{x}_2} \mathbf{F} \cdot d\mathbf{x}, \quad (10.10)$$

where

$$\mathbf{x}_1 = \mathbf{X}(t = t_1) \quad \text{and} \quad \mathbf{x}_2 = \mathbf{X}(t = t_2) \quad (10.11)$$

are the spatial coordinates of the endpoints for the trajectory at times t_1 and t_2 , and $d\mathbf{x}$ is the differential vector increment along the trajectory. Since the particle is moving along its dynamical trajectory, we can write $d\mathbf{x} = \mathbf{V} dt$ and make use of Newton's equation of motion (10.4) to reach

$$W = \int_{\mathbf{x}_1}^{\mathbf{x}_2} \mathbf{F} \cdot d\mathbf{x} = \frac{m}{2} \int_{\mathbf{x}_1}^{\mathbf{x}_2} \frac{d(\mathbf{V} \cdot \mathbf{V})}{dt} dt = K(t_2) - K(t_1), \quad (10.12)$$

where we identified the kinetic energy of the particle

$$K = \frac{m}{2} \mathbf{V} \cdot \mathbf{V}. \quad (10.13)$$

We conclude that the work done on the particle over a time interval is equal to its change in kinetic energy. This result is called the *work-energy theorem*.

10.1.6 Conservative forces

If the work done on the particle between two points in space is independent of the path taken between the points, then the force is said to be *conservative*. Recalling our discussion of exact differentials in Section 2.8, we know that a conservative force can be written as the gradient of a *force potential*¹

$$\mathbf{F} = -\nabla V, \quad (10.14)$$

where V is called the *potential* or the *potential energy*. Inserting the potential into the work equation (10.12) leads to an expression of mechanical energy conservation

$$W = \int_{\mathbf{x}_1}^{\mathbf{x}_2} \mathbf{F} \cdot d\mathbf{x} = \Delta K = -\Delta V \implies K(t_2) + V(t_2) = K(t_1) + V(t_1). \quad (10.15)$$

That is, the sum of the kinetic plus potential energy remains constant for a particle moving in a conservative force field. Gravity and the planetary centrifugal force are two conservative forces encountered in our studies of geophysical particle and fluid mechanics. The conservation of kinetic plus potential energy within a conservative force field offers our third conservation law: the conservation of mechanical energy.

10.1.7 Friction as a non-conservative force

Friction is the canonical non-conservative force that typically depends on the velocity field. For example, a common form of friction is given by *Rayleigh drag*

$$\mathbf{F}_{\text{Rayleigh}} = -\gamma m \mathbf{V}, \quad (10.16)$$

¹Do not confuse the potential, V , in equation (10.14) with the magnitude of the velocity, $|\mathbf{V}|$. They are distinct.

where $\gamma > 0$ is a constant with dimensions of inverse time. Newton's equation of motion with Rayleigh drag (and no other forces) takes the form

$$\frac{d\mathbf{V}}{dt} = -\gamma \mathbf{V}. \quad (10.17)$$

Notably, Rayleigh drag is not Galilean invariant since it is dependent on the velocity. We can understand this lack of invariance by noting that the friction force distinguishes the state of rest ($\mathbf{V} = 0$) in a particular reference frame from motion in a Galilean boosted reference frame.

The solution to the first order ordinary differential equation (10.17) is the exponential decay

$$\mathbf{V}(t) = \mathbf{V}(0) e^{-\gamma t}, \quad (10.18)$$

with $\mathbf{V}(0)$ the initial condition. We thus see that Rayleigh drag exponentially drives the velocity towards zero. Correspondingly, Rayleigh drag dissipates the kinetic energy according to twice the exponential decay

$$\frac{dK}{dt} = -2\gamma K \implies K(t) = K(0) e^{-2\gamma t}, \quad (10.19)$$

where $K(0) = \mathbf{V}(0) \cdot \mathbf{V}(0)/2$.

10.1.8 Comments

Every classical mechanics textbook has some form of the material presented in this section. The book from [Goldstein \(1980\)](#) is targeted at the entering physics graduate student, whereas [Marion and Thornton \(1988\)](#) is targeted at second year undergraduates.

We now move from generalities to the special case of a particle moving around a rotating and gravitating sphere. Throughout that discussion it can be useful to return to the general considerations presented in this section as a reminder of the fundamental physical notions, which are quite simple in their essence. Such reminders are particularly useful when encountering the somewhat tedious mathematical manipulations with spherical coordinates in a rotating reference frame.

10.2 The rotating earth

The earth's angular velocity is comprised of two main contributions: the spin of the earth about its axis and the orbit of the earth about the sun (see Figure 10.1). Other astronomical motions can be neglected for geophysical fluid mechanics. Therefore, in the course of a single period of 24 hours, or $24 \times 3600 = 86400$ seconds, the earth experiences an angular rotation of $(2\pi + 2\pi/365.24)$ radians. As such, the angular rotation rate is given by

$$\Omega = \frac{2\pi + 2\pi/365.24}{86400\text{s}} = \left[\frac{\pi}{43082} \right] \text{s}^{-1} = 7.2921 \times 10^{-5} \text{ s}^{-1}. \quad (10.20)$$

The earth's angular velocity, both its direction through the polar axis and its magnitude, is assumed constant in time for purposes of geophysical fluid mechanics

$$\frac{d\Omega}{dt} = 0. \quad (10.21)$$

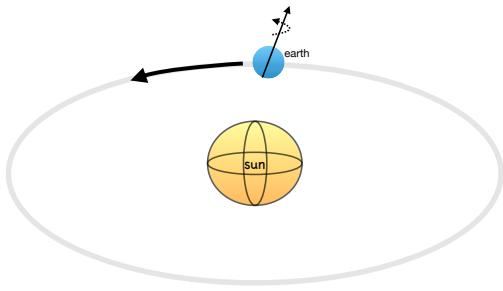


FIGURE 10.1: The angular velocity of the earth arises from the spin about the polar axis plus the orbit of the planet around the sun. This angular velocity determines the strength of the Coriolis acceleration and the planetary centrifugal acceleration.

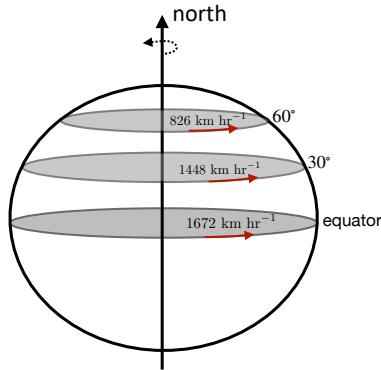


FIGURE 10.2: For an observer at rest on the earth's surface, the solid body speed is $\Omega R_e \cos \phi = 7.2921 \times 10^{-5} \text{ s}^{-1} \times 6.371 \times 10^6 \text{ m} \cos \phi \approx 465 \text{ m s}^{-1} \cos \phi = 1672 \text{ km hr}^{-1} \cos \phi$. We here display solid-body speeds for motion at the equator, 30° latitude, and 60° latitude. These speeds are much larger than motion of the ocean and atmosphere relative to the earth, thus motivating a description of geophysical fluid motion from the rotating terrestrial frame moving with the solid-body.

The angular speed (10.20) seems quite small. However, a terrestrial reference frame on the surface of the earth, undergoing *solid body motion*, moves with linear speed

$$U_{\text{solid-body}}(\phi) = \Omega R_e \cos \phi \approx 465 \text{ m s}^{-1} \cos \phi = 1672 \text{ km hr}^{-1} \cos \phi, \quad (10.22)$$

where we set the earth's radius to

$$R_e = 6.371 \times 10^6 \text{ m}. \quad (10.23)$$

This speed is quite large relative to a fixed frame outside the planet (see Figure 10.2)

$$U_{\text{solid-body}}(0) = 1672 \text{ km hr}^{-1} \quad (10.24a)$$

$$U_{\text{solid-body}}(30^\circ) = 1448 \text{ km hr}^{-1} \quad (10.24b)$$

$$U_{\text{solid-body}}(60^\circ) = 826 \text{ km hr}^{-1}. \quad (10.24c)$$

Even so, motion of the atmosphere and ocean are generally quite close to the solid-body motion, so that their speeds relative to the planet are much smaller than the speed of the rotating planet itself. Therefore, it is sensible to describe geophysical fluid motion from the non-inertial rotating terrestrial reference frame.

10.3 Reference frames and non-inertial accelerations

Motion of geophysical fluids deviates relatively little from solid-body motion. That is, motion of fluid elements deviates relatively little from motion of the earth itself. For these reasons, the preferred frame for studying geophysical motion is the rotating planetary frame (a non-inertial frame) rather than a frame fixed relative to the stars (an approximate inertial frame). Figure 10.3 illustrates the case for the position vector of a particle relative to the origin of a rotating sphere.

A set of basis vectors is needed in order to use coordinates for representing vectors. The basis vectors hold two key pieces of information, with the first being details of the coordinates. The second concerns the reference frame, whereby non-inertial accelerations arise from time dependence to the basis vectors. When represented in terms of non-inertial reference frame coordinates, the inertial acceleration (a vector) is decomposed into the sum of relative acceleration (relative to the non-inertial frame), centrifugal acceleration, and Coriolis acceleration.

When multiplied by mass, non-inertial accelerations can be interpreted as non-inertial forces. However, these forces are not imparted by an external force. Rather, they arise from accelerated motion of the non-inertial reference frame. In this sense, non-inertial accelerations are often termed “fictitious” or “kinematic”. Nonetheless, a terrestrial observer interprets motion that includes non-inertial accelerations, so that non-inertial accelerations are central to rationalizing observed planetary fluid motions. Indeed, as seen in Section 11.8, the Coriolis acceleration that appears in a non-inertial reference frame has its counterpart in axial angular momentum conservation, with angular momentum conservation a property of all reference frames. Hence, there is nothing imaginary about non-inertial accelerations.

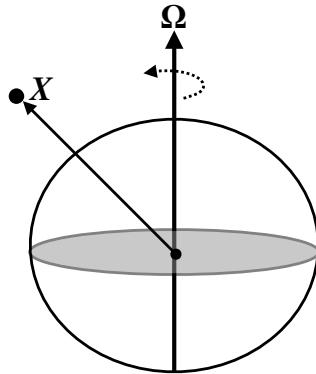


FIGURE 10.3: The position vector, $\mathbf{X}(t)$, for a particle moving around a rotating sphere with coordinate origins at the center of the sphere. The rotation axis is through the north pole, with angular velocity vector $\boldsymbol{\Omega}$. The sphere rotates in a positive right hand sense through the north polar axis (counter clockwise from above). The rotating frame has a solid-body velocity $\mathbf{U}_{\text{solid}} = \boldsymbol{\Omega} \wedge \mathbf{X}$ (equation (10.52)).

10.4 Rotation of a vector

How does a vector change under a solid-body rotation such as that shown in Figure 10.4? That is, when a vector is assumed to not translate nor to change its magnitude, how does it change as a result of solid-body rotation? To answer this question, observe that a pure rotation does not change the magnitude of a vector so that

$$|\mathbf{X}(t)| = |\mathbf{X}(t + \delta t)|. \quad (10.25)$$

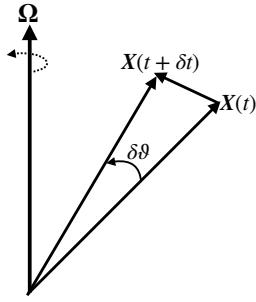


FIGURE 10.4: The change in a vector under a pure rotation leaves its origin fixed and the vector magnitude unchanged, $|\mathbf{X}(t)| = |\mathbf{X}(t + \delta t)|$. Only the vector direction changes by the angle $\delta\vartheta = \Omega \delta t$. Infinitesimal changes generated by the angular velocity $\boldsymbol{\Omega}$ lead to the vector differences $\mathbf{X}(t + \delta t) - \mathbf{X}(t) = \delta t \boldsymbol{\Omega} \wedge \mathbf{X}(t)$.

The condition (10.25) can be written as

$$\frac{d(\mathbf{X} \cdot \mathbf{X})}{dt} = 0, \quad (10.26)$$

which leads to the constraint

$$\mathbf{X} \cdot \frac{d\mathbf{X}}{dt} = 0. \quad (10.27)$$

That is, the velocity generated by a pure rotation is itself perpendicular to the position. We encountered this result in Section 2.1.4 when showing that unit vectors (vectors of a fixed unit length) can only change through rotations. What we see here is that any vector of fixed magnitude and fixed origin can change only through a rotation of its direction.

10.4.1 Brief derivation

Referring to Figure 10.4, we see that the infinitesimal difference $\mathbf{X}(t + \delta t) - \mathbf{X}(t)$ equals to the vector cross product of the angular velocity with the position vector

$$\mathbf{X}(t + \delta t) - \mathbf{X}(t) = \delta t \boldsymbol{\Omega} \wedge \mathbf{X}(t). \quad (10.28)$$

Dividing by δt leads to

$$\frac{d\mathbf{X}}{dt} = \boldsymbol{\Omega} \wedge \mathbf{X} \quad \text{pure rotation.} \quad (10.29)$$

Note that this evolution equation satisfies the constraint (10.27) since $\mathbf{X} \cdot (\boldsymbol{\Omega} \wedge \mathbf{X}) = 0$, meaning that the magnitude of the vector indeed remains fixed. We sometimes say that $\boldsymbol{\Omega}$ is the generator of rotations of \mathbf{X} .

10.4.2 Detailed derivation for planar motion

To determine the change in direction generated by a pure rotation, first consider the simplified case of planar rotation. Let $\boldsymbol{\Omega}$ be entirely in the vertical, and let \mathbf{X} be confined to the horizontal plane. In a time increment δt , the vector $\mathbf{X}(t)$ is rotated by an angle

$$\delta\vartheta = |\boldsymbol{\Omega}| \delta t \quad (10.30)$$

relative to $\mathbf{X}(t + \delta t)$. In the limit of small $\delta\vartheta$, the difference vector, $\mathbf{X}(t + \delta t) - \mathbf{X}(t)$, is perpendicular to $\mathbf{X}(t)$ and is of magnitude equal to the arc length

$$\delta s = |\mathbf{X}(t)| \delta\vartheta = |\mathbf{X}(t)| |\boldsymbol{\Omega}| \delta t. \quad (10.31)$$

We observe that the vector $\boldsymbol{\Omega} \wedge \mathbf{X}(t)$ points in the same direction as $\mathbf{X}(t + \delta t) - \mathbf{X}(t)$ and is of length $|\mathbf{X}(t)| |\boldsymbol{\Omega}|$. We conclude that

$$\mathbf{X}(t + \delta t) - \mathbf{X}(t) = \boldsymbol{\Omega} \wedge \mathbf{X}(t) \delta t. \quad (10.32)$$

Dividing through by δt and taking the limit $\delta t \rightarrow 0$ gives

$$\frac{d\mathbf{X}}{dt} = \boldsymbol{\Omega} \wedge \mathbf{X} \quad \text{pure rotation.} \quad (10.33)$$

The proof for the case in which $\boldsymbol{\Omega}$ has a component along \mathbf{X} is a straightforward generalization. The trajectory is still confined to a plane, but only the component of $\boldsymbol{\Omega}$ normal to the trajectory generates rotation.

10.4.3 Changes to Cartesian basis vectors under rotation

We make use of the result (10.29) in many places when working with rotating physics. One place of particular note concerns the changes to unit vectors under rotation. Cartesian unit vectors exhibit changes solely due to pure rotations. With $\boldsymbol{\Omega} = \Omega \hat{\mathbf{z}}$ as per a rotating spherical earth, the change in planetary Cartesian unit vectors under rotation is given by

$$\frac{d\hat{\mathbf{x}}}{dt} = \boldsymbol{\Omega} \wedge \hat{\mathbf{x}} = \Omega \hat{\mathbf{y}} \quad \text{and} \quad \frac{d\hat{\mathbf{y}}}{dt} = \boldsymbol{\Omega} \wedge \hat{\mathbf{y}} = -\Omega \hat{\mathbf{x}} \quad \text{and} \quad \frac{d\hat{\mathbf{z}}}{dt} = \boldsymbol{\Omega} \wedge \hat{\mathbf{z}} = 0. \quad (10.34)$$

10.4.4 Changes to polar coordinate unit vectors under rotation

As another example of how unit vectors change under rotation, consider the polar coordinates from Section 8.3 with rotation assumed to be about the vertical axis, $\hat{\mathbf{z}}$. In the horizontal plane we have the radial and angular unit vectors

$$\hat{\mathbf{r}} = \hat{\mathbf{x}} \cos \vartheta + \hat{\mathbf{y}} \sin \vartheta \quad (10.35a)$$

$$\hat{\boldsymbol{\vartheta}} = -\hat{\mathbf{x}} \sin \vartheta + \hat{\mathbf{y}} \cos \vartheta. \quad (10.35b)$$

We can directly verify that the time derivative of these unit vectors is given by

$$\frac{d\hat{\mathbf{r}}}{dt} = (\Omega + \dot{\vartheta}) \hat{\boldsymbol{\vartheta}} = (\Omega + \dot{\vartheta}) (\hat{\mathbf{z}} \wedge \hat{\mathbf{r}}) \quad (10.36a)$$

$$\frac{d\hat{\boldsymbol{\vartheta}}}{dt} = -(\Omega + \dot{\vartheta}) \hat{\mathbf{r}} = (\Omega + \dot{\vartheta}) (\hat{\mathbf{z}} \wedge \hat{\boldsymbol{\vartheta}}). \quad (10.36b)$$

Notice how these unit vectors change both due to the rotation of the reference frame, Ω , plus the change in the angular position relative to the reference frame, $\dot{\vartheta}$. We thus see that unit vectors, such as those for spherical coordinates and cylindrical-polar coordinates, change due to the solid-body rotation, just like the Cartesian unit vectors. Additionally, non-Cartesian unit vectors change when their orientation is modified relative to the Cartesian coordinate axes at a rate distinct from the solid-body. This situation occurs when a trajectory moves relative to the solid-body. We further discuss this notion in Section 10.6.2 and detail the spherical coordinate case in Sections 10.9.2 and 10.11.

10.5 Some tensor algebra

In Part I of this book, we detailed the use of tensor analysis for geophysical motions. We here summarize some of the salient points that will be of use later in this chapter when dealing with spherical coordinates.

10.5.1 Why we need general tensors

Cartesian tensors are sufficient for many purposes of fluid mechanics, such as when using Cartesian coordinates for a tangent plane approximation to study geophysical fluid motion (e.g., Section 23.2). However, we make routine use of spherical coordinates when describing geophysical motion, and cylindrical-polar coordinates for studies of rotating tank experiments (see Section 32.7). Finally, we use generalized vertical coordinates in the description of stratified flows (Chapters 43, 44, and 45). The basis vectors for curvilinear coordinates and generalized vertical coordinates change direction when moving through space. In contrast, Cartesian basis vectors always point in the same direction. This distinction between the basis vectors is the key reason curvilinear coordinates and generalized vertical coordinates require a more general formalism than afforded by Cartesian tensors.

10.5.2 The coordinate representation of a vector

The coordinate representation of a vector follows from decomposing the vector into components aligned according to a set of basis vectors. In particular, the coordinate representation of the position vector is given by

$$\vec{X} = \sum_{a=1}^3 \xi^a \vec{e}_a = \xi^a \vec{e}_a, \quad (10.37)$$

where the Einstein summation convention is defined by the final equality. In this equation,

$$\vec{e}_a = (\vec{e}_1, \vec{e}_2, \vec{e}_3) \quad (10.38)$$

is a set of linearly independent basis vectors, and ξ^a are the corresponding coordinate representations of the position vector \vec{X} . The basis vectors may be normalized to unit magnitude, as in the case of Cartesian coordinates, or may be unnormalized as for spherical coordinates (see Section 8.2.2). Note that we commonly make use of the boldface notation for a vector rather than the arrowed symbol (Section 6.5)

$$\boldsymbol{X} = \vec{X} = \xi^a \vec{e}_a. \quad (10.39)$$

The basis vectors in equation (10.37) have a lower index while the coordinate representation of a vector has an upper index. Why? For arbitrary coordinates (e.g., spherical), we make a distinction between a coordinate representation with an index upstairs (contravariant) versus the downstairs (covariant) representation. Moving between the covariant and contravariant representations requires a metric tensor. For much of this book we can keep this mathematical framework at a modest distance, with exposure only in selected places. The key crutch we rely on is that the planet is assumed to be embedded in a background Euclidean space \mathbb{R}^3 . That is, we are not considering the curved space-time of general relativity nor the marriage of space and time afforded by special relativity. These assumptions simplify our use of general tensors.

When working with general coordinates, it is necessary to distinguish between a basis vector \vec{e}_a and its dual partner known as a *one-form*, \tilde{e}^a . Duality here is defined using the familiar (Euclidean) inner product that leads to the bi-orthogonality relation

$$\vec{e}_a \cdot \tilde{e}^b = \delta_a^b, \quad (10.40)$$

with δ_a^b the Kronecker delta tensor

$$\delta_a^b = \begin{cases} 1 & \text{if } b = a \\ 0 & \text{if } b \neq a. \end{cases} \quad (10.41)$$

In linear algebra, a row vector is dual to its column vector, with that analog appropriate for the present context. Cartesian basis vectors equal to the basis one-forms, in which case there is no distinction between contravariant and covariant. However, the distinction is important for the general coordinates used in geophysical fluids such as encountered in Chapters 8 and 43.

10.5.3 Transformation of a coordinate representation

A vector is a geometric object; i.e., a line determined by its length and direction. More general tensors are also geometric objects. The coordinate representation of a tensor is a subjective description of the object. More specifically, the coordinate representation of a vector as given by equation (10.37) does not alter the vector, it only alters the representation of the vector. Hence, we can represent the vector using arbitrary coordinates

$$\vec{X} = \xi^a \vec{e}_a = \xi^{\bar{a}} \vec{e}_{\bar{a}}, \quad (10.42)$$

where $\xi^{\bar{a}}$ is the representation of the vector in a coordinate system defined by the basis vector $\vec{e}_{\bar{a}}$, whereas ξ^a is the representation in the unbarred coordinate system with basis vectors \vec{e}_a . In this chapter we choose the unbarred coordinates as planetary Cartesian and the barred coordinates as planetary spherical

$$(\xi^1, \xi^2, \xi^3) = (x, y, z) \quad \text{and} \quad (\xi^{\bar{1}}, \xi^{\bar{2}}, \xi^{\bar{3}}) = (r, \lambda, \phi), \quad (10.43)$$

with the coordinates related by (see Section 8.2 and Figure 8.1)

$$x = r \cos \phi \cos \lambda \quad \text{and} \quad y = r \cos \phi \sin \lambda \quad \text{and} \quad z = r \sin \phi. \quad (10.44)$$

The transformation of coordinate representations for a vector and basis vectors is provided by the transformation matrix and its inverse. For example, the relation between the coordinate representation of the velocity vector and acceleration vector, as well as the coordinate basis vectors, are given by

$$V^{\bar{a}} = \Lambda_{\bar{a}}^{\bar{a}} V^a \quad \text{and} \quad A^{\bar{a}} = \Lambda_{\bar{a}}^{\bar{a}} A^a \quad \text{and} \quad \vec{e}_{\bar{a}} = \Lambda_{\bar{a}}^a \vec{e}_a, \quad (10.45)$$

where the transformation matrix and its inverse are given by (see Section 8.2.1)

$$\Lambda_{\bar{a}}^a = \begin{bmatrix} \partial \xi^1 / \partial \xi^{\bar{1}} & \partial \xi^1 / \partial \xi^{\bar{2}} & \partial \xi^1 / \partial \xi^{\bar{3}} \\ \partial \xi^2 / \partial \xi^{\bar{1}} & \partial \xi^2 / \partial \xi^{\bar{2}} & \partial \xi^2 / \partial \xi^{\bar{3}} \\ \partial \xi^3 / \partial \xi^{\bar{1}} & \partial \xi^3 / \partial \xi^{\bar{2}} & \partial \xi^3 / \partial \xi^{\bar{3}} \end{bmatrix} = \begin{bmatrix} -r \cos \phi \sin \lambda & -r \sin \phi \cos \lambda & \cos \phi \cos \lambda \\ r \cos \phi \cos \lambda & -r \sin \phi \sin \lambda & \cos \phi \sin \lambda \\ 0 & r \cos \phi & \sin \phi \end{bmatrix} \quad (10.46)$$

$$\Lambda_{\bar{a}}^{\bar{a}} = \frac{1}{r^2 \cos \phi} \begin{bmatrix} -r \sin \lambda & r \cos \lambda & 0 \\ -r \cos \phi \sin \phi \cos \lambda & -r \cos \phi \sin \phi \sin \lambda & r \cos^2 \phi \\ r^2 \cos^2 \phi \cos \lambda & r^2 \cos^2 \phi \sin \lambda & r^2 \cos \phi \sin \phi \end{bmatrix}, \quad (10.47)$$

with

$$\Lambda_{\bar{a}}^{\bar{a}} \Lambda_{\bar{b}}^a = \delta_{\bar{b}}^{\bar{a}} \quad \text{and} \quad \Lambda_{\bar{a}}^a \Lambda_{\bar{b}}^{\bar{a}} = \delta_b^a. \quad (10.48)$$

We thus see that if we have a Cartesian representation to a vector, such as the velocity or acceleration, then we can use the above transformation rules to derive the spherical representation.

10.6 The velocity vector

The velocity is the time derivative of the position vector

$$\mathbf{V} = \frac{d\mathbf{X}}{dt}. \quad (10.49)$$

This equation is manifestly geometric in that the velocity at a space-time point is the tangent vector to the trajectory at that point. When expressing this relation using a coordinate representation we note that time dependence lives with both the coordinate representation of the position vector as well as the basis vectors

$$\mathbf{V} = \frac{d\mathbf{X}}{dt} = \frac{d(\xi^a \vec{e}_a)}{dt} = \frac{d\xi^a}{dt} \vec{e}_a + \xi^a \frac{d\vec{e}_a}{dt}. \quad (10.50)$$

10.6.1 Coordinate velocity

The first term on the right hand side of equation (10.50) is the velocity as measured within the chosen reference frame using the chosen coordinates

$$\mathbf{V}_{\text{coord}} \equiv \frac{d\xi^a}{dt} \vec{e}_a. \quad (10.51)$$

This is the contribution to velocity as measured in the reference frame that moves with the basis vectors. In the context of geophysical motions, this is the velocity measured in the rotating terrestrial reference frame.

10.6.2 Changes to the basis vectors

The second term on the right hand side of equation (10.50) arises from changes to the basis vectors. There are three means for a basis vector to change, and we encounter them when considering coordinate representations later in this chapter.

- **SOLID-BODY ROTATION:** For a solid-body rotation of the reference frame, the solid-body velocity is given by (see Section 10.4)

$$\mathbf{U}_{\text{solid}} = \boldsymbol{\Omega} \wedge \mathbf{X}. \quad (10.52)$$

- **ROTATION RELATIVE TO SOLID-BODY:** We can also change a vector through rotation by changing its direction at a rate that differs from the solid-body.
- **CHANGE IN MAGNITUDE:** Finally, if the basis vectors are not normalized, then they can change their magnitude during motion (stretching or compression).

10.7 Inertial acceleration and its decomposition

The inertial acceleration is given by the time derivative of the inertial velocity, which is the second derivative of the position vector

$$\mathbf{A} = \frac{d\mathbf{V}}{dt} = \frac{d^2\mathbf{X}}{dt^2}. \quad (10.53)$$

This equation is independent of any coordinate representation so that the physical and geometrical content are manifest. When introducing a coordinate representation, the resulting expression

becomes subject to details of the chosen coordinates and those details possibly obscure the underlying geometric basis. Hence, it is important to keep the geometric form in mind when offering an interpretation for coordinate dependent terms that arise in a coordinate representation.

Introducing a coordinate representation $\mathbf{X} = \xi^a \vec{e}_a$ into the acceleration (10.53), and making use of the chain rule, leads to

$$\mathbf{A} = \frac{d}{dt} \frac{d\mathbf{X}}{dt} \quad (10.54a)$$

$$= \frac{d}{dt} \frac{d(\xi^a \vec{e}_a)}{dt} \quad (10.54b)$$

$$= \frac{d}{dt} \left[\frac{d\xi^a}{dt} \vec{e}_a + \xi^a \frac{d\vec{e}_a}{dt} \right] \quad (10.54c)$$

$$= \frac{d^2\xi^a}{dt^2} \vec{e}_a + 2 \frac{d\xi^a}{dt} \frac{d\vec{e}_a}{dt} + \xi^a \frac{d^2\vec{e}_a}{dt^2}. \quad (10.54d)$$

The first term on the right hand side is the acceleration of the coordinate representation as measured in the rotating reference frame

$$\mathbf{A}_{\text{coord}} \equiv \frac{d^2\xi^a}{dt^2} \vec{e}_a. \quad (10.55)$$

It is the acceleration measured by an observer in the rotating frame using coordinates ξ^a . The remaining two terms arise from changes to the basis vectors, and they give rise to the Coriolis and centrifugal accelerations associated with the rotating reference frame. In non-Cartesian coordinates, they also give rise to a “metric acceleration” arising from the change in directions of the unit vectors associated with motion of the particle relative to the rotating reference frame (Section 10.6.2).

Some presentations of the kinematic result (10.54d) suggest that the factor of two on the middle term (the Coriolis term) is mysterious. In fact, there is nothing mysterious. Rather, the factor of two merely results from the need to take two time derivatives of the basis vectors as part of a representation of acceleration.

10.8 Representing the position vector

We make use of some results from Section 8.2 relating Cartesian and spherical coordinates and as furthermore defined by Figure 8.1. Starting with the position vector, we introduce the *planetary* Cartesian basis vectors, $(\hat{\mathbf{x}}, \hat{\mathbf{y}}, \hat{\mathbf{z}})$, and corresponding spherical basis vectors, $(\hat{\lambda}, \hat{\phi}, \hat{r})$. We thus have the suite of equivalent expressions for the position of a particle moving around the sphere

$$\mathbf{X} = x \hat{\mathbf{x}} + y \hat{\mathbf{y}} + z \hat{\mathbf{z}} \quad (10.56a)$$

$$= (r \cos \phi \cos \lambda) \hat{\mathbf{x}} + (r \cos \phi \sin \lambda) \hat{\mathbf{y}} + (r \sin \phi) \hat{\mathbf{z}} \quad (10.56b)$$

$$= r \hat{r} \quad (10.56c)$$

$$= |\mathbf{X}| \hat{r}. \quad (10.56d)$$

Note how the expression for the position vector is quite simple when written in spherical coordinates, as it is merely the distance from the origin with a direction that points radially from the origin to the particle.

10.9 Representing the velocity vector

As seen in Section 10.6, the inertial velocity vector has a coordinate representation written as

$$\mathbf{V} = \frac{d\mathbf{X}}{dt} = \frac{d(\xi^a \vec{e}_a)}{dt} = \frac{d\xi^a}{dt} \vec{e}_a + \xi^a \frac{d\vec{e}_a}{dt}. \quad (10.57)$$

Contributions arise from both the time changes in the coordinates, ξ^a , and time changes to the basis vectors, \vec{e}_a . We now consider the Cartesian and spherical forms for these changes.

10.9.1 Planetary Cartesian coordinate representation

The basis vectors for the Cartesian coordinates, $(\vec{e}_1, \vec{e}_2, \vec{e}_3) = (\hat{\mathbf{x}}, \hat{\mathbf{y}}, \hat{\mathbf{z}})$, are normalized, so they do not change their magnitude. Furthermore, they move only through solid-body motion of the rotating reference frame. We refer to these coordinates as *planetary Cartesian coordinates* since they are oriented according to the rotating planet. In Section 23.2 we introduce the distinct *tangent plane Cartesian coordinates*. Tangent plane Cartesian coordinates are also moving with the rotating planet. Yet they are defined according to a tangent plane at a point on the surface of the sphere. It is important to distinguish to two uses for Cartesian coordinates in geophysical fluid mechanics.

The angular velocity is oriented around the polar axis

$$\boldsymbol{\Omega} = \Omega \hat{\mathbf{z}}, \quad (10.58)$$

so that the solid-body velocity only has components in the $\hat{\mathbf{x}}$ and $\hat{\mathbf{y}}$ directions

$$\mathbf{U}_{\text{solid}} = \boldsymbol{\Omega} \wedge \mathbf{X} = \Omega (-\hat{\mathbf{x}} y + \hat{\mathbf{y}} x). \quad (10.59)$$

The inertial velocity thus has the following representation in terms of planetary Cartesian coordinates

$$\mathbf{V} = \frac{d}{dt} [\hat{\mathbf{x}} x + \hat{\mathbf{y}} y + \hat{\mathbf{z}} z] \quad (10.60a)$$

$$= \left[\hat{\mathbf{x}} \frac{dx}{dt} + \hat{\mathbf{y}} \frac{dy}{dt} + \hat{\mathbf{z}} \frac{dz}{dt} \right] + x \frac{d\hat{\mathbf{x}}}{dt} + y \frac{d\hat{\mathbf{y}}}{dt} + z \frac{d\hat{\mathbf{z}}}{dt} \quad (10.60b)$$

$$= \left[-y \Omega + \frac{dx}{dt} \right] \hat{\mathbf{x}} + \left[x \Omega + \frac{dy}{dt} \right] \hat{\mathbf{y}} + \frac{dz}{dt} \hat{\mathbf{z}} \quad (10.60c)$$

$$= \mathbf{V}_{\text{Cartesian}} + \boldsymbol{\Omega} \wedge \mathbf{X} \quad (10.60d)$$

$$= \mathbf{V}_{\text{Cartesian}} + \mathbf{U}_{\text{solid}}, \quad (10.60e)$$

where we defined the Cartesian velocity vector

$$\mathbf{V}_{\text{Cartesian}} \equiv \frac{dx}{dt} \hat{\mathbf{x}} + \frac{dy}{dt} \hat{\mathbf{y}} + \frac{dz}{dt} \hat{\mathbf{z}}, \quad (10.61)$$

which is the velocity as measured in the rotating reference frame when using planetary Cartesian coordinates. We also made use of equation (10.34) to express the time rate of change for the planetary Cartesian unit vectors, with this change arising solely from the planetary rotation.

10.9.2 Planetary spherical coordinate representation

The position vector in the planetary spherical coordinate representation is given by

$$\mathbf{X} = r \hat{\mathbf{r}}. \quad (10.62)$$

The basis vector $\hat{\mathbf{r}}$ is normalized, so that its evolution arises just from rotations. It can rotate either through solid-body motion of the rotating reference frame, or through changes in the spherical angles, λ, ϕ relative to the rotating reference frame. We see these two forms of time changes by taking the time derivative of $\hat{\mathbf{r}}$ as given by equation (8.30c)

$$\frac{d\hat{\mathbf{r}}}{dt} = \frac{d}{dt} [\hat{\mathbf{x}} \cos \lambda \cos \phi + \hat{\mathbf{y}} \sin \lambda \cos \phi + \hat{\mathbf{z}} \sin \phi]. \quad (10.63)$$

Expanding the right hand side leads to

$$\frac{d}{dt} [\hat{\mathbf{x}} \cos \lambda \cos \phi] = \frac{d\hat{\mathbf{x}}}{dt} \cos \lambda \cos \phi - \hat{\mathbf{x}} \dot{\lambda} \sin \lambda \cos \phi - \hat{\mathbf{x}} \dot{\phi} \cos \lambda \sin \phi \quad (10.64a)$$

$$\frac{d}{dt} [\hat{\mathbf{y}} \sin \lambda \cos \phi] = \frac{d\hat{\mathbf{y}}}{dt} \sin \lambda \cos \phi + \hat{\mathbf{y}} \dot{\lambda} \cos \lambda \cos \phi - \hat{\mathbf{y}} \dot{\phi} \sin \lambda \sin \phi \quad (10.64b)$$

$$\frac{d}{dt} [\hat{\mathbf{z}} \sin \phi] = \hat{\mathbf{z}} \dot{\phi} \cos \phi. \quad (10.64c)$$

Making use of equation (10.34) for the change in planetary Cartesian unit vectors due to rotation, and substituting the expressions (8.30a)-(8.30c) for the spherical unit vectors, leads to²

$$\frac{d\hat{\mathbf{r}}}{dt} = \cos \phi \left[\frac{d\lambda}{dt} + \Omega \right] \hat{\mathbf{\lambda}} + \frac{d\phi}{dt} \hat{\mathbf{\phi}}. \quad (10.65)$$

Consequently, the inertial velocity has the following spherical coordinate representation

$$\mathbf{V} = \frac{d\mathbf{X}}{dt} \quad (10.66a)$$

$$= \frac{d(r \hat{\mathbf{r}})}{dt} \quad (10.66b)$$

$$= \frac{dr}{dt} \hat{\mathbf{r}} + r \frac{d\hat{\mathbf{r}}}{dt} \quad (10.66c)$$

$$= r_{\perp} \left[\frac{d\lambda}{dt} + \Omega \right] \hat{\mathbf{\lambda}} + r \frac{d\phi}{dt} \hat{\mathbf{\phi}} + \frac{dr}{dt} \hat{\mathbf{r}} \quad (10.66d)$$

$$= (u + r_{\perp} \Omega) \hat{\mathbf{\lambda}} + v \hat{\mathbf{\phi}} + w \hat{\mathbf{r}} \quad (10.66e)$$

$$= \mathbf{V}_{\text{spherical}} + \mathbf{U}_{\text{solid}}. \quad (10.66f)$$

In this equation we introduced the spherical coordinate velocity vector

$$\mathbf{V}_{\text{spherical}} = r_{\perp} \frac{d\lambda}{dt} \hat{\mathbf{\lambda}} + r \frac{d\phi}{dt} \hat{\mathbf{\phi}} + \frac{dr}{dt} \hat{\mathbf{r}} = u \hat{\mathbf{\lambda}} + v \hat{\mathbf{\phi}} + w \hat{\mathbf{r}}, \quad (10.67)$$

where

$$u = r_{\perp} \frac{d\lambda}{dt} \quad \text{and} \quad v = r \frac{d\phi}{dt} \quad \text{and} \quad w = \frac{dr}{dt}, \quad (10.68)$$

are components to the spherical velocity vector, and with

$$r_{\perp} = r \cos \phi \quad (10.69)$$

²The reader is encouraged to work through the details to derive equation (10.65).

the distance to the polar axis. The spherical velocity, $\mathbf{V}_{\text{spherical}}$, is the velocity measured in the rotating reference frame when using planetary spherical coordinates. We also noted that the solid-body velocity has the spherical coordinate representation

$$\mathbf{U}_{\text{solid}} = \boldsymbol{\Omega} \wedge \mathbf{X} = r_{\perp} \Omega \hat{\lambda}. \quad (10.70)$$

That is, the solid-body velocity is purely zonal.

10.9.3 Transforming from Cartesian to spherical

We can make use of the tensor algebra from Section 10.5 to transform from the Cartesian representation of the velocity vector to the spherical representation. This approach leads to an equivalent result to that pursued thus far in this section, but it is somewhat more telescopic. In particular, we make use of the transformation rule (10.45) along with the transformation matrix (10.46) and its inverse (10.47) to have

$$V^{\bar{1}} = V^r = \Lambda^{\bar{1}}_1 V^1 + \Lambda^{\bar{1}}_2 V^2 + \Lambda^{\bar{1}}_3 V^3 \quad (10.71a)$$

$$V^{\bar{2}} = V^{\lambda} = \Lambda^{\bar{2}}_1 V^1 + \Lambda^{\bar{2}}_2 V^2 + \Lambda^{\bar{2}}_3 V^3 \quad (10.71b)$$

$$V^{\bar{3}} = V^{\phi} = \Lambda^{\bar{3}}_1 V^1 + \Lambda^{\bar{3}}_2 V^2 + \Lambda^{\bar{3}}_3 V^3, \quad (10.71c)$$

where the Cartesian components are

$$V^1 = \dot{x} - \Omega y = \dot{r} \cos \phi \cos \lambda - r \dot{\phi} \sin \phi \cos \lambda - r (\dot{\lambda} + \Omega) \cos \phi \sin \lambda \quad (10.72a)$$

$$V^2 = \dot{y} + \Omega x = \dot{r} \cos \phi \sin \lambda - r \dot{\phi} \sin \phi \sin \lambda + r (\dot{\lambda} + \Omega) \cos \phi \cos \lambda \quad (10.72b)$$

$$V^3 = \dot{z} = \dot{r} \sin \phi + r \dot{\phi} \cos \phi. \quad (10.72c)$$

Making use of the inverse transformation matrix components $\Lambda^{\bar{a}}_b$ given by equation (10.47), as well as the relation (10.45) between the coordinate basis vectors, leads to

$$\vec{V} = \mathbf{V} = V^{\bar{a}} \vec{e}_{\bar{a}} = r_{\perp} (\dot{\lambda} + \Omega) \hat{\lambda} + r \dot{\phi} \hat{\phi} + \dot{r} \hat{r}, \quad (10.73)$$

which is the same expression (10.66f) determined without the formalism of tensor algebra.

10.9.4 Axial angular momentum

As seen in Section 11.7, the zonal component of the inertial velocity equals to the axial angular momentum per unit mass

$$L^z = m \hat{\lambda} \cdot \mathbf{V} = m (u + r_{\perp} \Omega). \quad (10.74)$$

The distance to the rotational axis is given by r_{\perp} , and this is the *moment arm* for the axial angular momentum. For cases with rotational symmetry around polar axis, as for motion of a particle around a smooth sphere, the axial angular momentum is a constant of the motion. As discussed in Section 11.7, this conservation law offers a very important constraint on the particle trajectory and it plays a role in the motion of geophysical fluids (Section 18.4).

10.10 Planetary Cartesian representation of acceleration

The inertial acceleration vector is given by the second time derivative of the position vector

$$\mathbf{A} = \frac{d\mathbf{V}}{dt} = \frac{d^2\mathbf{X}}{dt^2}. \quad (10.75)$$

We here consider its representation using planetary Cartesian coordinates (x, y, z) and the Cartesian basis $(\vec{e}_1, \vec{e}_2, \vec{e}_3) = (\hat{x}, \hat{y}, \hat{z})$.

10.10.1 Planetary Cartesian representation

For our study of geophysical fluid motion, we assume the planetary angular velocity, $\boldsymbol{\Omega}$, is a constant in time

$$\frac{d\boldsymbol{\Omega}}{dt} = 0. \quad (10.76)$$

Making use of the results from Section 10.4 leads to

$$\frac{d\vec{e}_a}{dt} = \boldsymbol{\Omega} \wedge \vec{e}_a \quad (10.77)$$

and

$$\frac{d^2\vec{e}_a}{dt^2} = \frac{d}{dt}(\boldsymbol{\Omega} \wedge \vec{e}_a) = \boldsymbol{\Omega} \wedge \frac{d\vec{e}_a}{dt} = \boldsymbol{\Omega} \wedge (\boldsymbol{\Omega} \wedge \vec{e}_a), \quad (10.78)$$

which yields the inertial acceleration

$$\mathbf{A} = \frac{d}{dt} \frac{d\mathbf{X}}{dt} \quad (10.79a)$$

$$= \frac{d^2\xi^a}{dt^2} \vec{e}_a + 2 \frac{d\xi^a}{dt} (\boldsymbol{\Omega} \wedge \vec{e}_a) + \xi^a \boldsymbol{\Omega} \wedge (\boldsymbol{\Omega} \wedge \vec{e}_a) \quad (10.79b)$$

$$= \hat{\mathbf{x}} [\ddot{x} - 2\Omega \dot{y} - \Omega^2 x] + \hat{\mathbf{y}} [\ddot{y} + 2\Omega \dot{x} - \Omega^2 y] + \hat{\mathbf{z}} \ddot{z} \quad (10.79c)$$

$$= \ddot{x} \hat{\mathbf{x}} + \ddot{y} \hat{\mathbf{y}} + \ddot{z} \hat{\mathbf{z}} + 2\Omega (-\dot{y} \hat{\mathbf{x}} + \dot{x} \hat{\mathbf{y}}) - \Omega^2 (x \hat{\mathbf{x}} + y \hat{\mathbf{y}}) \quad (10.79d)$$

$$= \mathbf{A}_{\text{Cartesian}} + 2\boldsymbol{\Omega} \wedge \mathbf{V}_{\text{Cartesian}} + \boldsymbol{\Omega} \wedge (\boldsymbol{\Omega} \wedge \mathbf{X}) \quad (10.79e)$$

$$= \mathbf{A}_{\text{Cartesian}} - \mathbf{A}_{\text{Coriolis}} - \mathbf{A}_{\text{centrifugal}}. \quad (10.79f)$$

The inertial acceleration is thus decomposed into three terms. The first contribution is the Cartesian expression

$$\mathbf{A}_{\text{Cartesian}} = \frac{d^2x}{dt^2} \hat{\mathbf{x}} + \frac{d^2y}{dt^2} \hat{\mathbf{y}} + \frac{d^2z}{dt^2} \hat{\mathbf{z}} = \ddot{x} \hat{\mathbf{x}} + \ddot{y} \hat{\mathbf{y}} + \ddot{z} \hat{\mathbf{z}}, \quad (10.80)$$

which is the coordinate acceleration measured in the rotating frame using planetary Cartesian coordinates. The second contribution on the right hand side of equation (10.79f) is minus the Coriolis acceleration

$$\mathbf{A}_{\text{Coriolis}} = -2\boldsymbol{\Omega} \wedge \mathbf{V}_{\text{Cartesian}}. \quad (10.81)$$

The Coriolis acceleration plays a fundamental role in geophysical fluid mechanics and will be central to our development in this book. One key feature of the Coriolis acceleration is that it vanishes when there is no motion relative to the rotating reference frame. The third contribution is the *centripetal acceleration*, which is also minus the *centrifugal acceleration*

$$\mathbf{A}_{\text{centripetal}} = -\mathbf{A}_{\text{centrifugal}} = \boldsymbol{\Omega} \wedge (\boldsymbol{\Omega} \wedge \mathbf{X}) = -\Omega^2 \mathbf{X}. \quad (10.82)$$

The centrifugal acceleration points outward from (perpendicular to) the polar axis of rotation whereas the centripetal acceleration points inward; they are action/reaction pairs. They can be written as the gradient of a potential

$$\mathbf{A}_{\text{centrifugal}} = -\nabla \Phi_{\text{centrifugal}} \quad (10.83)$$

with

$$-\Phi_{\text{centrifugal}} \equiv \frac{(\boldsymbol{\Omega} \wedge \mathbf{x}) \cdot (\boldsymbol{\Omega} \wedge \mathbf{x})}{2} = \frac{\Omega^2 r_\perp^2}{2} = \frac{(\Omega r \cos \phi)^2}{2} = \frac{\Omega^2 (x^2 + y^2)}{2}. \quad (10.84)$$

The centripetal acceleration (pointing towards the rotational axis) is that part of the inertial acceleration that keeps the rotating particle from flying away from the rotational axis. On a massive rotating sphere, the centripetal acceleration is provided by that component of the gravitational force (Section 10.12) directed towards the rotation axis. The centripetal acceleration's opposing partner, the centrifugal acceleration, is a non-inertial acceleration that accounts for the slight equatorial bulge of the planet. It is what pulls one outward from the center of a rotating merry-go-round.

10.10.2 Summary of acceleration representation

For the purpose of formulating the equation of motion in the rotating terrestrial frame, we write the rotating frame acceleration as

$$\mathbf{A}_{\text{Cartesian}} = \mathbf{A} + \mathbf{A}_{\text{Coriolis}} + \mathbf{A}_{\text{centrifugal}} \quad (10.85a)$$

$$= \mathbf{A} - 2\boldsymbol{\Omega} \wedge \mathbf{V}_{\text{Cartesian}} - \nabla\Phi_{\text{centrifugal}} \quad (10.85b)$$

and summarize the following accelerations (force per unit mass).

- **INERTIAL:** Newton's law of motion is formulated within an inertial reference frame. It is the inertial acceleration, \mathbf{A} , that is directly affected by forces such as gravitation.
- **CORIOLIS:** The Coriolis acceleration

$$\mathbf{A}_{\text{Coriolis}} = -2\boldsymbol{\Omega} \wedge \mathbf{V}_{\text{Cartesian}} = -2\Omega \hat{\mathbf{z}} \wedge \mathbf{V}_{\text{Cartesian}} = -2\Omega \left[-\frac{dy}{dt} \hat{\mathbf{x}} + \frac{dx}{dt} \hat{\mathbf{y}} \right], \quad (10.86)$$

arises from our choice to describe motion within the rotating reference frame. The Coriolis acceleration gives rise to a rich suite of fundamentally new phenomena relative to non-rotating motion. It has components only in the horizontal planetary Cartesian plane spanned by the unit vectors $(\hat{\mathbf{x}}, \hat{\mathbf{y}})$. That is, the Coriolis acceleration occurs in a plane parallel to the equatorial plane. This geometric results is to be expected since the Coriolis acceleration arises from rotation about the polar $\hat{\mathbf{z}}$ axis.

- **CENTRIFUGAL:** The centrifugal acceleration

$$\mathbf{A}_{\text{centrifugal}} = -\nabla\Phi_{\text{centrifugal}} = \Omega^2 \mathbf{r}_\perp = \Omega^2 (x \hat{\mathbf{x}} + y \hat{\mathbf{y}}) \quad (10.87)$$

is another term arising from the rotating reference frame. As for the Coriolis acceleration, the centrifugal acceleration has components only in the horizontal planetary Cartesian plane spanned by the unit vectors $(\hat{\mathbf{x}}, \hat{\mathbf{y}})$; i.e., the centrifugal acceleration occurs in a plane parallel to the equatorial plane, which is again expected since the centrifugal acceleration arises from rotation about the polar axis. The centrifugal acceleration is directed outward from (perpendicular to) the polar axis of rotation. We see this orientation in Figure 10.6 to be discussed later. Furthermore, the centrifugal acceleration is nonzero even when the particle is fixed relative to the rotating planet, whereas the Coriolis acceleration is zero when the particle has zero motion relative to the planet.

The centrifugal acceleration can be written as the gradient of a scalar potential, $\mathbf{A}_{\text{centrifugal}} = -\nabla\Phi_{\text{centrifugal}}$ where $\Phi_{\text{centrifugal}} = -\Omega^2(x^2 + y^2)/2$ (equation (10.84)). Hence, the centrifugal acceleration can be combined with the gravitational acceleration in the equation of motion (see Section 10.12). The resulting “effective gravity” leads to a conservative force field that is modified relative to the central gravitational field of the non-rotating spherical planet. We detail these points in Section 10.12.4.

10.10.3 Further study

Section 3.5 of [Apel \(1987\)](#) offers an insightful presentation of the Coriolis acceleration. Visualizations from rotating tank experiments are useful to garner an intuitive understanding of the Coriolis acceleration. The first few minutes of [this video from Prof. Dave Fultz of the University of Chicago](#) is particularly insightful. We will further build up our understanding of the Coriolis acceleration as the book develops.

10.11 Spherical representation of acceleration

The spherical representation of the inertial velocity is given by equation (10.66f)

$$\mathbf{V} = \frac{d\mathbf{X}}{dt} = (u + r_{\perp} \Omega) \hat{\lambda} + v \hat{\phi} + w \hat{r} = \mathbf{V}_{\text{sphere}} + r_{\perp} \Omega \hat{\lambda}, \quad (10.88)$$

where we introduced the spherical velocity from equation (10.68)

$$\mathbf{V}_{\text{sphere}} \equiv u \hat{\lambda} + v \hat{\phi} + w \hat{r}. \quad (10.89)$$

We will also make use of the notation for the zonal component of the inertial velocity,

$$u_I = u + r_{\perp} \Omega. \quad (10.90)$$

Just as for computing the inertial velocity vector, the inertial acceleration must take into account changes in both the spherical coordinates and spherical basis vectors

$$\mathbf{A} = \frac{d}{dt} (u_I \hat{\lambda} + v \hat{\phi} + w \hat{r}) \quad (10.91a)$$

$$= \frac{du_I}{dt} \hat{\lambda} + \frac{dv}{dt} \hat{\phi} + \frac{dw}{dt} \hat{r} + u_I \frac{d\hat{\lambda}}{dt} + v \frac{d\hat{\phi}}{dt} + w \frac{d\hat{r}}{dt}. \quad (10.91b)$$

The spherical unit vectors change due to both the solid-body rotation of the rotating reference frame, plus motion of the particle relative to the rotating frame. Making use of the expressions given in Section (8.2.2), a bit of algebra yields the time derivatives

$$\frac{d\hat{\lambda}}{dt} = \left[\Omega + \frac{d\lambda}{dt} \right] (\hat{\phi} \sin \phi - \hat{r} \cos \phi) \quad (10.92)$$

$$\frac{d\hat{\phi}}{dt} = -\hat{\lambda} \left[\Omega + \frac{d\lambda}{dt} \right] \sin \phi - \hat{r} \dot{\phi} \quad (10.93)$$

$$\frac{d\hat{r}}{dt} = \hat{\lambda} \left[\frac{d\lambda}{dt} + \Omega \right] \cos \phi + \frac{d\phi}{dt} \hat{\phi}. \quad (10.94)$$

We are thus led to the inertial acceleration components

$$\mathbf{A} \cdot \hat{\lambda} = \frac{du_I}{dt} + \left[\frac{d\lambda}{dt} + \Omega \right] (w \cos \phi - v \sin \phi) \quad (10.95a)$$

$$\mathbf{A} \cdot \hat{\phi} = \frac{dv}{dt} + \left[\frac{d\lambda}{dt} + \Omega \right] u_I \sin \phi + w \frac{d\phi}{dt} \quad (10.95b)$$

$$\mathbf{A} \cdot \hat{r} = \frac{dw}{dt} - \left[\frac{d\lambda}{dt} + \Omega \right] u_I \cos \phi - v \frac{d\phi}{dt}. \quad (10.95c)$$

Use of the identities

$$u = r_{\perp} \frac{d\lambda}{dt} \quad \text{and} \quad u_I = u + r_{\perp} \Omega \quad \text{and} \quad \frac{du_I}{dt} = \frac{du}{dt} + \Omega (w \cos \phi - v \sin \phi) \quad (10.96)$$

and some reorganization leads to the spherical coordinate representation of the inertial acceleration

$$\begin{aligned} \mathbf{A} = & \hat{\lambda} \left[\frac{du}{dt} + \frac{u(w - v \tan \phi)}{r} + 2\Omega(w \cos \phi - v \sin \phi) \right] \\ & + \hat{\phi} \left[\frac{dv}{dt} + \frac{vw + u^2 \tan \phi}{r} + 2\Omega u \sin \phi + r_{\perp} \Omega^2 \sin \phi \right] \\ & + \hat{r} \left[\frac{dw}{dt} - \frac{u^2 + v^2}{r} - 2\Omega u \cos \phi - r_{\perp} \Omega^2 \cos \phi \right]. \end{aligned} \quad (10.97)$$

When encountering this expression for the first time it can be rather intimidating. The reader is encouraged to remember the underlying kinematic origin of the various terms, in which case the expression can be seen as both familiar and aesthetically pleasing.

10.11.1 Transforming from Cartesian to spherical

As in Section 10.9.3, we can make use of the tensor algebra from Section 10.5 to transform from the Cartesian representation of the acceleration vector to the spherical representation. Following the same steps as for the velocity leads to

$$A^{\bar{a}} = \Lambda^{\bar{a}}_a A^a \quad (10.98)$$

where the Cartesian components to the inertial acceleration are

$$A^1 = \ddot{x} - 2\Omega \dot{y} - \Omega^2 x \quad (10.99a)$$

$$A^2 = \ddot{y} + 2\Omega \dot{x} - \Omega^2 y \quad (10.99b)$$

$$A^3 = \ddot{z}. \quad (10.99c)$$

Making use of the coordinate transformation (10.44) allows us to express these Cartesian components of the acceleration in terms of spherical coordinates. Then we make use of the inverse transformation matrix components $\Lambda^{\bar{a}}_b$ given by equation (10.47), as well as the relation (10.45) between the coordinate basis vectors, which leads to

$$\vec{A} = \mathbf{A} = A^a \vec{e}_a = A^{\bar{a}} \vec{e}_{\bar{a}} \quad (10.100)$$

with the spherical coordinate representation given by equation (10.97) as derived without the formalism of tensor algebra. Both approaches require algebraic manipulations, so that it is useful to have two approaches to double-check results.

10.11.2 Decomposing the acceleration

We decompose the inertial acceleration (10.97) into the following terms

$$\mathbf{A} = \mathbf{A}_{\text{sphere}} + \mathbf{A}_{\text{metric}} - \mathbf{A}_{\text{Coriolis}} - \mathbf{A}_{\text{centrifugal}}, \quad (10.101)$$

with signs chosen so that in the rotating frame the acceleration is written

$$\underbrace{\mathbf{A}_{\text{sphere}} + \mathbf{A}_{\text{metric}}}_{\text{net spherical acceleration}} = \mathbf{A} + \mathbf{A}_{\text{Coriolis}} + \mathbf{A}_{\text{centrifugal}}. \quad (10.102)$$

We identify the net spherical acceleration as the sum of the coordinate acceleration and metric acceleration. In the absence of rotation, this sum provides an expression for the inertial acceleration as represented by spherical coordinates. The Coriolis and centrifugal terms arise from rotation.

10.11.3 Spherical coordinate acceleration

The spherical coordinate acceleration is given by the time change in the spherical velocity components

$$\mathbf{A}_{\text{sphere}} = \frac{du}{dt} \hat{\lambda} + \frac{dv}{dt} \hat{\phi} + \frac{dw}{dt} \hat{r}. \quad (10.103)$$

This term has no contribution from changes to the spherical unit vectors.

10.11.4 Metric acceleration

We define the metric acceleration as that contribution to the acceleration arising from the time dependence of the spherical unit vectors that appears when taking the time derivative of the velocity vector. For spherical coordinates we have

$$\mathbf{A}_{\text{metric}} = \hat{\lambda} \left[\frac{u(w - v \tan \phi)}{r} \right] + \hat{\phi} \left[\frac{v w + u^2 \tan \phi}{r} \right] - \hat{r} \left[\frac{u^2 + v^2}{r} \right] \quad (10.104a)$$

$$= \hat{\lambda} \left[\frac{u(w \cos \phi - v \sin \phi)}{r \cos \phi} \right] + \hat{\phi} \left[\frac{v w \cos \phi + u^2 \sin \phi}{r \cos \phi} \right] - \hat{r} \left[\frac{u^2 + v^2}{r} \right] \quad (10.104b)$$

$$= \frac{1}{r} [u \tan \phi (\hat{r} \wedge \mathbf{V}_{\text{sphere}}) + w \mathbf{U}_{\text{sphere}} - \hat{r} \mathbf{U}_{\text{sphere}} \cdot \mathbf{U}_{\text{sphere}}], \quad (10.104c)$$

where we wrote the horizontal (angular) and vertical (radial) components of the spherical velocity according to

$$\mathbf{V}_{\text{sphere}} = \mathbf{U}_{\text{sphere}} + \hat{r} w = \hat{\lambda} u + \hat{\phi} v + \hat{r} w. \quad (10.105)$$

Note that

$$\mathbf{V}_{\text{sphere}} \cdot \mathbf{A}_{\text{metric}} = 0, \quad (10.106)$$

so that the metric acceleration is orthogonal to the spherical velocity, in which case with (see equation (10.66f)) $\mathbf{V} = \mathbf{V}_{\text{sphere}} + r_{\perp} \Omega \hat{\lambda}$ we have

$$\mathbf{V} \cdot \mathbf{A}_{\text{metric}} = \Omega u (-v \sin \phi + w \cos \phi). \quad (10.107)$$

Furthermore, the metric acceleration vanishes when the curvature of the sphere vanishes (i.e., $r \rightarrow \infty$), as per a flat plane.

10.11.5 Centrifugal acceleration

The spherical coordinate representation of the centrifugal acceleration is given by

$$\mathbf{A}_{\text{centrifugal}} = -\nabla \Phi_{\text{centrifugal}} = \Omega^2 (x \hat{x} + y \hat{y}) = r_{\perp} \Omega^2 (-\hat{\phi} \sin \phi + \hat{r} \cos \phi). \quad (10.108)$$

The centrifugal acceleration points outward from the axis of rotation (see Figure 10.6 to be discussed later), so that it has no component in the longitudinal direction. Furthermore, note that

$$\mathbf{V} \cdot \mathbf{A}_{\text{centrifugal}} = \mathbf{V}_{\text{sphere}} \cdot \mathbf{A}_{\text{centrifugal}} = r_{\perp} \Omega^2 (-v \sin \phi + w \cos \phi). \quad (10.109)$$

10.11.6 Coriolis acceleration

The spherical coordinate representation of the Coriolis acceleration makes use of the spherical representation of the earth's angular velocity

$$\boldsymbol{\Omega} = \Omega \hat{\mathbf{z}} = \Omega (\hat{\phi} \cos \phi + \hat{\mathbf{r}} \sin \phi) \quad (10.110)$$

to reach the form

$$\mathbf{A}_{\text{Coriolis}} = -2 \boldsymbol{\Omega} \wedge \mathbf{V}_{\text{sphere}} \quad (10.111a)$$

$$= -2 \Omega (\hat{\phi} \cos \phi + \hat{\mathbf{r}} \sin \phi) \wedge \mathbf{V}_{\text{sphere}} \quad (10.111b)$$

$$= -2 \Omega (\hat{\phi} \cos \phi + \hat{\mathbf{r}} \sin \phi) \wedge (u \hat{\lambda} + v \hat{\phi} + w \hat{\mathbf{r}}) \quad (10.111c)$$

$$= -2 \Omega [\hat{\lambda} (w \cos \phi - v \sin \phi) + \hat{\phi} u \sin \phi - \hat{\mathbf{r}} u \cos \phi]. \quad (10.111d)$$

As for the metric acceleration, we note that the Coriolis acceleration is orthogonal to the spherical velocity

$$\mathbf{V}_{\text{sphere}} \cdot \mathbf{A}_{\text{Coriolis}} = 0, \quad (10.112)$$

so that

$$\mathbf{V} \cdot \mathbf{A}_{\text{Coriolis}} = 2 r_{\perp} \Omega^2 (v \sin \phi - w \cos \phi). \quad (10.113)$$

We find it convenient to introduce a shorthand notation

$$\mathbf{f} = (2 \Omega \sin \phi) \hat{\mathbf{r}} \quad \text{and} \quad \mathbf{f}^* = (2 \Omega \cos \phi) \hat{\phi}, \quad (10.114)$$

so that the Coriolis acceleration takes the form

$$\mathbf{A}_{\text{Coriolis}} = -(\mathbf{f} + \mathbf{f}^*) \wedge \mathbf{V}_{\text{sphere}}. \quad (10.115)$$

There are two contributions to the Coriolis acceleration: one from the radial and one from the meridional component of the earth's rotation vector.

10.11.7 Coriolis acceleration for large-scale motions

Let us again write the Coriolis acceleration in equation (10.111d), only now underlining two terms

$$\mathbf{A}_{\text{Coriolis}} = -2 \Omega [\hat{\lambda} (\underline{w \cos \phi} - v \sin \phi) + \hat{\phi} \underline{u \sin \phi} - \hat{\mathbf{r}} \underline{u \cos \phi}]. \quad (10.116)$$

For many applications in geophysical fluid dynamics, the term $\hat{\mathbf{r}} (2 \Omega u \cos \phi)$ is much smaller than the competing gravitational acceleration that also contributes to the radial acceleration, thus prompting $\hat{\mathbf{r}} (2 \Omega u \cos \phi)$ to be dropped from the $\hat{\mathbf{r}}$ equation of motion.³ Furthermore, the vertical velocity term is generally much smaller than the horizontal velocity term appearing in the $\hat{\lambda}$ component. Dropping these two terms results in the form for the Coriolis acceleration used for large-scale dynamics, such as when considering the hydrostatic primitive equations for geophysical fluids (Section 23.1)

$$\mathbf{A}_{\text{Coriolis}}^{\text{large-scale}} \equiv -2 \Omega \sin \phi (-\hat{\lambda} v + \hat{\phi} u) \equiv -f \hat{\mathbf{r}} \wedge \mathbf{V}_{\text{sphere}}. \quad (10.117)$$

For the last equality we introduced the Coriolis parameter

$$f \equiv 2 \Omega \sin \phi. \quad (10.118)$$

³The term $\hat{\mathbf{r}} (2 \Omega u \cos \phi)$ is called the Eötvös correction in the study of marine gravity.

As illustrated in Figure 10.5, we see that it is the radial (i.e., the local vertical) component of the earth's angular rotation that plays the most important role in large-scale geophysical fluid mechanics

$$\boldsymbol{\Omega} \cdot \hat{\mathbf{r}} = \Omega \hat{\mathbf{z}} \cdot \hat{\mathbf{r}} = [\Omega (\hat{\phi} \cos \phi + \hat{\mathbf{r}} \sin \phi)] \cdot \hat{\mathbf{r}} = \Omega \sin \phi \hat{\mathbf{r}} = (f/2) \hat{\mathbf{r}}. \quad (10.119)$$

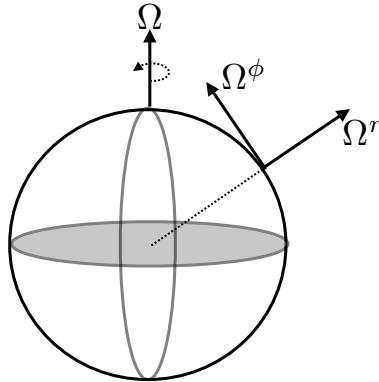


FIGURE 10.5: This figure illustrates the two components of the earth's rotational velocity, $\boldsymbol{\Omega} = \Omega \hat{\mathbf{z}} = \Omega (\hat{\phi} \cos \phi + \hat{\mathbf{r}} \sin \phi)$. The radial component (also known as the local vertical component), $\boldsymbol{\Omega} \approx \Omega \sin \phi \hat{\mathbf{r}}$, is the most important component for large scale geophysical fluid dynamics.

10.12 Newtonian gravity

Thus far we have focused on kinematics of a particle moving around a rotating sphere. We now acknowledge that the particle is moving in the gravitational field of the planet. The gravitational force acting on the particle is the only inertial force felt by the point particle. Since the point particle contains no internal structure and it has no surface area, the total energy for the particle equals to the mechanical energy (Chapter 11). We here discuss the gravitational potential energy and the associated gravitational force, all within the context of Newtonian mechanics.

10.12.1 Newtonian gravity from Poisson's equation

The Newtonian gravitational potential, Φ , in the presence of a mass distribution with density, ρ , satisfies Poisson's equation (see Section 3.5)

$$\nabla^2 \Phi = 4\pi G \rho, \quad (10.120)$$

where

$$G = 6.674 \times 10^{-11} \text{ N m}^2 \text{ kg}^{-2} = 6.674 \times 10^{-11} \text{ m}^3 \text{ kg}^{-1} \text{ s}^{-2} \quad (10.121)$$

is Newton's gravitational constant. Gravity is a conservative force, so that the gradient of the gravitational potential gives the gravitational acceleration

$$\mathbf{g} = -\nabla \Phi. \quad (10.122)$$

The gravitational potential has dimensions of $\text{L}^2 \text{T}^{-2}$ so that its gradient indeed has the dimensions of acceleration, L T^{-2} . Note that the gravitational potential responds instantaneously to any changes in the mass distribution. It took a few centuries, and the genius of Einstein, to provide a local connection between mass and gravity as rendered by space-time curvature and gravitational waves.

A volume integral of Poisson's equation (10.120) over a region in space leads to

$$\int_{\mathcal{R}} \nabla^2 \Phi \, dV = 4\pi G \int_{\mathcal{R}} \rho \, dV. \quad (10.123)$$

The integral on the right hand side is the mass contained in the region,

$$M = \int_{\mathcal{R}} \rho \, dV, \quad (10.124)$$

whereas the divergence theorem (Section 2.7) allows us to write the left hand side as a surface integral

$$\int_{\partial\mathcal{R}} \nabla \Phi \cdot \hat{\mathbf{n}} \, d\mathcal{S} = 4\pi G M. \quad (10.125)$$

We now consider the special case of a spherical mass distribution as viewed from a point outside of the mass.

10.12.2 Gravitational field for a spherical earth

In this book we assume the mass density of the planet to be spherically symmetric, in which case the gravitational potential is a function only of the radial distance from the center of the mass distribution, $\Phi = \Phi(r)$. Letting the integration region \mathcal{R} be a sphere of radius $r > R_e$, where R_e is the radius of the planet, and making use of spherical coordinates from Section 8.2.8, brings equation (10.125) to

$$4\pi r^2 \frac{\partial \Phi}{\partial r} = 4\pi G M. \quad (10.126)$$

Integration leads to the gravitational potential for an arbitrary point outside the spherically symmetric mass distribution

$$\Phi_e = -\frac{G M}{r}, \quad (10.127)$$

where we set the integration constant to zero. We see that when sampling the gravity field at a radius equal to or larger than the spherical planet radius, the gravitational potential is identical to that of a point mass at the origin. The gradient of the gravitational potential (10.127) yields the gravitational acceleration

$$\mathbf{g}_e = -\nabla \Phi_e = -\frac{G M}{r^2} \hat{\mathbf{r}} \quad (10.128)$$

along with the gravitational force acting on a point particle of mass m

$$\mathbf{F}_{\text{gravity}} = m \mathbf{g}_e = -m \nabla \Phi_e. \quad (10.129)$$

Furthermore, the gravitational potential energy of the particle (dimensions $M L^2 T^{-2}$) is given by

$$P = m \Phi_e. \quad (10.130)$$

10.12.3 Approximate gravitational acceleration

For most applications of atmospheric and oceanic fluid dynamics, it is sufficient to assume the gravitational acceleration is constant and equal to its value at the earth's surface. This assumption holds so long as the radial position of the particle is a distance from the earth surface that is small

relative to the earth radius. We generally make this assumption throughout this book.⁴ In this case we can assume the earth's gravitational acceleration, g_e , is a constant so that

$$\mathbf{g}_e = -g_e \hat{\mathbf{r}}, \quad (10.131)$$

where

$$g_e = \frac{GM_e}{R_e^2} \approx 9.8 \text{ m s}^{-2}. \quad (10.132)$$

To reach this value, we assumed a sphere of mass equal to the earth mass

$$M_e = 5.977 \times 10^{24} \text{ kg}, \quad (10.133)$$

and radius

$$R_e = 6.371 \times 10^6 \text{ m} \quad (10.134)$$

determined so that the sphere has the same volume as the earth.

The corresponding gravitational potential for the particle is given by

$$\Phi_e = g_e r, \quad (10.135)$$

with the gravitational acceleration

$$\mathbf{g}_e = -\nabla\Phi_e = -g_e \hat{\mathbf{r}}, \quad (10.136)$$

and the gravitational potential energy

$$m\Phi_e = m g_e r. \quad (10.137)$$

We emphasize that the expression for the gravitational potential, (10.135), and potential energy, (10.137), are accurate only so long as the radial position of the particle is a distance from the earth surface that is small relative to the earth radius. Furthermore, note that the approximate gravitational potential (10.135) is positive whereas the unapproximated potential (10.127) is negative. However, the absolute zero of the potential has no physical significance. Instead, what is relevant is the change between two points in space, with both potentials increasing when moving away from the earth center.

10.12.4 Effective gravitational force from the geopotential

We can combine the potential for the planetary centrifugal acceleration as given by equation (10.84) with the gravitational potential (10.127), thus resulting in the *geopotential*

$$\Phi = r [g_e - \mathbf{U}_{\text{solid}}^2/(2r)]. \quad (10.138)$$

Again, this form for the geopotential is relevant for motion that is close enough to the earth surface that we can assume the earth's gravitational acceleration, g_e , is constant as discussed in Section 10.12.3. In Exercise 10.5 we examine the geopotential when this assumption is not made.

The contribution from the planetary centrifugal term in the geopotential (10.138) can be estimated by making use of terrestrial values, in which $R = R_e = 6.371 \times 10^6 \text{ m}$ (equation (10.134)),

⁴The assumption of constant gravitational field is not appropriate when considering details of oceanic or atmospheric tidal motions or when aiming for precise measures of sea level. We consider more general gravitational fields in Chapter 29.

and $\Omega_e = 7.292 \times 10^{-5} \text{ s}^{-1}$ (Section 10.2). The planetary centrifugal term is its largest at the equator, $\phi = 0$, where

$$\frac{U_{\text{solid}}^2}{2R_e} \approx 0.017 \text{ m s}^{-2}, \quad (10.139)$$

so that the ratio of the gravitational to centrifugal accelerations is (at most)

$$\frac{g_e}{U_{\text{solid}}^2/(2R_e)} = \frac{M_e G / R_e^2}{\Omega_e^2 R_e / 2} \approx 576. \quad (10.140)$$

The geopotential is thus dominated by the earth's gravitational potential. Even so, the centrifugal acceleration leads to a slight equatorial bulge on the earth. To account for this slight non-sphericity, geophysical fluid models generally interpret the radial direction \hat{r} as pointing parallel to $\nabla\Phi$ rather than parallel to $\nabla\Phi_e$. We have more to say on this topic in Section 10.13.3.

10.12.5 Further study

Newton's gravitational law is standard material from freshman physics. Some commonly used physical properties of the earth are summarized in Appendix Two of [Gill \(1982\)](#).

10.13 Newton's law of motion

As seen in Section 10.1, Newton's law of motion says that in an inertial reference frame, temporal changes to the linear momentum arise only from externally applied forces. With gravity the only inertial force acting on the constant mass particle, Newton's equation of motion says that

$$m \mathbf{A} = -m \nabla\Phi_e. \quad (10.141)$$

This is a relatively simple equation of motion. However, it becomes more complex, yet useful for practical purposes, when moving to the rotating reference frame of terrestrial observers and furthermore when represented using spherical coordinates.

10.13.1 Cartesian coordinate representation

The inertial acceleration using planetary Cartesian coordinates is given by equation (10.85b)

$$\mathbf{A} = \mathbf{A}_{\text{Cartesian}} - \mathbf{A}_{\text{Coriolis}} - \mathbf{A}_{\text{centrifugal}} \quad (10.142a)$$

$$= \mathbf{A}_{\text{Cartesian}} + 2\boldsymbol{\Omega} \wedge \mathbf{V}_{\text{Cartesian}} + \nabla\Phi_{\text{centrifugal}}, \quad (10.142b)$$

so that the rotating frame Cartesian equation of motion is given by

$$\mathbf{A}_{\text{Cartesian}} = -\nabla\Phi_e - 2\boldsymbol{\Omega} \wedge \mathbf{V}_{\text{Cartesian}} - \nabla\Phi_{\text{centrifugal}} \quad (10.143a)$$

$$= -2\boldsymbol{\Omega} \wedge \mathbf{V}_{\text{Cartesian}} - \nabla\Phi, \quad (10.143b)$$

where the geopotential is the sum of the gravitational and centrifugal potentials (equation (10.138))

$$\Phi = \Phi_e + \Phi_{\text{centrifugal}}. \quad (10.144)$$

We can write the equation of motion in component form by exposing indices (Cartesian tensors) and using a dot for time derivative

$$\ddot{X}_a + 2\epsilon_{abc}\boldsymbol{\Omega}_b \dot{X}_c = -\partial_a\Phi. \quad (10.145)$$

We can write this equation in the standard vector form within the rotating reference frame

$$\frac{d^2 \mathbf{X}}{dt^2} + 2\boldsymbol{\Omega} \wedge \dot{\mathbf{X}} = -\nabla\Phi. \quad (10.146)$$

Note that the basis vectors need not be time differentiated again since their change has already been taken care of when exposing the Coriolis and centrifugal accelerations. This equation of motion is the standard form that will recur for a fluid, with the addition of contact forces from pressure and friction (Chapter 18).

Since the rotation of the reference frame is assumed to be constant in time, the equation of motion (10.146) can be written

$$\frac{d\mathbf{M}}{dt} = \frac{d}{dt}(\mathbf{V}_{\text{Cartesian}} + 2\boldsymbol{\Omega} \wedge \mathbf{X}) = -\nabla\Phi, \quad (10.147)$$

where we introduced the *potential momentum* per mass

$$\mathbf{M} = \mathbf{V}_{\text{Cartesian}} + 2\boldsymbol{\Omega} \wedge \mathbf{X}. \quad (10.148)$$

We further discuss potential momentum in Section 11.4.

10.13.2 Spherical coordinate representation

We now make use of the acceleration written in planetary spherical coordinates as given in Section 10.11

$$\mathbf{A}_{\text{sphere}} + \mathbf{A}_{\text{metric}} = \mathbf{A}_{\text{Coriolis}} + \mathbf{A} + \mathbf{A}_{\text{centrifugal}} = -2\boldsymbol{\Omega} \wedge \mathbf{V}_{\text{sphere}} - \nabla\Phi. \quad (10.149)$$

The effective gravitational force is not a central force due to the contribution from the centrifugal acceleration. We see this fact more explicitly by using the equations in Section 10.11 to write the spherical equations of motion

$$\dot{u} + \frac{u(w - v \tan \phi)}{r} + 2\Omega(w \cos \phi - v \sin \phi) = 0 \quad (10.150)$$

$$\dot{v} + \frac{v w + u^2 \tan \phi}{r} + 2\Omega u \sin \phi = -r_{\perp} \Omega^2 \sin \phi \quad (10.151)$$

$$\dot{w} - \frac{u^2 + v^2}{r} - 2\Omega u \cos \phi = r_{\perp} \Omega^2 \cos \phi - g_e. \quad (10.152)$$

The Ω^2 term in both the meridional equation, (10.151), and the radial equation, (10.152), are the two components of the planetary centrifugal acceleration, $\mathbf{A}_{\text{centrifugal}} = r_{\perp} \Omega^2 (-\hat{\phi} \sin \phi + \hat{r} \cos \phi)$. The centrifugal acceleration is directed outward from the planetary axis of rotation, and it is balanced by an inward directed planetary centripetal acceleration provided by that portion of the gravitational acceleration directed oppositely to the centrifugal. Notably, a particle initially at rest on a smooth spherical planet will accelerate meridionally toward the equator due to the meridional component of the centrifugal acceleration (from equation (10.151) we find $\dot{v} = -r_{\perp} \Omega^2 \sin \phi$).

Imagine setting the centripetal acceleration to zero, in which case the particle would still feel the central force from gravity but its trajectory would differ. As seen in Section 10.12.4, the earth's gravitational acceleration is much larger than the planetary centrifugal acceleration, so that in the absence of the planetary centrifugal acceleration the particle would still be bound to the planet. But in more extreme conditions where the rotational rate is much higher (e.g., a rotating neutron star), removing the centripetal acceleration causes a huge modification to the particle trajectory. Indeed, with enough rotation, the particle can overcome gravitational effects from the planet to leave its orbit.

10.13.3 Geopotential coordinate representation

As we saw in Section 10.12.2, the radius of a sphere that best fits the volume of the earth is given by $R_e = 6.371 \times 10^6$ m. The non-central nature of the effective gravitational force (arising from central gravity plus planetary centrifugal) leads to an oblate spheroidal shape for planets such as the earth. The result is a distinction between the earth's equatorial and polar radii (Appendix Two of [Gill \(1982\)](#))

$$R_{\text{equator}} = 6.378 \times 10^6 \text{ m} \quad \text{and} \quad R_{\text{pole}} = 6.357 \times 10^6 \text{ m}, \quad (10.153)$$

with a corresponding ratio

$$1 - \frac{R_{\text{pole}}}{R_{\text{equator}}} \approx 3 \times 10^{-3}. \quad (10.154)$$

An oblate spheroid shape does a better job fitting the actual earth shape than a sphere, thus motivating the use of oblate spheroid coordinates for describing planetary scale mechanics. In this case, the radial coordinate is constant on the oblate spheroid shaped geopotential, and the effective gravitational acceleration is precisely aligned with the geopotential direction.

Even though oblate spheroidal coordinates are better than spherical for describing geopotentials, it is possible, to a high degree of accuracy, to describe the earth's geometry as spherical. Doing so simplifies the mathematics since oblate spheroidal coordinates are less convenient and less familiar than standard spherical coordinates. We are thus led to assume that the radial coordinate measures distances perpendicular to the geopotential, yet to use geometric/metric functions based on spherical coordinates. The error in this approach is small for the earth, and well worth the price since we no longer have a component to the effective gravitational force that is within the surface.

We illustrate the change in coordinates in Figure 10.6, with the figure caption also explaining how the force balances are reorganized. Absorbing the centrifugal term into an effective gravitational potential then leads to the effective gravitational acceleration vector

$$-\nabla\Phi = -g\hat{\mathbf{r}}, \quad (10.155)$$

with g the effective gravitational acceleration. Using this convention, the particle equations of motion take the following form

$$\dot{u} + \frac{u(w - v \tan \phi)}{r} + 2\Omega(w \cos \phi - v \sin \phi) = 0 \quad (10.156)$$

$$\dot{v} + \frac{v w + u^2 \tan \phi}{r} + 2\Omega u \sin \phi = 0 \quad (10.157)$$

$$\dot{w} - \frac{u^2 + v^2}{r} - 2\Omega u \cos \phi = -g. \quad (10.158)$$

Notably, in geopotential vertical coordinates the effective gravitational acceleration only impacts the radial equation of motion.

10.13.4 Comments

Figure 10.6, including its caption, offers a view on the transition from spherical coordinates to geopotential coordinates. The use of geopotential coordinates is rather accurate and extremely convenient for most purposes of geophysical fluid mechanics, with a notable exception being the study of tides and sea level, in which detailed models of the earth's mass distribution and gravity field are used (e.g., see [Gregory et al. \(2019\)](#) for a review). Although precise, the transition to

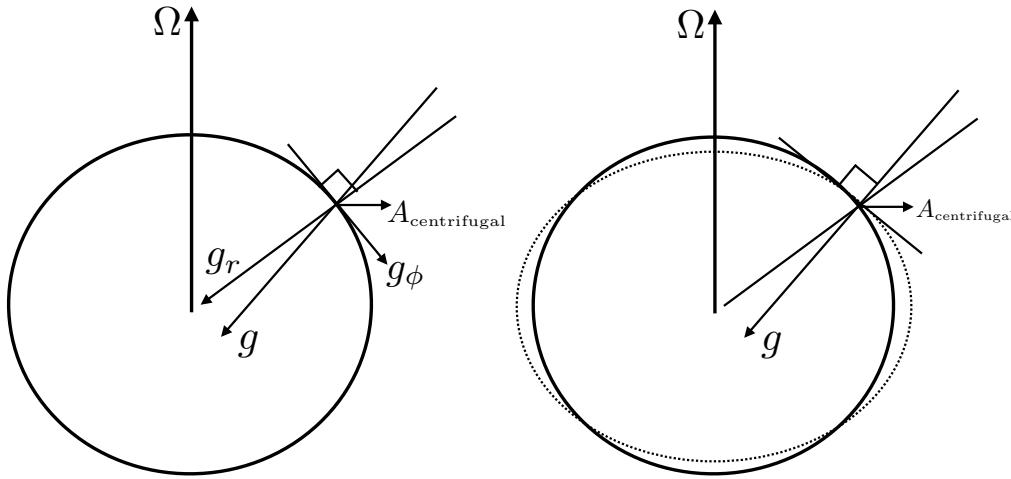


FIGURE 10.6: This figure illustrates the spherical (left panel) versus geopotential (right panel) coordinate systems used to study geophysical fluids. The left panel shows the non-central nature of the effective gravitational acceleration \mathbf{g} on a rotating spherical planet, with the effective gravity given by the sum of the gravitational acceleration, $\mathbf{g}_r = -g_e \hat{\mathbf{r}}$, plus centrifugal acceleration (equation (10.108)), $\mathbf{A}_{\text{centrifugal}} = r_{\perp} \Omega^2 (-\hat{\phi} \sin \phi + \hat{\mathbf{r}} \cos \phi)$. The gravitational acceleration points radially to the center of the sphere whereas the centrifugal acceleration points outward away from the polar axis of rotation. A particle initially at rest on a smooth spherical planet will accelerate meridionally toward the equator due to the meridional component of the centrifugal acceleration (from equation (10.151) we have $\dot{v} = -r_{\perp} \Omega^2 \sin \phi$). The right panel shows a *geopotential vertical coordinate*, $r = R + z$, that measures the distance perpendicular to the oblate spheroid shaped geopotential surface (dotted surface). The geopotential vertical coordinate precisely aligns the effective gravitational force with the vertical coordinate, so that there is no component of the effective gravity along the surface directions ($g_{\phi} = 0$). Equivalently, the central gravitational acceleration now has a meridional component on the oblate spheroid that exactly balances the meridional component to the centrifugal acceleration, leaving an effective gravity that is only vertical. Hence, a frictionless oblate spheroidal planet allows for a particle at rest on the planet's surface to remain at rest (see equation (10.157)). Note that this figure is not drawn to scale, with the oblate nature highly exaggerated compared to the real earth system (see equation (10.154)), and the centrifugal acceleration much smaller than the gravitational (see equation (10.140)).

geopotential coordinates is somewhat subtle in principle since we are reorganizing how the centrifugal acceleration appears. It is through this reorganization that we can largely ignore the centrifugal acceleration in our further studies of geophysical fluid mechanics.⁵

10.13.5 Further study

Section 4.12 of [Gill \(1982\)](#) and section 2.2.1 of [Vallis \(2017\)](#) present the terrestrial scaling needed to justify spherical coordinates with a radial effective gravitational potential. [Morse and Feshbach \(1953\)](#) and [Veronis \(1973\)](#) present details of spheroidal coordinates.

10.14 Exercises

EXERCISE 10.1: WORKING THROUGH THE SPHERICAL ACCELERATION

Convince yourself that the spherical form of the acceleration given by equation (10.97) is indeed correct.

EXERCISE 10.2: VELOCITY AND ACCELERATION IN CYLINDRICAL-POLAR COORDINATES

In Section 8.3 we worked through the transformation from Cartesian coordinates to cylindrical-

⁵An exception is in our study of rotating laboratory tank experiments in which it is more convenient to expose the centrifugal acceleration (Section 23.4).

polar coordinates for describing motion in a rotating reference frame. We also made use of polar coordinates in Section 10.4.4 to illustrate how polar unit vectors change under rotations. Here we develop an expression for the position, velocity, and acceleration vectors in a reference frame rotating with a constant rate about the vertical axis ($\boldsymbol{\Omega} = \Omega \hat{z}$) using rotating cylindrical-polar coordinates. The cylindrical-polar coordinates are useful when describing physical systems such as rotating fluid columns (e.g., fluids in a rotating circular tank as in Section 32.7) or when studying cyclostrophically balanced flow (Section 27.5).

- (A) Determine the representation of the inertial velocity vector, $\mathbf{V} = d\mathbf{X}/dt$, in terms of cylindrical-polar coordinates. Hint: do not forget to include the solid body motion of the rotating reference frame, which was discussed in Section 10.4.4.
- (B) Determine the representation of the inertial acceleration vector, $\mathbf{A} = d\mathbf{V}/dt$, in terms of cylindrical-polar coordinates.
- (C) Writing the inertial acceleration in the form

$$\mathbf{A} = \mathbf{A}_{\text{cylindrical-polar}} + \mathbf{A}_{\text{metric}} - \mathbf{A}_{\text{centrifugal}} - \mathbf{A}_{\text{Coriolis}}, \quad (10.159)$$

give the mathematical expressions for these terms:

- $\mathbf{A}_{\text{cylindrical-polar}}$ = acceleration in the rotating reference frame using cylindrical-polar coordinates;
- $\mathbf{A}_{\text{metric}}$ = acceleration due to motion of the cylindrical-polar unit vectors relative to the rotating reference frame;
- $\mathbf{A}_{\text{centrifugal}}$ = centrifugal acceleration;
- $\mathbf{A}_{\text{Coriolis}}$ = Coriolis acceleration.

EXERCISE 10.3: VELOCITY PROJECTED ONTO ACCELERATION

The kinetic energy per mass of a particle is given by

$$\mathcal{K} = \mathbf{V} \cdot \mathbf{V}/2, \quad (10.160)$$

where \mathbf{V} is the inertial velocity of a particle. In an inertial reference frame it is trivial to show that

$$\frac{d\mathcal{K}}{dt} = \mathbf{V} \cdot \mathbf{A} \quad (10.161)$$

through use of the chain rule, where $\mathbf{A} = d\mathbf{V}/dt$ is the inertial acceleration. Verify that this identity also holds in the rotating reference frame. For simplicity make use of planetary Cartesian coordinates.

EXERCISE 10.4: GEOMETRY OF CONSTANT GEOPOTENTIAL SURFACES

Here we examine some properties of the geopotential given by equation (10.138), where the squared solid-body speed is $\mathbf{U}_{\text{solid}}^2 = (\Omega r \cos \phi)^2$. We only consider geopotentials that are close to the radius of the planet, so that we can assume the gravitational acceleration, g_e , is constant and takes on its value at R_e as in Sections 10.12.2 and 10.12.4.

- (A) Sketch surfaces of constant geopotential according to the expression (10.138). Draw both meridional (north-south) lines and zonal (east-west) lines.

- (B) By equating the geopotential going around the pole to that going around the equator, show that the polar radius is less than the equatorial radius when $\Omega > 0$.
- (C) Taking the terrestrial values of g_e , R_{equator} , and Ω , what is the polar radius R_{pole} ? Compare to the measured value of the polar radius given by equation (10.153).

EXERCISE 10.5: GENERAL FORM OF THE GEOPOTENTIAL

In Exercise 10.4, as in Sections 10.12.3 and 10.12.4, we only considered geopotentials that are close to the radius of the planet. Show that geopotentials have larger radius at the equator than at the poles even when not making this assumption. Hint: maintain the general form of the gravitational potential as given by equation (10.127), then add to the potential for the centrifugal acceleration (10.84). Evaluate the geopotential at the pole and then show that this same geopotential has a larger radial position anywhere equatorward of the pole.

EXERCISE 10.6: SCALING TO JUSTIFY USE OF GEOPOTENTIAL COORDINATES

Summarize the argument that justifies the use of geopotential coordinates while retaining the spherical geometry. Make use of your favorite textbook discussion such that given in Chapter 2 of *Vallis (2017)*.

EXERCISE 10.7: ACCELERATIONS ACTING ON A RESTING PARTICLE

In this exercise we consider the accelerations acting on a particle at rest on a smooth/frictionless rotating spherical planet and a rotating oblate spheroidal planet. We also consider similar questions in Section 11.10.

- (A) Motion of a particle on a rotating spherical planet is described using the spherical coordinates from Section 10.13.2 with the corresponding equations of motion (10.150)-(10.152). Suppose that we drop a particle on a rotating sphere and that we describe its motion using the spherical coordinates. Discuss the initial acceleration of the particle that is released from rest.
- (B) Motion of a particle on a rotating oblate spheroid planet is described using the geopotential coordinates from Section 10.13.3 with the corresponding equations of motion (10.156)-(10.158). Suppose that we drop a particle on a rotating oblate spheroid and that we describe its motion using the geopotential coordinates. Discuss the initial acceleration of the particle that is released from rest.

EXERCISE 10.8: GRAVITATIONAL POTENTIAL INSIDE A SPHERE

In Section 10.12.1 we found the gravitational, Φ , for a point outside of the spherical earth. Here we find the gravitational potential inside of the earth. That is, solve the Poisson equation

$$\nabla^2 \Phi = 4\pi G \rho, \quad (10.162)$$

for a radial position $r < R_e$. Hint: follow the approach in Section 10.12.1 and then match the potential at $r = R_e$. Note: the solution can be found in most first year undergraduate physics texts.



Symmetry and conservation

A symmetry is a discrete or continuous operation that leaves a physical system unchanged. For example, let \mathbf{X} be a trajectory satisfying Newton's equation of motion and \mathcal{A} be an operation, then if $\mathcal{A}[\mathbf{X}]$ also satisfies Newton's equation of motion then \mathcal{A} is a symmetry of the physical system. There is a connection between symmetries and conservation laws and these connections are explored in this chapter. These ideas are fundamental to how we garner a qualitative understanding of motion in which, for many purposes, it is more useful to know the dynamically conserved properties shared by all trajectories rather than details of any particular trajectory. Furthermore, conservation laws provide predictive capabilities and guidance for designing analytical and numerical solution methods.

READER'S GUIDE TO THIS CHAPTER

This chapter relies on the particle mechanics material studied in Chapter 10, and here we explore many facets of rotational physics that are central to geophysical fluid motions. We make use of ideas from this chapter throughout the book.

11.1	Loose threads	208
11.2	Trajectories and dynamical constraints	208
11.2.1	Connecting symmetries to conservation laws	209
11.2.2	Further study	209
11.3	Time reversal symmetry	209
11.4	Potential momentum	210
11.4.1	Basics	210
11.4.2	Comment about terminology	211
11.5	Inertial oscillations	211
11.5.1	Oscillator equation for inertial oscillations	212
11.5.2	Particle trajectory and velocity	212
11.5.3	Period of inertial oscillations	213
11.5.4	Comments and further study	213
11.6	Dynamical constraints from spatial symmetries	214
11.6.1	Linear momentum conservation	214
11.6.2	Potential momentum conservation	214
11.6.3	Angular momentum conservation	215
11.7	Axial angular momentum conservation	215
11.7.1	Angular momentum	216
11.7.2	Conservation of axial angular momentum	216
11.8	Constraints from axial angular momentum conservation	217
11.8.1	Deriving the constraints	217

11.8.2	Axial angular momentum conservation yields zonal Coriolis acceleration	218
11.8.3	Zonal acceleration induced by meridional motion	218
11.8.4	Zonal acceleration induced by radial motion	219
11.8.5	Zonal acceleration derived from axial angular momentum conservation	219
11.8.6	Coriolis acceleration from meridional motion	220
11.8.7	When lateral motions dominate vertical motions	221
11.8.8	Comments	221
11.9	Mechanical energy conservation	222
11.9.1	Some properties of kinetic energy	222
11.9.2	Planetary Cartesian mechanical energy	224
11.9.3	Planetary spherical mechanical energy	224
11.9.4	Geopotential mechanical energy	225
11.10	Sample trajectories using spherical coordinates	226
11.10.1	Free fall to the center	226
11.10.2	Free fall to the equatorial plane	226
11.11	Exercises	227

11.1 Loose threads

- None ATT.

11.2 Trajectories and dynamical constraints

From Newton's law of motion, the trajectory of a particle is specified so long as we know the forces acting on the particle and the particle's initial position and initial velocity. The trajectory encapsulates all dynamical information about the moving particle. However, it is often difficult to unpack that information to understand the nature of the motion. Knowledge of the trajectory is not always the best route to dynamical insight nor to predictive capabilities.

For dynamical insight it is generally more useful to develop an understanding of constraints respected by the motion, with dynamical constraints manifesting as conservation laws. For example, does the motion conserve mechanical energy? What about angular momentum? If dynamical constraints are present, then all trajectories, regardless their complexity, satisfy the constraints. Knowledge of the constraints can reveal information often hidden when just having knowledge of the trajectory. Constraints also provide predictive statements of value when studying the stability of motion and for developing numerical methods for simulations.

An example is useful to illustrate the differences noted above. Consider the following two situations.

- We have the analytic expression for the trajectory of a particle moving around a rotating sphere. So we know this trajectory for all time.
- Instead imagine we know that all particles, no matter what their trajectory, conserve both the axial angular momentum around the rotational axis and the mechanical energy.

Knowledge of the trajectory for a single particle does not readily reveal that this particle is constrained in its motion to conserve angular momentum and mechanical energy. In contrast, knowing that all particles, no matter what trajectory, conserve angular momentum and mechanical energy gives us the ability to understand basic properties of the motion and to furthermore predict the response under perturbations. We offer specifics to this predictive capability later in this chapter

and provide more examples throughout this book. Dynamical constraints imposed by conservation laws are especially relevant for fluids since it is rare to determine the analytical expression for fluid motion, making the knowledge of constraints incredibly valuable.

11.2.1 Connecting symmetries to conservation laws

The discovery of conservation laws often comes from inspired manipulations of the equations of motion. However, there is a more robust and fundamental means to deduce conservation laws through their connection to symmetries, with a symmetry manifesting as an operation that leaves the physical system unchanged. For example, does the physical system remain unchanged when shifting the origin of time? If so, then mechanical energy is a constant of the motion. Likewise, if there is rotational symmetry around an axis, then the associated angular momentum is a constant of the motion.

The connection between symmetries (kinematics) and conservation laws (dynamics) was made by [Noether \(1918\)](#) (see [Noether and Tavel \(2018\)](#) for an English translation). Noether's Theorem is fundamental to nearly all areas of theoretical physics. It is sufficient for our study to make use of this theorem as a conceptual framework for understanding conservation laws and their connections to symmetries. Quite simply, if there is a symmetry then there is a corresponding conservation law, and vice versa.

It is very useful to identify conserved quantities as a means to understand and to constrain the motion. This perspective holds even when the symmetries giving rise to conserved quantities are broken in realistic cases. For example, as seen in Section 10.1.7, friction breaks time translation symmetry and so leads to the dissipation of mechanical energy. Nonetheless, understanding the frictionless motion, and the associated energy conservation law, offers insights for the frictional case. Indeed, for many purposes, knowledge of the trajectory is less important than knowledge of conserved, or partially conserved, dynamical quantities. In this chapter, we offer two examples to support this point: the case of mechanical energy conservation and axial angular momentum conservation. These conservation laws also hold in a modified form when moving to the continuum fluid (e.g., Chapter 18). Additional conservation properties also arise that are unique to the continuum, with conservation of potential vorticity the most notable one for geophysical fluids (Chapter 38).

11.2.2 Further study

The notions of conservation laws and symmetries in classical mechanics are lucidly discussed in Chapters 1 and 2 of [Landau and Lifshitz \(1976\)](#). A pedagogical lecture on these topics can be found in this [online lecture from the Space Time series](#). Finally, [this essay about Emmy Noether](#) provides insights into this mathematician whose work, conducted under some very unfortunate circumstances, forever connected symmetry and conservation laws.

11.3 Time reversal symmetry

A deterministic process is time-reversible if the time-reversed process satisfies the same dynamic equations as the forward-time process. That is, the dynamical equations are symmetric under a change in the sign of time so that the time-reversed evolution of one state is equivalent to the forward-time evolution of a corresponding state. We here discuss time reversal symmetry in the context of the point particle with trajectory, $\mathbf{X}(t)$, and velocity, $\mathbf{V}(t) = d\mathbf{X}(t)/dt$.

If $\mathbf{X}(t)$ is a solution to the equations of motion, then what is needed for

$$\mathbf{X}^*(t^*) = \mathbf{X}(-t) \quad \text{and} \quad \mathbf{V}^*(t^*) = d\mathbf{X}^*(t^*)/dt^* = -d\mathbf{X}(-t)/dt \quad (11.1)$$

to define the same trajectory traversed backwards in time, where $t^* = -t$? We answer this question by recalling from Section 10.13.1 that with planetary rotation a constant in time then the Cartesian coordinate equation of motion is given by

$$\frac{d}{dt} \left[\dot{\mathbf{X}}(t) + 2\boldsymbol{\Omega} \wedge \mathbf{X}(t) \right] = -\nabla\Phi(t). \quad (11.2)$$

We are interested in the constraints that ensure the following equation to also satisfied

$$\frac{d}{dt^*} \left[\frac{d\mathbf{X}^*(t^*)}{dt^*} + 2\boldsymbol{\Omega}^* \wedge \mathbf{X}^*(t^*) \right] = -\nabla\Phi(t^*). \quad (11.3)$$

The effective gravitational acceleration remains time reversible if

$$\Phi^*(t^*) = \Phi(t), \quad (11.4)$$

which is trivially satisfied for $\Phi = g z$. The Coriolis acceleration is velocity dependent so that it generally breaks time reversal symmetry. However, we can recover time symmetry by assuming that the rotation direction switches when time reverses so that

$$\boldsymbol{\Omega}^* = -\boldsymbol{\Omega}. \quad (11.5)$$

With this transformation, a trajectory, $\mathbf{X}(t)$, that solves the forward equation (11.2) yields a trajectory $\mathbf{X}(-t)$ that solves the same equation but with time reversed. We return to this question of time reversal symmetry for the Euler equations of perfect fluid mechanics in Exercise 18.2.

11.4 Potential momentum

In Section 10.13.1 we introduced the *potential momentum*. In this section we study its conservation properties given that it is a constant of the motion for a particle moving on a time independent geopotential in a direction where the geopotential does not change. That is, the conservation of potential momentum arises from a symmetry of the geopotential.

11.4.1 Basics

We again start with the Cartesian coordinate equation of motion

$$\frac{d}{dt} \left[\dot{\mathbf{X}} + 2\boldsymbol{\Omega} \wedge \mathbf{X} \right] = -\nabla\Phi. \quad (11.6)$$

As written it suggests we introduce the *potential momentum* per mass

$$\mathbf{M} \equiv \dot{\mathbf{X}} + 2\boldsymbol{\Omega} \wedge \mathbf{X} = \hat{\mathbf{x}}(u - 2\Omega y) + \hat{\mathbf{y}}(v + 2\Omega x) + \hat{\mathbf{z}} w, \quad (11.7)$$

in which case

$$\dot{\mathbf{M}} = -\nabla\Phi. \quad (11.8)$$

Now let $\hat{\mathbf{s}}$ be a unit vector tangent to the geopotential surface so that $\hat{\mathbf{s}} \cdot \nabla \Phi = 0$. Assuming the geopotential surface is time independent so that $\hat{\mathbf{s}}$ is also time independent, then the equation of motion (11.8) leads to

$$\frac{d(\hat{\mathbf{s}} \cdot \mathbf{M})}{dt} = 0. \quad (11.9)$$

That is, the projection of the potential momentum onto a static geopotential surface is a constant of motion. This dynamical constraint arises since we cannot distinguish one point on the geopotential from another; i.e., there is a symmetry associated with motion along the static geopotential. Noether's Theorem (Section 11.2.1) then says that this geometric symmetry leads to a constant of the motion, here given by that component of potential momentum within the geopotential surface. We illustrate this situation in Figure 11.1 with a horizontal geopotential surface.

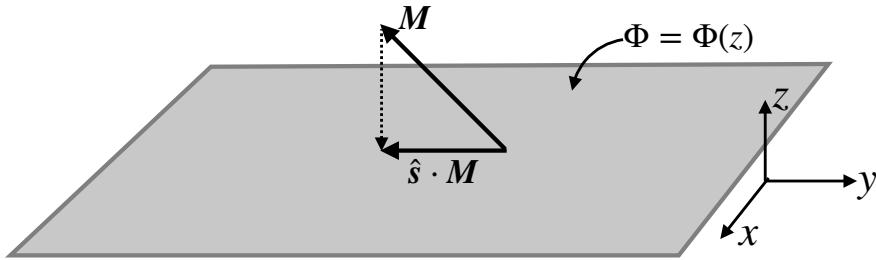


FIGURE 11.1: The projection of the potential momentum onto the geopotential surface is a constant of the motion, $d(\hat{\mathbf{s}} \cdot \mathbf{M})/dt = 0$. Here the geopotential surface is the x-y-plane so that $\hat{\mathbf{x}} \cdot \mathbf{M} = u - 2\Omega y$ and $\hat{\mathbf{y}} \cdot \mathbf{M} = v + 2\Omega x$ are the two conserved components of potential momentum.

Consider a particle with potential momentum \mathbf{M} and move it from an arbitrary point to a reference position with $\mathbf{X} = 0$. Upon reaching the reference position, the horizontal velocity of the particle must equal to \mathbf{M} in order to maintain the same potential momentum. This example motivates the name “potential momentum”, since \mathbf{M} measures the potential for relative motion contained in the particle as it moves along a geopotential.

11.4.2 Comment about terminology

As noted on page 51 of [Markowski and Richardson \(2010\)](#), one might see potential momentum referred to as *pseudo angular momentum*, with some dropping the “pseudo” portion to the name. In either case, it is important to note that potential momentum is distinct from angular momentum. In particular, there is no moment-arm as part of the potential momentum.

Furthermore, most authors use the term “absolute” momentum rather than potential momentum, perhaps in reference to the momentum measured in the absolute or inertial reference frame. However, that connection is incorrect since the inertial velocity is (Section 10.9.1)

$$\mathbf{V} = \mathbf{V}_{\text{Cartesian}} + \mathbf{U}_{\text{solid-body}} = \mathbf{V}_{\text{Cartesian}} + \boldsymbol{\Omega} \wedge \mathbf{X}. \quad (11.10)$$

The factor of two multiplying the rotation rate in the potential momentum arises from the Coriolis acceleration. In contrast, solid-body rotation velocity contributes to the inertial velocity and it has a factor of unity multiplying the rotation rate.

11.5 Inertial oscillations

In Section 23.2 we introduce the tangent plane approximation for motion on a rotating sphere. In this approximation, motion occurs on a rotating geopotential surface with the surface approximated

as horizontal. Furthermore, we use local tangent plane Cartesian coordinates (which are distinct from the planetary Cartesian coordinates used in Chapter 10 and illustrated in Figure 8.1). The f -plane approximation furthermore sets the Coriolis parameter to a constant,

$$f = 2\Omega \sin \phi_0, \quad (11.11)$$

where ϕ_0 is a chosen latitude. Consequently, a particle moving on the f -plane maintains a constant horizontal potential momentum

$$\frac{dM_x}{dt} = \frac{d(u - f y)}{dt} = 0 \quad (11.12a)$$

$$\frac{dM_y}{dt} = \frac{d(v + f x)}{dt} = 0, \quad (11.12b)$$

where we introduced the horizontal velocity components $(u, v) = (\dot{x}, \dot{y})$. These two conservation laws greatly constrain the motion of the particle moving on a constant geopotential surface.

11.5.1 Oscillator equation for inertial oscillations

Taking the time derivative of the zonal equation (11.12a) and using the meridional equation (11.12b) leads to

$$\ddot{u} - f \dot{v} = \ddot{u} + f^2 u = 0. \quad (11.13)$$

Similar manipulations for the meridional velocity equation render the free oscillator equation for each component of the horizontal velocity

$$\frac{d^2u}{dt^2} + f^2 u = 0 \quad \text{and} \quad \frac{d^2v}{dt^2} + f^2 v = 0. \quad (11.14)$$

Motions that satisfy this equation are termed *inertial oscillations*.

11.5.2 Particle trajectory and velocity

Time integrating the equation of motion (11.14) renders the particle trajectory and its velocity

$$\mathbf{X}(t) = (U/f) [\hat{\mathbf{x}} \sin(ft) + \hat{\mathbf{y}} \cos(ft)] \quad (11.15a)$$

$$\mathbf{U}(t) = U [\hat{\mathbf{x}} \cos(ft) - \hat{\mathbf{y}} \sin(ft)], \quad (11.15b)$$

where $U > 0$ is the particle speed, which is a constant, and we assumed the initial conditions

$$\mathbf{X}(0) = (U/f) \hat{\mathbf{y}} \quad \text{and} \quad \mathbf{U}(0) = U \hat{\mathbf{x}}. \quad (11.16)$$

From the particle trajectory equation (11.15a), we see that motion of a particle exhibiting inertial oscillations is circular with radius

$$R = U/|f|. \quad (11.17)$$

As depicted in Figure 11.2, northern hemisphere ($f > 0$) inertial oscillations occur in the clockwise direction whereas southern hemisphere motion is counter-clockwise. Consequently, particle motion undergoing inertial oscillations occurs in an anti-cyclonic sense (opposite to the sense of the rotating reference frame). As we discuss in more detail in Section 27.4, inertial oscillations arise from a balance between the Coriolis acceleration of the rotating frame and the centrifugal

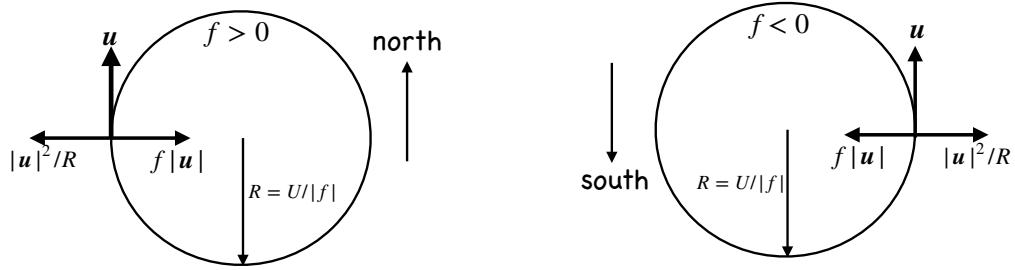


FIGURE 11.2: Inertial oscillation of a particle on a plane occurs when the centrifugal acceleration balances the Coriolis acceleration. Left panel: $f > 0$ for the northern hemisphere, revealing that inertial motion is an anti-cyclonic (clockwise) circular motion with radius $|R| = U/|f|$. The Coriolis acceleration is to the right, pointing into the center of the inertial circle, whereas the centrifugal acceleration points away from the center. Right panel: Counterclockwise inertial motion in the southern hemisphere with the same balance between Coriolis and centrifugal accelerations.

acceleration arising from the particle's circular motion.¹ The only way to realize this balance is for the particle to move anti-cyclonically, with the Coriolis acceleration pointing towards the inside of the inertial circle and the centrifugal acceleration pointing outside. Finally, note that the potential momentum for inertial oscillations vanishes since

$$\mathbf{M}(t) = \mathbf{U}(t) + f \hat{\mathbf{z}} \wedge \mathbf{X}(t) = 0. \quad (11.18)$$

Adding an arbitrary constant to the initial position makes the potential momentum equal to a nonzero constant.

11.5.3 Period of inertial oscillations

Inertial oscillations possess a constant speed and move around the inertial circle with a period

$$T_{\text{inertial}} = \frac{2\pi}{f} = \frac{11.97}{|\sin \phi_0|} \text{ hour}, \quad (11.19)$$

where we set $\Omega = 7.292 \times 10^{-5} \text{ s}^{-1}$ (equation (10.20)). This period is the time it takes to go around the circle. It is smallest at the poles, where the latitude $\phi_0 = \pm\pi/2$ and $T_{\text{smallest}} \approx 12$ hour. At the equator, $\phi_0 = 0$, so that the radius of the inertial circle is infinite and inertial oscillations are unavailable. Furthermore, T_{inertial} is the time for a Foucault pendulum to turn through π radians, so that T_{inertial} is sometimes referred to as one-half a pendulum day.

11.5.4 Comments and further study

Inertial oscillations of fluid elements are described by the above constant potential momentum equation of motion. Such oscillations are commonly measured by ocean current meters, especially in higher latitude regions where diurnal (day-night) variations in wind forcing have a strong projection onto the inertial period. This resonant forcing puts energy into inertial or near-inertial motions. It is quite amazing that such oscillations are indeed found in the ocean, given that we have ignored pressure and friction, which are two forces that impact on fluid elements (as opposed to point particles). The main reason we can observe this motion in the ocean is that upper ocean currents are often generated by winds even in the absence of horizontal pressure gradients, so that there are occasions when the motion is not affected much by pressure gradients.

¹This centrifugal acceleration is distinct from the planetary centrifugal acceleration absorbed into the geopotential (Section 10.12.4). The planetary centrifugal acceleration plays no role in inertial oscillations.

We study inertial waves in Section 26.6 as part of our study of rapidly rotating fluid motion. We also encounter inertial motions in Section 27.4 as part of our characterization of horizontal fluid motion according to the balance between forces. As noted above, we arrive at inertial motion by balancing the Coriolis acceleration with the centrifugal acceleration due to the curved motion of a fluid particle. The name “inertial” does not here refer to motion in an inertial reference frame (Section 10.3). Instead, it refers to the balance between accelerations arising only when the particle is in motion (i.e., has inertia), with these accelerations being the Coriolis and centrifugal.

A rotating tank fluid experiment offers a useful controlled setting to observe inertial waves in a fluid, such as shown near the 18 minute mark in [this video from Prof. Fultz](#).

11.6 Dynamical constraints from spatial symmetries

We here position the conservation of potential momentum among two other dynamical conservation laws by summarizing the spatial symmetries that lead to the conservation laws via Noether’s Theorem. We follow this discussion in Section 11.9, which studies mechanical energy conservation arising from time symmetry.

11.6.1 Linear momentum conservation

As seen in Section 10.1.1, linear momentum remains constant for a particle moving without any forces acting on it; i.e., a *free particle*. This type of motion is not common for geophysical particles or fluids since they feel gravity and so are not free. Even so, we consider this limiting case as a point of comparison for the other conservation laws.

The conservation of linear momentum is simply viewed within the particle’s inertial reference frame, where a vanishing inertial acceleration leads to a constant inertial velocity

$$\mathbf{A} = d\mathbf{V}/dt = 0. \quad (11.20)$$

In a Euclidean space, a vanishing acceleration means the particle is moving in a straight line with constant velocity. When viewed from a rotating frame using Cartesian coordinates, a vanishing inertial acceleration means that the Cartesian acceleration balances Coriolis and centrifugal accelerations

$$\ddot{\mathbf{X}} = -2\boldsymbol{\Omega} \wedge \dot{\mathbf{X}} - \nabla\Phi_{\text{centrifugal}}. \quad (11.21)$$

This equation is clearly not a very useful means to describe unaccelerated free particle motion. We can provide a bit more compactness to this equation by introducing the potential momentum (11.7) so that

$$d\mathbf{M}/dt = -\nabla\Phi_{\text{centrifugal}}. \quad (11.22)$$

While the inertial acceleration vanishes ($\mathbf{A} = d\mathbf{V}/dt = 0$), time changes to the potential momentum are balanced by the gradient of the centrifugal potential.

11.6.2 Potential momentum conservation

Conservation of potential momentum arises from symmetry of particle motion on a constant geopotential surface. The conservation law is most readily viewed within the rotating frame, whereby (equation (11.9))

$$\frac{d(\hat{\mathbf{s}} \cdot \mathbf{M})}{dt} = 0. \quad (11.23)$$

A geopotential is a two-dimensional surface so that this conservation law corresponds to two dynamical constraints such as shown in Figure 11.1.

11.6.3 Angular momentum conservation

As detailed in Section 11.7, the angular momentum computed with respect to the axis of rotation is a constant of the motion (Figure 11.3). This conservation law arises from rotational symmetry of the system about the rotational axis. Axial (z -axis) angular momentum conservation takes the form

$$dL^z/dt = 0, \quad (11.24)$$

where the axial angular momentum is

$$L^z = m r_{\perp}^2 (\dot{\lambda} + \Omega) \quad \text{with} \quad r_{\perp} = \sqrt{x^2 + y^2} = r \cos \phi. \quad (11.25)$$

The distance from the rotation axis, r_{\perp} , is the *moment arm* for the axial angular momentum. The longitude, λ , measures the angle in the counter-clockwise direction from the positive x -axis, and $\dot{\lambda}$ is the time change of the longitude.

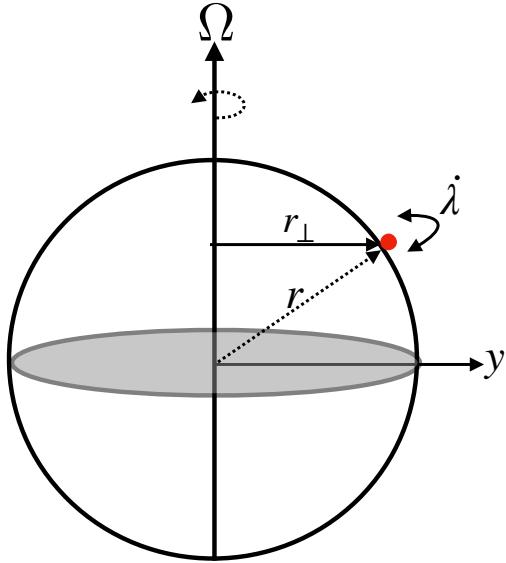


FIGURE 11.3: Axial angular momentum, $L^z = m r_{\perp}^2 (\dot{\lambda} + \Omega)$, is the moment of the zonal momentum around the sphere, with $r_{\perp} = r \cos \phi = \sqrt{x^2 + y^2}$ the moment arm length; i.e., the distance of the particle to the axis of rotation, whereas $r^2 = \sqrt{x^2 + y^2 + z^2}$ is the radial distance to the center of the sphere. L^z is a constant of the motion for the particle moving in the absence of friction.

11.7 Axial angular momentum conservation

Does the particle know anything about the longitudinal angle, λ ? Since we assume that the sphere is smooth (i.e., no mountains), and since the sphere is rotating around the polar axis, there is an arbitrariness in how we choose the zero for the longitudinal angle. That is, the physical system remains unchanged if we shift the longitudinal angle by a constant. Noether's Theorem then says that this rotational symmetry leads to a corresponding angular momentum conservation. Hence, the particle's angular momentum around the rotational axis remains fixed by the initial conditions. We here prove that axial angular momentum is constant by manipulating the equations of motion. Many of the same manipulations also occur when considering angular conservation for a continuous fluid (see Section 18.4).

11.7.1 Angular momentum

The angular momentum of the particle, computed with respect to the origin of the sphere, is given by

$$\mathbf{L} = m \mathbf{X} \wedge \mathbf{V}. \quad (11.26)$$

This is the moment of the linear momentum computed relative to the origin. We write the angular momentum computed along the polar axis as

$$L^z = \mathbf{L} \cdot \hat{\mathbf{z}} \quad (11.27a)$$

$$= m (\mathbf{X} \wedge \mathbf{V}) \cdot \hat{\mathbf{z}} \quad (11.27b)$$

$$= m (\hat{\mathbf{z}} \wedge \mathbf{X}) \cdot \mathbf{V} \quad (11.27c)$$

$$= m r \cos \phi (\hat{\lambda} \cdot \mathbf{V}) \quad (11.27d)$$

$$= m r_{\perp} (\hat{\lambda} \cdot \mathbf{V}). \quad (11.27e)$$

Hence, the angular momentum about the polar axis equals to the component of the linear momentum in the longitudinal direction, multiplied by the distance to the polar rotational axis (the moment-arm)

$$r_{\perp} = r \cos \phi. \quad (11.28)$$

In deriving equation (11.27e), we made use of the identity

$$\hat{\mathbf{z}} \wedge \mathbf{X} = r_{\perp} \hat{\lambda}, \quad (11.29)$$

which will prove useful below for proving that axial angular momentum is a constant of the motion.

We now write the axial angular momentum in equation (11.27e) in terms of the rotating frame quantities. To do so, introduce the inertial velocity written using spherical coordinates according to equation (10.66f), which yields

$$L^z = m r_{\perp} (\hat{\lambda} \cdot \mathbf{V}) = m r_{\perp}^2 (\dot{\lambda} + \Omega) = m r_{\perp} (u + r_{\perp} \Omega). \quad (11.30)$$

When measured from the rotating terrestrial frame, the axial angular momentum consists of two terms: one from the zonal velocity of the particle relative to the planet and another from the solid-body motion of the planet.

11.7.2 Conservation of axial angular momentum

The time derivative of the axial angular momentum is given by

$$\frac{dL^z}{dt} = m \frac{d}{dt} [(\mathbf{X} \wedge \mathbf{V}) \cdot \hat{\mathbf{z}}] \quad (11.31a)$$

$$= m \frac{d}{dt} [(\hat{\mathbf{z}} \wedge \mathbf{X}) \cdot \mathbf{V}] \quad (11.31b)$$

$$= m (\hat{\mathbf{z}} \wedge \mathbf{V}) \cdot \mathbf{V} + m (\hat{\mathbf{z}} \wedge \mathbf{X}) \cdot \mathbf{A} \quad (11.31c)$$

$$= m (\hat{\mathbf{z}} \wedge \mathbf{X}) \cdot \mathbf{A} \quad (11.31d)$$

$$= m r_{\perp} \hat{\lambda} \cdot \mathbf{A}. \quad (11.31e)$$

To reach this result we noted that the polar axis direction, $\hat{\mathbf{z}}$, is time independent, and we used the identity (11.29) for the final step. The inertial acceleration arises just from the central-force gravitational field (equation (10.141))

$$\mathbf{A} = -\nabla \Phi_e = -g_e \hat{\mathbf{r}}. \quad (11.32)$$

Since $\hat{\lambda} \cdot \hat{r} = 0$, we have axial angular momentum conservation

$$\frac{dL^z}{dt} = 0. \quad (11.33)$$

11.8 Constraints from axial angular momentum conservation

Axial angular momentum conservation constrains the particle motion and in turn it offers a means for interpreting and predicting motion on the sphere. It is also a natural means for describing the motion when viewed from a non-rotating inertial reference frame. We here explore how these constraints manifest through thought experiments illustrated in Figure 11.4. In doing so, we show that the acceleration induced by axial angular momentum conserving motion corresponds to the Coriolis acceleration that is part of a rotating non-inertial reference frame description. That is, angular momentum constrained motion described in an inertial reference frame corresponds to the Coriolis acceleration appearing in a rotating non-inertial frame description.

Throughout the following we write the axial angular momentum as

$$L^z = m l^z = m r_{\perp}^2 (\dot{\lambda} + \Omega) = m r_{\perp} (u + r_{\perp} \Omega), \quad (11.34)$$

where

$$l^z = L^z/m \quad \text{and} \quad u = r_{\perp} \dot{\lambda} \quad (11.35)$$

are the axial angular momentum per mass and zonal velocity component. Since the mass of the particle is constant, conservation of $L^z = m l^z$ implies conservation of l^z . Note that for most geophysical applications, $l^z > 0$ since solid-body motion dominates over the relative zonal velocity:

$$r_{\perp} \Omega > |u| \quad \text{for most terrestrial motions,} \quad (11.36)$$

with $r_{\perp} \Omega = 465 \text{ m s}^{-1}$ at the equator (see Figure 10.2). In Exercise 11.1 we consider the interesting, but geophysically uncommon, case where $l^z \leq 0$.

11.8.1 Deriving the constraints

Conservation of axial angular momentum says that it is not possible to change u or $\dot{\lambda}$ without also changing r_{\perp} in such a manner to ensure that l^z remains fixed. To determine the relation between these changes we set $\delta l^z = 0$, where δ refers to any small change. When writing $l^z = r_{\perp}^2 (\dot{\lambda} + \Omega)$ we have

$$\delta l^z = 2 r_{\perp} \delta r_{\perp} (\dot{\lambda} + \Omega) + r_{\perp}^2 \delta \dot{\lambda} = 2 l^z \frac{\delta r_{\perp}}{r_{\perp}} + r_{\perp}^2 \delta \dot{\lambda}, \quad (11.37)$$

where we set $\delta \Omega = 0$ since the earth's rotation rate is assumed to be fixed. Furthermore, we find it useful to write the changes in terms of l^z since it is a constant of the motion. Likewise, when writing $l^z = r_{\perp} (u + r_{\perp} \Omega)$ we have

$$\delta l^z = \delta r_{\perp} (u + r_{\perp} \Omega) + r_{\perp} (\delta u + \Omega \delta r_{\perp}) = \Omega r_{\perp} \delta r_{\perp} \left[1 + \frac{l^z}{\Omega r_{\perp}^2} \right] + r_{\perp} \delta u. \quad (11.38)$$

Setting $\delta l^z = 0$ then leads to the constraints

$$\delta \dot{\lambda} = -\frac{2 l^z}{r_{\perp}^2} \frac{\delta r_{\perp}}{r_{\perp}} \iff \delta u = -\Omega \delta r_{\perp} \left[1 + \frac{l^z}{\Omega r_{\perp}^2} \right]. \quad (11.39)$$

As noted above, $l^z > 0$ is generally the case for geophysical fluid motion. Consequently, axial angular momentum conserving motion that brings the particle closer to the rotation axis ($\delta r_\perp < 0$) leads to an eastward velocity change ($\delta \dot{\lambda} > 0$ and $\delta u > 0$). The opposite occurs for motion with $\delta r_\perp > 0$. These results hold in both the northern and southern hemispheres.

Since $r_\perp = r \cos \phi$, the distance to the rotational axis can change either by changing the radial position or the meridional position

$$\delta r_\perp = (\cos \phi) \delta r - (r \sin \phi) \delta \phi. \quad (11.40)$$

Assuming these changes occur over a small time increment δt allows us to write

$$\delta r_\perp / \delta t = w \cos \phi - v \sin \phi, \quad (11.41)$$

where

$$v = r \delta \phi / \delta t = r \dot{\phi} \quad \text{and} \quad w = \delta r / \delta t = \dot{r} \quad (11.42)$$

introduced the meridional and vertical components to the particle velocity according to equation (10.68). For geophysical fluid motion, changes in vertical distance δr are far smaller than the distance to the earth's center, $\delta r \ll r$. In this case, when $\phi \neq 0$, then δr_\perp is affected much more by meridional motion at constant radial position (second term on right hand side of equation (11.41)) than by vertical motion at constant latitude (first term). We return to this observation in Section 11.8.7 when discussing nearly the shallow fluid approximation used to develop the primitive equations for the atmosphere and ocean in Section 23.1.2.

11.8.2 Axial angular momentum conservation yields zonal Coriolis acceleration

Consider the zonal acceleration felt by a particle at rest (in the rotating frame) whose angular momentum per mass equals $l^z = r_\perp^2 \Omega$. In this case the zonal acceleration in equation (11.39) is

$$\delta \dot{\lambda} = -2 \Omega \frac{\delta r_\perp}{r_\perp} \iff \dot{u} = -2 \Omega (\delta r_\perp / \delta t) = -2 \Omega (w \cos \phi - v \sin \phi), \quad (11.43)$$

where we used equation (11.41) for $\delta r_\perp / \delta t$. The expression for \dot{u} is precisely the same as the Coriolis acceleration appearing in the zonal momentum equation (10.156)

$$\dot{u} + \frac{u (w - v \tan \phi)}{r} = \underbrace{-2 \Omega (w \cos \phi - v \sin \phi)}_{\text{Coriolis acceleration}}. \quad (11.44)$$

Hence, the Coriolis acceleration appearing in the zonal momentum equation is identical to the zonal acceleration induced by constraining the motion to conserve axial angular momentum. That is, by unpacking the constraint of axial angular momentum conservation to reveal the zonal momentum equation, the Coriolis acceleration is revealed as part of that package. We pursue this connection in Section 11.8.5 by deriving the zonal momentum equation from the axial angular momentum equation.

11.8.3 Zonal acceleration induced by meridional motion

Consider a particle moving meridionally ($\delta \phi \neq 0$) while maintaining a constant radial position ($\delta r = 0$). The axial angular momentum constraint (11.43) induces a zonal acceleration

$$\dot{u} = 2 \Omega v \sin \phi, \quad (11.45)$$

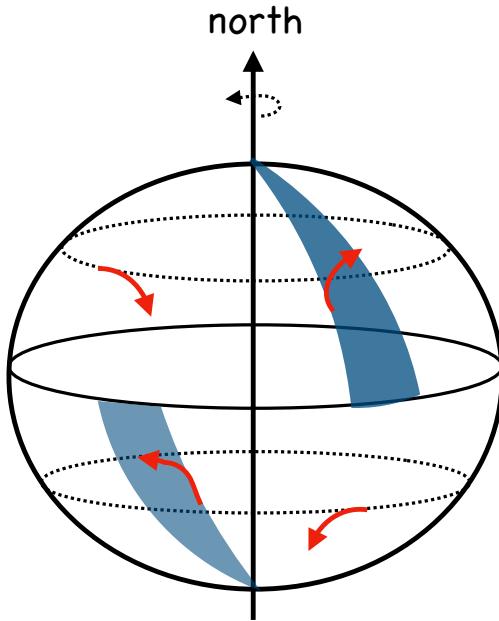


FIGURE 11.4: In the northern hemisphere, the Coriolis acceleration viewed from the rotating reference frame deflects a particle to its right whereas it deflects a particle to the left in the southern hemisphere. This deflection is also a result of axial angular momentum conservation, which is the natural way to describe the motion from a non-rotating inertial reference frame. We can summarize the ideas developed in Section 11.8 as follows. For a meridionally displaced trajectory, the spinning earth causes the trajectory to pick up motion relative to the earth surface, zonally to the right in the northern hemisphere and zonally to the left in the southern hemisphere. For a zonally displaced trajectory to conserve angular momentum, the particle must pick up meridional motion; poleward if the zonal displacement is westward and equatorward if the displacement is eastward.

which, as seen by equation (11.44), is the Coriolis acceleration appearing in the zonal momentum equation arising from the meridional motion. For poleward motion in either hemisphere, $v \sin \phi > 0$. Hence, axial angular momentum conserving motion towards either pole induces an eastward acceleration, whereas a westward acceleration is induced for equatorward motion. For the northern hemisphere, the induced acceleration deflects the particle to the right when looking downstream whereas in the southern hemisphere the induced acceleration deflects the particle to the left. These deflections are illustrated in Figure 11.4.

11.8.4 Zonal acceleration induced by radial motion

Now consider a particle moving radially while holding the latitude fixed ($\delta r \neq 0$ and $\delta\phi = 0$). The axial angular momentum constraint (11.43) induces a zonal acceleration

$$\dot{u} = -2\Omega w \cos \phi, \quad (11.46)$$

which, as seen by equation (11.44), is the Coriolis acceleration appearing in the zonal momentum equation arising from the vertical motion. Hence, for vertically downward motion ($w < 0$), axial angular momentum conservation induces a positive zonal acceleration, $\dot{u} > 0$, which we expect from axial angular momentum conservation since the particle is moving closer to the rotation axis.

11.8.5 Zonal acceleration derived from axial angular momentum conservation

The above discussion can be formalized by analyzing how the conservation of axial angular momentum leads to an expression for the zonal acceleration. For this purpose, compute the time

derivative of the first form of the axial angular momentum in equation (11.34), in which case

$$\frac{1}{m} \frac{dL^z}{dt} = \frac{d[(r \cos \phi)^2 (\dot{\lambda} + \Omega)]}{dt} \quad (11.47a)$$

$$= 2(\dot{r} \cos \phi - r \dot{\phi} \sin \phi)(\dot{\lambda} r \cos \phi + r \Omega \cos \phi) + (r \cos \phi)^2 \ddot{\lambda} \quad (11.47b)$$

$$= 2(w \cos \phi - v \sin \phi)(u + r \Omega \cos \phi) + (r \cos \phi)^2 \ddot{\lambda}, \quad (11.47c)$$

where we introduced the (u, v, w) velocity components according to equation (10.68). With the zonal velocity $u = \dot{\lambda} r_{\perp} = \dot{\lambda} r \cos \phi$, we have

$$r \cos \phi \ddot{\lambda} = \dot{u} + \frac{u}{r \cos \phi} (v \sin \phi - w \cos \phi), \quad (11.48)$$

so that equation (11.47c) takes the form

$$\frac{1}{m} \frac{dL^z}{dt} = 2(w \cos \phi - v \sin \phi)(u + r \Omega \cos \phi) + (r \cos \phi)^2 \ddot{\lambda} \quad (11.49a)$$

$$= (w \cos \phi - v \sin \phi)(u + 2r \Omega \cos \phi) + \dot{u} r \cos \phi. \quad (11.49b)$$

Setting $dL^z/dt = 0$ and rearranging then leads to a prognostic equation for the zonal velocity

$$\frac{du}{dt} = \left[\frac{u}{r \cos \phi} + 2\Omega \right] (v \sin \phi - w \cos \phi). \quad (11.50)$$

The first term in the bracket arises from curvature of the sphere (the “metric acceleration”) whereas the second term is the Coriolis acceleration.

The same result can be obtained by performing the time derivative on the second form of the axial angular momentum in equation (11.34), in which case

$$\frac{1}{m} \frac{dL^z}{dt} = \frac{d[u r \cos \phi + \Omega(r \cos \phi)^2]}{dt} \quad (11.51a)$$

$$= \dot{u} r \cos \phi + u \dot{r} \cos \phi - u r \dot{\phi} \sin \phi + 2\Omega r \cos \phi (\dot{r} \cos \phi - r \dot{\phi} \sin \phi). \quad (11.51b)$$

Again, setting $dL^z/dt = 0$ and rearranging leads to the zonal acceleration (11.50).

11.8.6 Coriolis acceleration from meridional motion

What happens when zonal motion is perturbed in the meridional direction? Following the angular momentum constraint (11.43), a poleward perturbation, $\delta\phi > 0$, to a northern hemisphere particle will have $\dot{u} > 0$, with this zonal acceleration corresponding to a rightward deflection relative to the $\delta\phi > 0$ perturbation. Likewise, for $\delta\phi < 0$ a northern hemisphere particle will have $\dot{u} < 0$, which corresponds to a rightward deflection relative to the $\delta\phi < 0$ perturbation.

We see the correspondence to the Coriolis acceleration when viewed in the rotating reference frame by examining the meridional momentum equation (10.157)

$$\dot{v} + \frac{v w + u^2 \tan \phi}{r} = -2\Omega \underline{u \sin \phi}, \quad (11.52)$$

with the Coriolis acceleration underlined on the right hand side. In the northern hemisphere ($\sin \phi > 0$), the Coriolis acceleration gives rise to a rightward (equatorward) acceleration, $-2\Omega \underline{u \sin \phi} < 0$, when the particle is moving eastward, $u > 0$ (super-rotating), thus inducing a negative meridional acceleration, $\dot{v} < 0$. Conversely, if the particle is moving to the west so that $u < 0$, then the Coriolis acceleration is again to the right, only this time it deflects the particle in the poleward direction with $\dot{v} > 0$. The analogous considerations hold in the southern hemisphere where the particle is deflected to the left by the Coriolis acceleration. These motions are reflected in Figure 11.4.

11.8.7 When lateral motions dominate vertical motions

We here consider two approximations relevant to large scale geophysical fluid motions.

1. The particle kinetic energy is dominated by lateral motions on the sphere (i.e., motion at constant radial position).
2. Vertical (radial) excursions are much smaller than the earth radius.

When applied to a fluid, the first assumption leads to the hydrostatic approximation (Section 23.3), and the second assumption leads to the shallow fluid approximation (Section 23.1.2). Self-consistency of the equations of motion means that these two assumptions must be applied together.

Dropping the vertical velocity component to the kinetic energy leads to

$$K \approx \frac{m}{2} [(u + r_{\perp} \Omega)^2 + v^2]. \quad (11.53)$$

The second assumption means that the axial angular momentum takes the approximate form

$$L^z \approx m R_{\perp} (u + \Omega R_{\perp}) = m R_{\perp}^2 (\dot{\lambda} + \Omega), \quad (11.54)$$

where

$$r = R + z \approx R \quad \text{and} \quad R_{\perp} = R \cos \phi. \quad (11.55)$$

The approximate angular momentum (11.54) ignores contributions from vertical motion in changing the moment-arm. Indeed, as noted in Section 11.8.1, vertical movements within the atmosphere and ocean (relatively thin fluid layers over the earth's surface) lead to a relatively small modification to the moment-arm, so the assumption that $r_{\perp} \approx R \cos \phi$ is reasonable. With $r \approx R$, the zonal acceleration (11.50) is modified to the form

$$\frac{du}{dt} = v \left[\frac{u \tan \phi}{R} + f \right] \quad \text{where} \quad f = 2 \Omega \sin \phi. \quad (11.56)$$

That is, we dropped the vertical velocity component, w , from the general form of the acceleration (11.50). Correspondingly, the meridional momentum equation takes the form

$$\frac{dv}{dt} = -u \left[\frac{u \tan \phi}{R} + f \right]. \quad (11.57)$$

These approximate forms for the zonal and meridional accelerations appear in the primitive equations of geophysical fluid dynamics presented in Section 23.1.

11.8.8 Comments

To orient oneself according to the Coriolis acceleration, one needs to remember one fact: in the northern hemisphere motion on the sphere at constant radius ($\delta r = 0$) leads to a rightward deflection whereas in the southern hemisphere the particle experiences a leftward deflection. Figure 11.4 depicts some sample deflections.

A concise summary of many features of rotating physics is provided by [this video from SciencePrimer](#).

11.9 Mechanical energy conservation

Does the particle know anything about the origin of time? Since the angular velocity of the planet and the gravitational acceleration are both assumed constant in time, then changing the time will leave the physical system unaltered. That is, the physical system remains unchanged if we shift all clocks by a constant amount. Through Noether's Theorem, this symmetry in time leads to mechanical energy conservation. That is, the particle's mechanical energy is fixed by the initial conditions. We here prove that mechanical energy is constant in time by manipulating the equations of motion. Many of the manipulations also occur when considering the mechanical energy conservation laws for a continuum fluid discussed in Chapter 18.

11.9.1 Some properties of kinetic energy

In this subsection we establish some basic properties of kinetic energy for a particle. As we saw in Section 10.1.5, changes in the kinetic energy of a particle equal to the *mechanical work* done on the particle as it moves along its trajectory

$$K(t_2) - K(t_1) = \int_{\mathbf{x}_1}^{\mathbf{x}_2} \mathbf{F} \cdot d\mathbf{x} = \int_{t_1}^{t_2} \mathbf{F} \cdot \mathbf{V} dt, \quad (11.58)$$

where $\mathbf{V} dt = d\mathbf{x}$ defines the vector increment along the trajectory, and $\mathbf{x}_{1,2}$ are the endpoints of the trajectory. The integrand, $\mathbf{F} \cdot \mathbf{V}$, is known as the *power*. Hence, equation (11.58) says that the time integral of the power equals to the difference in kinetic energy between the final and initial times.

The kinetic energy is *not* Galilean invariant since movement to another inertial reference frame leads to the kinetic energy change

$$\bar{\mathbf{V}} = \mathbf{V} + \mathbf{U} \implies \bar{K} = K + \frac{m}{2} (2\mathbf{V} \cdot \mathbf{U} + \mathbf{U} \cdot \mathbf{U}), \quad (11.59)$$

where \mathbf{U} is a constant boost velocity so that $d\mathbf{U}/dt = 0$. We do not expect kinetic energy to be Galilean invariant since it measures energy of motion relative to the chosen reference frame. Even so, the time change of the kinetic energy in a different inertial frame is given by the power in the new frame

$$\frac{d\bar{K}}{dt} = \frac{dK}{dt} + m \mathbf{A} \cdot \mathbf{U} = \mathbf{F} \cdot \mathbf{V} + \mathbf{F} \cdot \mathbf{U} = \mathbf{F} \cdot \bar{\mathbf{V}}. \quad (11.60)$$

Hence, we can directly connect kinetic energy changes to forces within an arbitrary inertial reference frame.

Cartesian expression for kinetic energy

Consider the expression for kinetic energy when introducing the velocity of the rotating reference frame. Writing the inertial velocity in the planetary Cartesian form

$$\mathbf{V} = \mathbf{V}_{\text{Cartesian}} + \mathbf{U}_{\text{solid}}, \quad (11.61)$$

leads to

$$K = \frac{m}{2} [\mathbf{V}_{\text{Cartesian}} \cdot \mathbf{V}_{\text{Cartesian}} + 2\mathbf{V}_{\text{Cartesian}} \cdot \mathbf{U}_{\text{solid}} + \mathbf{U}_{\text{solid}} \cdot \mathbf{U}_{\text{solid}}] \quad (11.62)$$

The first term arises from motion of the particle relative to the rotating sphere; the second arises from coupling between relative velocity and solid-body velocity; and the third arises from solid-body motion of the sphere.

Spherical expression for kinetic energy: Part I

To expose spherical symmetry of the physical system, we express the kinetic energy in terms of the planetary spherical coordinates defined in Figure 8.1. Doing so for the solid body velocity leads to equation (10.70)

$$\mathbf{U}_{\text{solid}} = \Omega r \cos \phi (-\sin \lambda \hat{\mathbf{x}} + \cos \lambda \hat{\mathbf{y}}). \quad (11.63)$$

Likewise, the velocity components measured in the rotating frame are given by

$$\dot{X} = \frac{d(r \cos \phi \cos \lambda)}{dt} = \dot{r} \cos \phi \cos \lambda - r \dot{\phi} \sin \phi \cos \lambda - r \dot{\lambda} \cos \phi \sin \lambda \quad (11.64a)$$

$$\dot{Y} = \frac{d(r \cos \phi \sin \lambda)}{dt} = \dot{r} \cos \phi \sin \lambda - r \dot{\phi} \sin \phi \sin \lambda + r \dot{\lambda} \cos \phi \cos \lambda \quad (11.64b)$$

$$\dot{Z} = \frac{d(r \sin \phi)}{dt} = \dot{r} \sin \phi + r \dot{\phi} \cos \phi. \quad (11.64c)$$

Bringing terms together then leads to the kinetic energy in terms of spherical coordinates

$$K = \frac{m}{2} \left[(\dot{r}^2 + r^2 \dot{\phi}^2 + \dot{\lambda}^2 r^2 \cos^2 \phi) + (2 \Omega r^2 \dot{\lambda} \cos^2 \phi) + (\Omega r \cos \phi)^2 \right]. \quad (11.65)$$

Spherical expression for kinetic energy: Part II

An alternative means for deriving the kinetic energy in equation (11.65) makes use of the spherical coordinate form of the inertial velocity given by equation (10.66f), in which case

$$\mathbf{V} = (u + r_{\perp} \Omega) \hat{\lambda} + v \hat{\phi} + w \hat{r}, \quad (11.66)$$

so that

$$K = \frac{m}{2} [(u + r_{\perp} \Omega)^2 + v^2 + w^2], \quad (11.67)$$

where $r_{\perp} = r \cos \phi$. Additionally, as discussed in Section 11.7, the axial angular momentum is given by

$$L^z = m r_{\perp} (u + r_{\perp} \Omega) \equiv m l^z, \quad (11.68)$$

and this property is a constant of the motion when there is azimuthal (zonal) symmetry. It is thus convenient to write the kinetic energy as

$$K = \frac{m}{2} [(l^z/r_{\perp})^2 + v^2 + w^2]. \quad (11.69)$$

Geopotential expression for mechanical energy

As seen in Section 11.9.4, the conserved form of the mechanical energy using geopotential coordinates takes on the form

$$M_{\text{geop}} = \frac{m}{2} [u^2 + v^2 + w^2 + 2 g r], \quad (11.70)$$

where each of these symbols takes on their geopotential interpretation.

11.9.2 Planetary Cartesian mechanical energy

The time derivative of the kinetic energy is given by

$$\frac{dK}{dt} = m \mathbf{V} \cdot \frac{d\mathbf{V}}{dt} = m \mathbf{V} \cdot \mathbf{A} = -m \mathbf{V} \cdot \nabla \Phi_e. \quad (11.71)$$

For the final equality we introduced the gravitational potential given that the particle only feels an external force from gravity as per equation (10.141). The gravitational potential is given by (see equation (10.135))

$$\Phi_e = g_e r, \quad (11.72)$$

so that

$$\frac{dK}{dt} = -m \mathbf{V} \cdot \nabla \Phi_e = -m g_e \dot{r}. \quad (11.73)$$

This result means that kinetic energy is reduced when moving the particle away from the earth center ($\dot{r} > 0$). Moving away from the earth requires work to overcome the gravitational attraction. This work to overcome the gravitational attraction is taken away from the kinetic energy of the particle. Furthermore, the work is added to the gravitational potential energy, whose evolution is given by (see equation (10.137))

$$\frac{dP_e}{dt} = m g_e \dot{r}, \quad (11.74)$$

where we assumed a constant gravitational acceleration g_e . Consequently, as the particle moves away from the earth center, its reduction in kinetic energy is exactly compensated by an increase in potential energy. Hence, the mechanical energy for the particle remains constant throughout the motion

$$\frac{d(K + P_e)}{dt} = 0, \quad (11.75)$$

where the mechanical energy is the sum of the inertial kinetic energy plus the gravitational potential energy

$$M = K + P_e \quad (11.76a)$$

$$= \frac{m}{2} \mathbf{V} \cdot \mathbf{V} + m \Phi_e \quad (11.76b)$$

$$= \frac{m}{2} [(u + r_\perp \Omega)^2 + v^2 + w^2] + m g_e r \quad (11.76c)$$

$$= \frac{m}{2} [(l^z/r_\perp)^2 + v^2 + w^2] + m g_e r. \quad (11.76d)$$

11.9.3 Planetary spherical mechanical energy

It is physically revealing to expose the exchange of mechanical energy between kinetic and gravitational potential energies. Furthermore, knowledge of the total mechanical energy at any time affords knowledge for all time since the mechanical energy remains constant. Following from the discussion in Section 11.8, where we examined the constraints on particle motion due to conservation of axial angular momentum, we here ask similar questions about mechanical energy conservation. We make use of the spherical form of the equations of motion, equations (10.150)-(10.152), which expose the planetary centrifugal and Coriolis accelerations, and pursue the calculation for geopotential coordinates in Section 11.9.4.

We find it convenient to write the momentum equations (10.150)-(10.152) in terms of the distance to the polar axis, $r_\perp = r \cos \phi$, and its time derivative, $\dot{r}_\perp = w \cos \phi - v \sin \phi$

$$\frac{d}{dt} [r_\perp u + \Omega r_\perp^2] = 0 \quad (11.77a)$$

$$\dot{v} = -\frac{v w}{r} - \frac{u \tan \phi}{r} (u + 2 \Omega r_\perp) - r_\perp \Omega^2 \sin \phi \quad (11.77b)$$

$$\dot{w} = \frac{u^2 + v^2}{r} + 2 \Omega u \cos \phi + r_\perp \Omega^2 \cos \phi - g_e. \quad (11.77c)$$

Equation (11.77a) expresses the conservation of axial angular momentum, $\dot{l}^z = 0$, where the axial angular momentum per mass is $l^z = r_\perp (u + r_\perp \Omega)$ (Section 11.7). For the mechanical energy, we have

$$\dot{M} = 0 \quad \text{with} \quad M = \frac{m}{2} [(l^z/r_\perp)^2 + v^2 + w^2] + m g_e r, \quad (11.78)$$

where $l^z/r_\perp = u + r_\perp \Omega$ is the zonal component to the inertial velocity (equation (10.66f)). We now show that $\dot{M} = 0$ arises from the momentum equations. Performing the time derivative, and setting $d\dot{l}^z/dt = 0$, leads to

$$\frac{1}{m} \dot{M} = \frac{1}{2} \frac{d}{dt} [(l^z/r_\perp)^2 + v^2 + w^2 + 2 g_e r] = -\frac{(l^z)^2 \dot{r}_\perp}{(r_\perp)^3} + v \dot{v} + w (\dot{w} + g_e). \quad (11.79)$$

Use of the meridional momentum equation (11.77b) renders

$$v \dot{v} = -v \left[\frac{v w}{r} + \frac{u \tan \phi}{r} (u + 2 \Omega r_\perp) + r_\perp \Omega^2 \sin \phi \right] \quad (11.80a)$$

$$= -\frac{v^2 w}{r} - \frac{v \tan \phi}{r} (u + r_\perp \Omega)^2. \quad (11.80b)$$

Likewise, the vertical momentum equation (11.77c) renders

$$w (\dot{w} + g_e) = \frac{v^2 w}{r} + \frac{w}{r} [u^2 + 2 \Omega u r_\perp + (r_\perp \Omega)^2] \quad (11.81a)$$

$$= \frac{v^2 w}{r} + \frac{w}{r} (u + r_\perp \Omega)^2, \quad (11.81b)$$

so that

$$v \dot{v} + w (\dot{w} + g_e) = r^{-1} (-v \tan \phi + w) (u + r_\perp \Omega)^2 = \frac{(l^z)^2 \dot{r}_\perp}{(r_\perp)^3}. \quad (11.82)$$

Combining this result with equation (11.79) leads to the expected $\dot{M} = 0$, so that angular momentum conservation combined with the meridional and vertical momentum equations is equivalent to mechanical energy conservation.

11.9.4 Geopotential mechanical energy

Now consider the mechanical energy conservation when using geopotential coordinates with the equations of motion (10.156)-(10.158)

$$\dot{u} + \frac{u (w - v \tan \phi)}{r} + 2 \Omega (w \cos \phi - v \sin \phi) = 0 \quad (11.83a)$$

$$\dot{v} + \frac{v w + u^2 \tan \phi}{r} + 2 \Omega u \sin \phi = 0 \quad (11.83b)$$

$$\dot{w} - \frac{u^2 + v^2}{r} - 2 \Omega u \cos \phi = -g. \quad (11.83c)$$

Multiplying each equation by its respective velocity component and summing yields

$$\frac{1}{m} \dot{M}_{\text{geop}} = \frac{1}{2} \frac{d}{dt} [u^2 + v^2 + w^2 + 2gr] = 0. \quad (11.84)$$

The conserved mechanical energy for the particle written in geopotential coordinates

$$M_{\text{geop}} = m(u^2 + v^2 + w^2)/2 + m gr \quad (11.85)$$

is compared to that written using spherical coordinates in equation (11.78). We note that since the coordinates are different (spherical versus geopotential), the velocity components (u, v, w) are slightly different in the two coordinate systems, as is the radial position.

11.10 Sample trajectories using spherical coordinates

We here examine two relatively simple trajectories using spherical coordinates. We do so to illustrate how the dynamical properties discussed in this chapter, particularly angular momentum conservation, help to determine trajectories. The discussion complements that considered in Exercise 10.7.

11.10.1 Free fall to the center

A particle freely falling to the center of the sphere has a trajectory given by

$$u = -r_{\perp} \Omega \quad \text{and} \quad v = 0 \quad \text{and} \quad \dot{w} = -g_e \implies w = -g_e t + w_0. \quad (11.86)$$

A zonal velocity of $u = -r_{\perp} \Omega$ (to the west) means that the particle has zero zonal inertial velocity; i.e., the particle has a fixed longitude in absolute space. It also means there is zero axial angular momentum, $l^z = 0$. To verify that axial angular momentum remains zero, note that

$$\dot{l}^z = 0 \implies \dot{u} r_{\perp} = -\dot{r}_{\perp} (u + 2\Omega r_{\perp}), \quad (11.87)$$

which is satisfied since $u = -r_{\perp} \Omega$ means that $\dot{u} = -\dot{r}_{\perp} \Omega$. Furthermore, note that the zonal acceleration is given by

$$\ddot{u} = -\Omega \dot{r}_{\perp} = -\Omega(w \cos \phi - v \sin \phi) = -\Omega w \cos \phi = (\Omega g_e \cos \phi) t > 0, \quad (11.88)$$

where we assumed the particle is released with $w_0 = 0$. As the particle falls, and thus reduces its distance from the polar axis, it zonally accelerates with $\ddot{u} > 0$ to maintain zero angular momentum. This acceleration slows the particle's relative speed, which is expected since $u = -r_{\perp} \Omega$ means the relative zonal velocity goes to zero as the particle approaches the axis of rotation.

11.10.2 Free fall to the equatorial plane

Now consider a trajectory with a constant zonal velocity, in which case $\dot{u} = 0$ so that the angular momentum equation (11.77a) leads to

$$\dot{r}_{\perp} (u + 2\Omega r_{\perp}) = 0. \quad (11.89)$$

There are two ways to satisfy this constraint, one being to set $u = -2\Omega r_{\perp}$. We here choose to instead examine the case with a fixed moment-arm,

$$\dot{r}_{\perp} = 0 \implies w \cos \phi = v \sin \phi. \quad (11.90)$$

This relation couples vertical and meridional motion in a manner that keeps the distance to the rotational axis fixed. For example, if the particle is falling towards the earth center so that $w \cos \phi < 0$, then there is an equatorward meridional velocity to keep the distance from the rotational axis fixed. In this manner, the particle is freely falling not towards the earth center but instead towards the equatorial plane, all while maintaining a fixed zonal velocity with $u = 0$.

Use of the relation (11.90) between v and w leads to the meridional acceleration

$$r_{\perp} \dot{v} = -v w \cos \phi - u \sin \phi (u + 2\Omega r_{\perp}) - (r_{\perp} \Omega)^2 \sin \phi = -\sin \phi [v^2 + (l^z/r_{\perp})^2], \quad (11.91)$$

which is negative in the northern hemisphere and positive in the southern; i.e., it is always directed to lower latitudes. Similarly, the vertical momentum equation leads to

$$r (\dot{w} + g_e) = v^2 + (l^z/r_{\perp})^2 > 0, \quad (11.92)$$

which is always positive, meaning the vertical acceleration is always dominated by the earth's gravitational acceleration. Finally, the meridional and vertical accelerations are related by

$$\dot{v} \cos \phi + (\dot{w} + g_e) \sin \phi = 0. \quad (11.93)$$

When the particle reaches the equator at $\phi = 0$, both the meridional acceleration and the vertical velocity vanish. In contrast, at the poles where $\phi = \pm\pi/2$, the vertical acceleration is $\dot{w} = -g_e$ whereas the meridional velocity vanishes.

11.11 Exercises

EXERCISE 11.1: NEGATIVE AXIAL ANGULAR MOMENTUM

Recall in Section 11.8 we assume the axial angular momentum is positive, which is geophysically the common situation since axial angular momentum from the solid-body motion is so large relative to motion of geophysical fluids. But let us consider the uncommon case where the particle moves zonally westward at a speed greater than the planetary rotation speed so that

$$\dot{\lambda} + \Omega < 0 \iff u + \Omega r_{\perp} < 0, \quad (11.94)$$

which means the axial angular momentum per mass of the particle is negative

$$l^z = r_{\perp} (u + r_{\perp} \Omega) = -|l^z| < 0. \quad (11.95)$$

- (A) Discuss what happens to $\delta\dot{\lambda}$ for the particle as it is deflected poleward with constant radius while conserving axial angular momentum. Hint: consider the first form of equation (11.39).
- (B) Discuss what happens to δu for the particle as it is deflected poleward with constant radius while conserving axial angular momentum. Discuss the cases where
 - (i) $|l^z| = 0$
 - (ii) $|l^z| < \Omega r_{\perp}^2$.
 - (iii) $|l^z| > \Omega r_{\perp}^2$

Hint: consider the second form of equation (11.39).

- (C) Is fluid particle motion with $u + \Omega r_{\perp} < 0$ relevant for the terrestrial atmosphere and ocean? Why? To help answer this question, what is Ωr_{\perp} for $\phi = \pi/4$ and $r = R_e$? Note, we already provided the result for the equator just after equation (11.36). Compare these speeds to that of a category 5 tropical cyclone (Google it).



Part III

Kinematics of fluid flow

In this part of the book we move from the discrete particle mechanics studied in Part II to now focus on continuous fluid media. We motivate the study of continuous media in Chapter 12 by noting the huge space and time scale separation between molecular motions and the motion of macroscopic fluid elements. Thereafter we develop kinematics for classical non-relativistic fluid flow. Quite conveniently, key kinematic properties of non-rotating flows also hold for the fluid flow on steady rotating planets such as considered in this book. The reason is that steady solid-body rotation does not directly impart any strain to the flow, where strain refers to the relative motion between fluid elements (Chapter 13). Rotation does impart a planetary component to the vorticity of geophysical fluids, with implications for our study of vorticity in Part VI. However, for the purpose of this part of the book we can ignore planetary rotation.

EULERIAN AND LAGRANGIAN REFERENCE FRAMES

The *Eulerian and Lagrangian reference frames* provide dual kinematic descriptions of fluid flows. The Eulerian frame describes fluid motion relative to a frame fixed in the laboratory whereas the Lagrangian frame follows a moving fluid particle. The Eulerian frame is inertial (when the laboratory is not accelerating), whereas the Lagrangian is non-inertial since fluid particles generally accelerate. Having two descriptions of the same motion provides a synergy that is missing with either alone, thus exemplifying a maxim:

If you can study, describe, solve, or understand a physics problem more than one way, then do so!

We require skills to move between the Eulerian and Lagrangian descriptions, with tools from mathematical transformation theory of Part I of great use for this purpose. Elements of Eulerian and Lagrangian kinematics are the focus of Chapter 13, and Chapter 14 further develops the kinematics of material lines, areas, and volumes.

MASS AND MATTER CONSERVATION ARE PART OF FLUID KINEMATICS

The conservation of mass plays a central role in physics. For fluids, mass conservation constrains the flow regardless what forces act on the fluid. Hence, we include mass conservation as a part of fluid kinematics rather than fluid dynamics. Mass conservation, and its expression as volume conservation for incompressible flows, are the topics of Chapters 15 and 17. Chapter 16 develops the allied study of matter conservation and matter flow, with this study forming the foundations for *tracer mechanics* that form much of Part IX in this book.

FLUID KINEMATICS + FLUID DYNAMICS = FLUID MECHANICS

Kinematics is concerned with the intrinsic properties of motion, including properties of the space and time in which motion occurs. It is the complement to dynamics, which is concerned with the causes of motion that arise through the action of forces. In one sense, kinematics deduces the acceleration whereas dynamics deduces the forces, with Newton's second law linking the two via the equation of motion: $\mathbf{F} = m \mathbf{a}$. In fluid mechanics, kinematics studies the flow of a fluid and its matter constituents, whereas dynamics studies the forces causing the motion. Furthermore, as discussed in Chapter 11, symmetries of a mechanical system lead, through Noether's Theorem, to dynamical conservation laws. That is, symmetries, which embody kinematic properties, lead to dynamical invariants maintained by the motion, with these invariants constraining the motion. The intellectual avenues pursued in developing a mechanical description of fluid motion are many and varied, with fluid kinematics and fluid dynamics intimately woven into the fabric of that description.

Continuum approximation

Viewed macroscopically, a fluid deforms continuously when applying a force so that a fluid has no preferred shape. Correspondingly, a fluid responds to a shearing stress by moving. Ordinary gases and liquids are canonical examples of fluids, with gases filling any container with its molecules widely separated whereas molecules in liquids are much closer together so that liquids are far less compressible than gases.

For geophysical fluid mechanics, we are concerned with the atmosphere (mostly a gas) and the ocean (mostly a liquid). We are furthermore interested in macroscopic properties of fluid motion, with no interest in describing molecular degrees of freedom. Nor do we consider rarefied gas dynamics, which is a subject appropriate for the upper bounds of the atmosphere. For these reasons we pursue a phenomenological approach that makes use of conservation laws describing the motion of a continuous fluid media. This treatment is based on the *continuum approximation*, which assumes that mathematical limits for fluid volumes tending to zero are reached on length and time scales very large compared to molecular scales.¹ The temporal version of the continuum approximation corresponds to *quasi-static processes*, which forms the basis for quasi-equilibrium thermodynamics. Quasi-static processes refer to macroscopic motion that evolves with time scales far larger than time scales of molecular motions, so that when treated thermodynamically a fluid element evolves smoothly from one local thermodynamic equilibrium state to another.²

The huge space and time scale separation that supports the continuum approximation allows us to make use of calculus for describing the mechanics of fluid motion. That is, the continuum approximation makes fluid mechanics a continuous field theory. Correspondingly, the conservation laws describing fluid motion are partial differential equations.

READER'S GUIDE TO THIS CHAPTER

This chapter presents salient points supporting our use of a space and time continuous description of fluid mechanics. Section 12.2 summarizes the key results and Section 12.3 provides a bit more detail by quoting from kinetic theory. Our overall goal here is to unpack the dictum *macroscopically small yet microscopically large*, which summarizes the regime considered by the continuum approximation and the quasi-static approximation. For this purpose, we borrow from the kinetic theory of gases as treated in statistical physics books such as [Reif \(1965\)](#) and [Huang \(1987\)](#). Chapter 1 of [Salmon \(1998\)](#) also provides a compelling discussion with application to

¹The continuum approximation used for continuum mechanics is not to be confused with the *continuum hypothesis* in mathematical set theory.

²We study equilibrium thermodynamics in Chapter 21.

geophysical fluid mechanics. No prior exposure to these treatments is necessary nor do we dive into the details.

12.1	Loose ends	232
12.2	A variety of length scales	232
12.2.1	Molecular and macroscopic length scales	232
12.2.2	Continuous fields rather than discrete molecules	233
12.2.3	Reynolds number and the continuum length scale	234
12.2.4	Resolution of measurements and simulations	235
12.2.5	Comments	236
12.3	Results from kinetic theory	236
12.3.1	A mole and Avogadro's number	236
12.3.2	Ideal gas law	236
12.3.3	Molecular mean free path	237
12.3.4	Root mean square molecular speed	238
12.3.5	Basis for the quasi-static approximation	238
12.3.6	Macroscopically small and microscopically large	239
12.3.7	Whence a rigorous treatment?	239
12.3.8	Further study	240

12.1 Loose ends

- Add a figure for Section 12.2.4.
- Can one be more precise about the scale in Figure 12.2 where measurements become fuzzy?

12.2 A variety of length scales

Matter is comprised of molecules. However, fluid mechanics is not concerned with the motion of individual molecular degrees of freedom. Rather, fluid mechanics is concerned with phenomenological conservation laws satisfied by a continuous fluid material. This approach represents an idealization that is supported by centuries of successful descriptions of macroscopic fluid motion.

We here outline the essential features of the continuum approximation. This approximation supports our macroscopic description of a fluid in terms of continuous fields rather than discrete molecules. More details are offered in Section 12.3, although a full discussion is outside the subject of fluid mechanics, instead resting deep within the field of statistical physics and kinetic theory.

12.2.1 Molecular and macroscopic length scales

A fluid mechanical description focuses on fluid regions that are macroscopically small (e.g., $L_{\text{macro}} \sim 10^{-3}$ m) yet microscopically large (e.g., $L_{\text{macro}} \gg L_{\text{mfp}} \sim 10^{-7}$ m, where L_{mfp} is the molecular mean free path). A region of air with volume L_{macro}^3 contains roughly 10^{16} air molecules at standard temperature ($T_{\text{stand}} = 0^\circ\text{C} = 273.15$ K) and standard atmospheric pressure ($p_{\text{stand}} = 101.325 \times 10^3$ Pa), whereas that same volume contains roughly 10^{19} water molecules. These numbers (justified in Section 12.3) illustrate the notions of macroscopically small yet microscopically large. It is only when reaching length scales on the order of the molecular mean free path that we need to be concerned with the discrete nature of matter. Figure 12.1 offers a schematic to illustrate these quite distinct length scales.

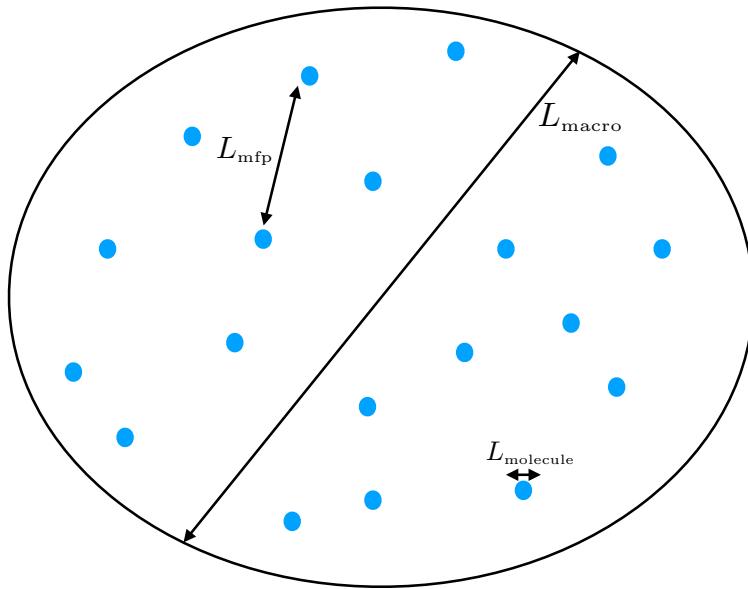


FIGURE 12.1: Schematic to illustrate the three length scales considered when making the continuum approximation. The blue circles represent molecules with diameter L_{molecule} . On average, molecules are separated by the mean free path, $L_{\text{mfp}} \approx 1000 L_{\text{molecule}}$. The smallest macroscopic length scale of interest for fluid mechanics is $L_{\text{macro}} \sim 10^{-3} \text{ m}$, which is roughly $L_{\text{macro}} = 10^4 L_{\text{mfp}}$ for an ideal gas at standard conditions. A region of air with volume L_{macro}^3 contains roughly 10^{16} air molecules, whereas that same volume contains roughly 10^{19} water molecules. For either case, the Law of Large Numbers greatly helps in taking the continuum limit. Note that this schematic is not drawn to scale!

The huge number of molecules within a macroscopically tiny region justifies our assumption that physical properties are homogeneous over regions of size L_{macro} . In essence, this *continuum approximation* works with small but finite sized fluid elements whose mean dynamical properties (e.g., velocity, vorticity) and thermodynamical properties (e.g., mass density, matter concentration, temperature, specific entropy) are defined locally at any point within the continuous fluid media.

12.2.2 Continuous fields rather than discrete molecules

When measured on length scales of the mean free path, material properties exhibit very large fluctuations on time scales of order $L_{\text{mfp}}/v_{\text{rms}}$, where v_{rms} is the root-mean-square speed of a fluid molecule (see Section 12.3.4). However, on macroscopic scales encompassing many molecular degrees of freedom, fluid matter appears continuous in both space and time.

Let us be a bit more precise by considering the measurement of mass density for a prescribed region of fluid, δV . To compute the mass density we simply take the ratio of the mass of fluid in the region, δm , to the region volume. When the region volume is sizable and thus containing many molecules, we can maintain a relatively fixed mass since molecular fluctuations have a relatively tiny effect on δm . Hence, we can maintain a precise measurement of the mass density, $\delta m/\delta V$. However, when the volume of the region reaches down to the order of $\delta V \sim L_{\text{mfp}}^3$, then molecular fluctuations lead to a relatively large fluctuation in the region mass. We thus lose the notion of a smooth and continuous mass density when the volume approaches that set by the molecular mean free path. This situation is depicted in Figure 12.2.

The ratio of the mean free path to the macroscopic length scale is known as the Knudsen number

$$\text{Kn} = \frac{L_{\text{mfp}}}{L_{\text{macro}}}. \quad (12.1)$$

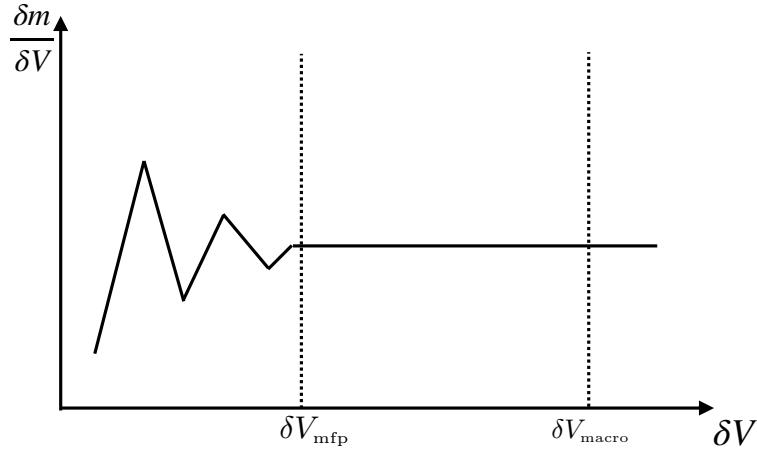


FIGURE 12.2: The measurement of mass density of a fluid becomes erratic for volumes smaller than that determined by the molecular mean free path, $\delta V_{\text{mfp}} \sim L_{\text{mfp}}^3$. For fluid mechanics we are concerned with macroscopic length scales much larger than the mean free path, $L_{\text{macro}} \gg L_{\text{mfp}}$, in which case the density is a smooth function of space and time. This figure is adapted from Figure 1.2.1 of [Batchelor \(1967\)](#).

For our purposes, we are concerned only with fluid conditions where the mean free path is microscopic so that the Knudsen number is tiny

$$\text{Kn} \ll 1. \quad (12.2)$$

For tiny Knudsen numbers, we are led to make use of the continuum approximation. The continuum approximation allows us to employ fluid properties that take values at each point within a space and time continuum, (\mathbf{x}, t) . For example, we make use of the mass density, $\rho(\mathbf{x}, t)$, fluid velocity, $\mathbf{v}(\mathbf{x}, t)$, pressure $p(\mathbf{x}, t)$, temperature $T(\mathbf{x}, t)$, tracer concentration, $C(\mathbf{x}, t)$, and other thermodynamic fields.

We contrast the above to the study of a rarefied gas, such as in the outer reaches of the earth's atmosphere. With a relatively small number density of molecules, rarefied gases have macroscopic mean free paths so that there are relatively few molecular collisions in a given time increment. Correspondingly, a rarefied gas is far from thermodynamic equilibrium and the continuum approximation is not well suited to its description.

12.2.3 Reynolds number and the continuum length scale

The continuum field equations of fluid mechanics are formally established for fluid motions with length scales on the order of L_{macro} and larger. We stated earlier that L_{macro} is on the order of a millimetre, with that length loosely based on noting that there are a huge number of molecules in a volume with this size. Furthermore, most macroscopic measurements in a fluid cannot distinguish features much smaller than a millimetre. We here describe another means to determine this length scale.

Namely, we set L_{macro} to the length scale at which the Reynolds number is order unity

$$\text{Re}_{\text{macro}} = \frac{U L_{\text{macro}}}{\nu} \sim 1. \quad (12.3)$$

In this equation, $\nu > 0$ is the kinematic viscosity (dimensions squared length per time), and U is the scale for a macroscopic fluid velocity fluctuation. The Reynolds number measures the ratio of inertial accelerations (accelerations felt by fluid elements) to frictional accelerations from viscous forces (forces due to the rubbing of fluid elements against one another). We provide more details

concerning the Reynolds number when studying fluid stresses in Chapter 19. For present purposes, we note that when the Reynolds number on the order of unity, viscous forces play a leading role in the acceleration of the fluid. In particular, it is at this scale that viscous accelerations act to dissipate kinetic energy, with this dissipation an important process in fluid turbulence. We are thus motivated to let the length scale where viscosity is important determine L_{macro} .

The kinematic viscosity is the ratio of the *dynamic viscosity* and the mass density. For air, the kinematic viscosity is (page 75 of [Gill \(1982\)](#))

$$\nu_{\text{air}} = \frac{1.7 \times 10^{-5} \text{ kg m}^{-1} \text{ s}^{-1}}{1.3 \text{ kg m}^{-3}} = 1.3 \times 10^{-5} \text{ m}^2 \text{ s}^{-1}, \quad (12.4)$$

and a typical fluid velocity fluctuation has a scale 10^{-1} m s^{-1} , so that

$$L_{\text{macro}} \approx 10^{-4} \text{ m} = 0.1 \text{ mm}. \quad (12.5)$$

Water has a kinematic viscosity (page 75 of [Gill \(1982\)](#))

$$\nu_{\text{water}} = \frac{10^{-3} \text{ kg m}^{-1} \text{ s}^{-1}}{1000 \text{ kg m}^{-3}} \approx 10^{-6} \text{ m}^2 \text{ s}^{-1}, \quad (12.6)$$

and a fluid velocity fluctuation about 10 times smaller than air. Hence, the macroscopic length scale for water is on the order of that for air, both of which are roughly 10^{-4} m . The Reynolds number based macroscale length is not too far from the millimetre scale proposed earlier. We are thus further compelled to consider the macroscopic length scale to be on the order $10^{-4} - 10^{-3} \text{ m}$.

12.2.4 Resolution of measurements and simulations

When we measure fluid motions in the laboratory or field, we generally do not measure the motions at scales on the order of L_{macro} . That is, our measurement devices generally have a spatial resolution much coarser than L_{macro} , so that $L_{\text{measure}} \gg L_{\text{macro}}$. Likewise, numerical simulations are generally made with discrete grids with length scales $L_{\text{numerical}} \gg L_{\text{macro}}$. The equations describing motions at the measurement/simulation length scales involve effects from coupled and nonlinear fluctuations occurring at the smaller (unmeasured) scales. These fluctuations, generally associated with turbulent or chaotic motions, have statistical correlations that can play a role, sometimes a dominant role, in the evolution of flow features at the measured/simulated scales. The parameterization of these correlations in terms of measured/simulated motions constitutes the *turbulence closure* problem. We do not study turbulence closure in this book though we do identify its implications at certain points.

It is important to acknowledge the limited ability of macroscopic measurements to accurately characterize fine scale motions. For this purpose define a gradient length scale

$$L_{\text{gradient}} = \frac{|\mathbf{v}|}{|\nabla \mathbf{v}|}, \quad (12.7)$$

where \mathbf{v} is the velocity of a fluid element and $|\nabla \mathbf{v}|$ is the magnitude of velocity gradients. Decomposing fluctuations into Fourier modes allows us to see that an accurate measurement of velocity fluctuations with length scales L_{gradient} requires a measurement length scale that satisfies

$$2\pi L_{\text{measure}} \leq L_{\text{gradient}}. \quad (12.8)$$

This constraint means that to measure velocity fluctuations on a scale L_{gradient} requires a finer measurement sampling with $L_{\text{measure}} = L_{\text{gradient}}/(2\pi)$.

12.2.5 Comments

The above discussion of length scales transfers seamlessly over to time scales through dividing the length scale by the velocity scale. Correspondingly, fluctuations with time scales shorter than $2\pi T_{\text{measure}}$ cannot be accurately measured.

12.3 Results from kinetic theory

If the reader is content to accept the continuum approximation on face value, then the material in this section can be readily skipped. For others, this section outlines results from the kinetic theory of ideal gases in support of the continuum approximation. Deductive treatments that transition from molecular mechanics to macroscopic fluid mechanics is a topic of the kinetic theory of gases and liquids, which is well outside our scope. In Section 12.3.8, we provide literature pointers for those wishing more rigor.

12.3.1 A mole and Avogadro's number

There are a tremendous number of molecules in the tiniest drop of water or puff of air. Just how many? To answer this question, we introduce the notion of a mole of matter. A mole is defined as the mass of a material substance that contains Avogadro's number of that substance, where

$$A_v = 6.022 \times 10^{23} \text{ mole}^{-1}. \quad (12.9)$$

Avogadro's number, A_v , is the proportionality constant converting from one molar mass of a substance to the mass of a substance. Avogadro's number is conventionally specified so that one mole of the carbon isotope ^{12}C contains exactly 12 grams. Hence, 12 grams of ^{12}C contains 6.022×10^{23} atoms of ^{12}C . Avogadro's number provides a connection between scales active in the microscopic world of molecules to the macroscopic world of everyday experience.

Dry air (air with no water vapor) is comprised of oxygen molecules O_2 , at roughly 22% by molecular mass, and nitrogen molecules N_2 , at roughly 78% molecular mass.³ The molar mass of dry air is thus

$$M_{\text{air}} = 0.22 * 32 \text{ g mole}^{-1} + 0.78 * 28 \text{ g mole}^{-1} \approx 28.8 \text{ g mole}^{-1}. \quad (12.10)$$

Pure (fresh) water is comprised of two hydrogen atoms and one oxygen atom. The molar mass of pure water is thus given by

$$M_{\text{water}} = 2 * 1 \text{ g mole}^{-1} + 16 \text{ g mole}^{-1} = 18 \text{ g mole}^{-1}. \quad (12.11)$$

12.3.2 Ideal gas law

The ideal gas law is given by

$$pV = nR_gT, \quad (12.12)$$

where p is the pressure, V is the volume, n is the number of moles, R_g is the universal gas constant,⁴ and T is the absolute or thermodynamic temperature (temperature relative to absolute zero). Measuring the temperature in Kelvin leads to the universal gas constant

$$R_g = 8.314 \text{ J mole}^{-1} \text{ K}^{-1} = 8.314 \text{ kg m}^2 \text{ s}^{-2} \text{ mole}^{-1} \text{ K}^{-1}, \quad (12.13)$$

³We here ignore the presence of other trace gases, such as CO_2 and H_2O , although these gases are critical for understanding atmospheric radiation and hence the earth's energy budget.

⁴We write R_g rather than the more conventional R to distinguish from R commonly used in this book for the radius of a spherical earth.

where the second equality replaced the energy unit, Joule, by its MKS equivalent,

$$J = \text{kg m}^2 \text{s}^{-2}. \quad (12.14)$$

Use of the ideal gas law (12.12) says that one mole of ideal gas at standard temperature ($T_{\text{stand}} = 0^\circ\text{C} = 273.15 \text{ K}$) and standard atmospheric pressure ($p_{\text{stand}} = 101.325 \times 10^3 \text{ Pa}$) occupies the following volume

$$V = \frac{n R_g T_{\text{stand}}}{p_{\text{stand}}} \quad (12.15a)$$

$$= \frac{(1 \text{ mole}) (8.314 \text{ kg m}^2 \text{s}^{-2} \text{mole}^{-1} \text{K}^{-1}) (273.15 \text{ K})}{101.325 \times 10^3 \text{ kg m}^{-1} \text{s}^{-2}} \quad (12.15b)$$

$$\approx 2.25 \times 10^{-2} \text{ m}^3, \quad (12.15c)$$

where we introduced the MKS units for pressure (force per unit area)

$$\text{Pa} = \text{N m}^{-2} = \text{kg m}^{-1} \text{s}^{-2}. \quad (12.16)$$

Hence, the number density (number of molecules) for a mole of ideal gas is given by

$$n_{\text{gas}} = \frac{\text{number per mole}}{\text{volume per mole}} \quad (12.17a)$$

$$= \frac{A_v}{V} \quad (12.17b)$$

$$= \frac{6.022 \times 10^{23}}{2.25 \times 10^{-2} \text{ m}^3} \quad (12.17c)$$

$$= 2.68 \times 10^{25} \text{ m}^{-3}. \quad (12.17d)$$

Specializing to air, we compute the mass density of air at standard temperature and pressure as

$$\rho_{\text{air}} = \frac{M_{\text{air}}}{V} = \frac{28.8 \times 10^{-3} \text{ kg}}{2.25 \times 10^{-2} \text{ m}^3} = 1.28 \text{ kg m}^{-3}, \quad (12.18)$$

where we set $M_{\text{air}} = 28.8 \times 10^{-3} \text{ kg}$ according to equation (12.10). This ideal gas density is close to the 1.225 kg m^{-3} density measured for air at standard conditions, thus giving us confidence for using the ideal gas law for dry air. Differences arise from trace constituents in air as well as inter-molecular forces (an ideal gas has no inter-molecular forces).

12.3.3 Molecular mean free path

We are in search of length scales relevant for molecular motion. One length scale is that of the molecule itself. Another is set by the distance between molecular collisions. The molecular mean free path is the mean distance that a molecule travels before colliding with another molecule. Arguments from kinetic theory of gases, applied to an ideal gas, lead to the expression

$$L_{\text{mfp}} = \frac{1}{\pi \sqrt{2} n_{\text{gas}} d^2} \quad (12.19)$$

where d is the diameter of the molecule. The mean diameter of air molecules is roughly

$$d_{\text{molecule air}} \approx 2 \times 10^{-10} \text{ m}. \quad (12.20)$$

Hence, the mean free path for air molecules at standard temperature and pressure is

$$L_{\text{mfp}} = \frac{1}{\pi \sqrt{2} n_{\text{gas}} d_{\text{molecule air}}^2} \quad (12.21\text{a})$$

$$= \frac{1}{\pi \sqrt{2} (2.68 \times 10^{25} \text{ m}^{-3}) (2 \times 10^{-10} \text{ m})^2} \quad (12.21\text{b})$$

$$= 2 \times 10^{-7} \text{ m}. \quad (12.21\text{c})$$

The mean free path for an air molecule is roughly 1000 times larger than the molecular diameter (e.g., Figure 12.1).

12.3.4 Root mean square molecular speed

What is the mean speed for molecules moving through a gas? Again, kinetic theory for ideal gases offers an explicit expression, here written in terms of the pressure and density of the gas

$$v_{\text{rms}} = \sqrt{\frac{3p}{\rho}} = \sqrt{\frac{3R_g T}{M}}. \quad (12.22)$$

Note the direct relation between pressure, temperature, and speed. That is, molecules move faster at higher temperature, and thus impart larger pressure on their surrounding. At standard pressure and temperature, the root-mean-square speed for an air molecule is given by

$$v_{\text{rms}} = \sqrt{\frac{3p_{\text{stand}}}{\rho_{\text{air}}}} \quad (12.23\text{a})$$

$$= \sqrt{\frac{3(101.325 \times 10^3 \text{ kg m}^{-1} \text{ s}^{-2})}{1.28 \text{ kg m}^{-3}}} \quad (12.23\text{b})$$

$$= 487 \text{ m s}^{-1}. \quad (12.23\text{c})$$

To get a sense for the relative scale of this speed, note that the speed of sound in air at standard temperature and pressure is 331 m s^{-1} . So these molecules are moving faster than sound! These speeds are correspondingly much higher than the speeds typical for fluid elements in the atmosphere and ocean.

12.3.5 Basis for the quasi-static approximation

Assuming one collision occurs within a mean free path, and the molecules are moving at the root-mean-square speed, we can estimate the time between collision according to

$$t_{\text{collision}} = \frac{L_{\text{mfp}}}{v_{\text{rms}}} \quad (12.24)$$

The corresponding time for air is given by

$$t_{\text{air}} = \frac{2 \times 10^{-7} \text{ m}}{487 \text{ m s}^{-1}} = 4.1 \times 10^{-10} \text{ s}. \quad (12.25)$$

Inverting this number, we see that there are roughly $t_{\text{air}}^{-1} = 2.5 \times 10^9 \text{ s}^{-1}$ collisions per second.

The huge number of collisions per second means that for all macroscopic processes, including geophysical fluid flow, the dynamical time scales for the macroscopic motion are much larger than

the time scales for molecular equilibration. We are thus led to define a *quasi-static macroscopic process* as one that occurs through a series of thermodynamic equilibrium states. Consequently, we can use equilibrium thermodynamic relations while allowing for time evolution of the macroscopic system. We have more to say on this topic in Chapters 21 and 22 when studying thermodynamics, with particular emphasis on the implications of the quasi-static approximation presented in Section 21.1.3.

12.3.6 Macroscopically small and microscopically large

For environmental measurements of the atmosphere and ocean, or for conventional measurements in laboratories, we can detect differences in fluid properties (e.g., mass density, velocity, tracer concentration, thermodynamic state properties) for length scales on the order of

$$L_{\text{macro}} = 10^{-3} \text{ m.} \quad (12.26)$$

For macroscopic purposes, fluid properties are homogeneous over regions with length scales on the order of L_{macro} . Although macroscopically rather tiny, a fluid region of volume L_{macro}^3 is huge microscopically. We can see so by computing the number of molecules in this region.

At standard conditions, a volume of air of size L_{macro}^3 contains

$$N_{\text{air molecules}} = V n_{\text{gas}} = (10^{-3} \text{ m})^3 (2.68 \times 10^{25} \text{ m}^{-3}) \approx 3 \times 10^{16} \text{ air molecules.} \quad (12.27)$$

To compute the number of water molecules in this same volume, we first use the water mass density of $\rho \approx 10^3 \text{ kg m}^{-3}$ to determine the water mass in this region

$$M_{\text{water}} = \rho_{\text{water}} V = (1000 \text{ kg m}^{-3}) (10^{-9} \text{ m}^3) = 10^{-6} \text{ kg.} \quad (12.28)$$

Water has a molar mass of $0.018 \text{ kg mole}^{-1}$, so a volume of $(10^{-3} \text{ m})^3$ contains⁵

$$N_{\text{water molecules}} = \left(\frac{10^{-6} \text{ kg}}{0.018 \text{ kg mole}^{-1}} \right) \times 6.022 \times 10^{23} \text{ molecules mole}^{-1} = 3 \times 10^{19} \text{ water molecules.} \quad (12.29)$$

Water thus has roughly 10^3 more molecules in this volume than air at standard pressure, which reflects the roughly 10^3 times larger mass density for water. Regardless, both water and air contain a huge number of molecules in this macroscopically tiny region.

12.3.7 Whence a rigorous treatment?

A rigorous derivation of continuum field theory, starting from molecular dynamics, is nontrivial even for an ideal gas, and largely non-existent for liquids. Indeed, some say a Nobel Prize awaits the person providing a fully deductive theory. For our purpose, we remain satisfied to postulate that a continuum description is valid for fluid mechanics of the atmosphere and ocean. A means for evaluating this postulate is to perform experimental measures and compare to the continuum theory. Centuries of experiments with fluid motions in the environment and laboratory lend credence to the continuum description. We consider these tests to be sufficient motivation to pursue the continuum approach for fluid mechanics and geophysical fluid dynamics.

⁵The calculation on page 9 of [Griffies \(2004\)](#) has a factor of 10^6 error.

12.3.8 Further study

Pedagogical treatments of the ideal gas law and kinetic theory can be found in most books on introductory physics or chemistry. [Vallis \(2017\)](#) provides extensions of the ideal gas law for an atmosphere with moisture.

For discussions of the continuum approximation reflecting that given here, see the discussion on page 1 of [Olbers et al. \(2012\)](#), or the more thorough treatments given in Section 1.2 of [Batchelor \(1967\)](#) and Section 1.4 of [Kundu et al. \(2016\)](#). Chapter 1 of [Salmon \(1998\)](#) offers an even more thorough treatment, touching on elements from kinetic theory and details for how to coarse grain average over molecular degrees of freedom (see his pages 3 and 4 and Sections 9, 10, and 11). A rigorous account of kinetic theory is offered in many treatments of statistical mechanics. That given by [Reif \(1965\)](#) and [Huang \(1987\)](#) are accessible to those with a physics undergraduate training. When reading the statistical mechanics literature, look for discussions of the “hydrodynamical limit,” which concerns the transition from discrete particle mechanics to continuum mechanics.



Fundamentals of fluid kinematics

Fluid motion is very complex and experience has shown that it is very useful to have more than one means to describe the motion. For this purpose we study fluid motion when viewed in both the Eulerian and Lagrangian reference frames, where the Eulerian frame is fixed in the laboratory and the Lagrangian frame moves with a fluid particle. These dual descriptions form the foundation for fluid kinematics. Although the Eulerian frame is more familiar to many, and often the only kinematics needed for many purposes, a Lagrangian description offers useful insights into the theoretical foundations of the subject. It does so by providing the natural reference frame to formulate dynamical laws given that it follows seamlessly from the formulation of point particle mechanics discussed in Part II of this book. We thus give attention to the needs of both Eulerian and Lagrangian kinematics in this chapter and elsewhere in this book.

READER'S GUIDE TO THIS CHAPTER

To keep the discussion focused, we assume that spatial positions and fluid particle trajectories are represented using Cartesian coordinates. Even so, we require elements of the tensor transformation theory from Part I to systematically move between the Eulerian and Lagrangian descriptions. We review the salient formalism in this chapter to keep the discussion reasonably self-contained. This is a relatively long chapter that introduces many concepts and tools used in nearly every subsequent chapter of this book.

The key references for this chapter include the text by [Salmon \(1998\)](#), who provides an elegant and accessible treatment of Eulerian and Lagrangian fluid mechanics. Chapter 4 of [Aris \(1962\)](#) offers a lucid treatment of fluid kinematics in the context of tensor analysis. Much of the treatment here follows Chapters 1 and 2 of the ocean fluid mechanics book of [Olbers et al. \(2012\)](#).

13.1 Strong and weak formulations	242
13.2 Conceptually partitioning the continuum	243
13.2.1 Fluid particle	243
13.2.2 Material fluid parcel in perfect fluids	244
13.2.3 Finite sized material objects in perfect fluids	244
13.2.4 Fluid element in real fluids	245
13.2.5 Test fluid element in real fluids	245
13.2.6 Finite sized fluid region in real fluids	245
13.3 Lagrangian and Eulerian reference frames	245
13.4 Material and position coordinates	246
13.4.1 Trajectories of fluid particles	246

13.4.2	Material coordinates	247
13.4.3	Transforming between material and position coordinates	248
13.4.4	Transformation matrix and its Jacobian	248
13.4.5	A discrete algorithm to compute the transformation matrix	249
13.4.6	Jacobian as the ratio of volumes	250
13.4.7	Further study	251
13.5	Lagrangian and Eulerian time derivatives	252
13.5.1	Infinitesimal space-time increment of a function	252
13.5.2	Total time derivative of a function	252
13.5.3	Eulerian: evolution measured in the laboratory frame	252
13.5.4	Lagrangian: evolution measured in the material frame	253
13.5.5	Lagrangian time derivative formulated from the material frame	253
13.5.6	Sample material time derivative operations	254
13.5.7	Worked example: velocity and acceleration from a trajectory	256
13.5.8	Summarizing some terminology	257
13.6	Galilean invariance	257
13.6.1	Galilean transformation	257
13.6.2	Transformation matrix	258
13.6.3	Transforming the differential operators	258
13.6.4	Comments	259
13.7	Invariance of the material time derivative	259
13.7.1	Invariance based on definition of the material time derivative	259
13.7.2	Invariance for a rotating reference frame	260
13.7.3	Invariance using space-time tensors	262
13.7.4	Comments	262
13.8	Fluid flow lines	262
13.8.1	Material pathlines from fluid particle trajectories	262
13.8.2	Fluid streamlines and streamtubes	263
13.8.3	Distinguishing streamlines and pathlines	264
13.8.4	Fluid streaklines	264
13.8.5	An analytic example of flow lines	265
13.8.6	Further study	267
13.9	Exercises	267

13.1 Strong and weak formulations

The continuum approximation (Chapter 12) allows us to consider fluid flow from a field theoretic perspective, whereby physical properties are described by fields that take on values at each point of a space and time continuum. Consequently, we make use of a differential equation formulation of the governing fluid equations as well as an integral formulation. The differential equation formulation is sometimes referred to as the *strong* formulation. This name is motivated by the need to make “strong” assumptions about the smoothness of the continuum fields. Without such smoothness assumptions the differential equations lack predictive skill. Even so, some fluid phenomena (e.g., shocks) do not satisfy the necessary smoothness assumptions, thus making the strong formulation unsuitable for their description. Non-smooth flows are where the integral formulation is essential, with the integral formulation known as the *weak* formulation since it requires fewer assumptions about smoothness.

In this book, we are not concerned with shocks or other discontinuities within the fluid continuum, though we do study discontinuities at boundaries. Consequently, we can make use of both the strong and weak formulations. Smooth fluid flows are afforded a connection between the

weak and strong formulations through the Leibniz-Reynolds transport theorem derived in Section 16.3.4. As we will see, each formulation is suited for particular needs. For example, the strong formulation provides a more concise view of the fluid equations and allows for familiar manipulations/transformations based on the rules of differential calculus studied in Part I. In contrast, the weak formulation is needed to develop budgets over finite fluid regions. Correspondingly, the weak formulation provides a starting point for the derivation of finite volume budgets that serve as the basis for certain numerical methods (e.g., [Griffies et al. \(2020\)](#)).

13.2 Conceptually partitioning the continuum

As part of a continuum description of fluid motion, we make use of conceptual physical systems to frame the mechanics. These physical systems are bounded by imagined partitions that help to organize questions about forces and transport within the continuum, and the partitions can be either open or closed to matter and energy exchange depending on the character of the fluid. There is no pretense that these partitions can be experimentally determined, nor are the fluid systems to be considered discrete molecules or solid objects. Rather, partitions are drawn within the continuum fluid by the theorist for purposes of formulation and conceptualization. We are afforded the ability to draw these partitions through our use of the continuum approach to fluid mechanics.¹

13.2.1 Fluid particle

A *fluid particle* is a zero dimensional point that follows the fluid flow as specified by the velocity field (left panel in Figure 13.1). Since it has zero spatial extent, a fluid particle does not affect the flow. Notably, a fluid particle is not a molecule since even a molecule has nonzero extent.

The position of a fluid particle is uniquely specified by its material coordinate and time (we discuss material coordinates in Section 13.4.1). The trajectory of a fluid particle represents an *integral curve* of the flow field, where each point along a trajectory has its tangent parallel to the velocity vector (Section 13.8). The accumulation of fluid particle trajectories define pathlines that prescribe the Lagrangian reference frame (Section 13.3).

A fluid particle is analogous to *test mass particles* in Newtonian gravitation that are used to map gravitational field lines, and *test electric charges* in electromagnetism used for mapping the electromagnetic field. However, fluid particles are entirely kinematic since they have zero mass and are fully defined kinematically through specifying the velocity field. Fluid particles can be used for both perfect fluids, which necessarily have a single matter constituent, as well as real fluids with multiple matter constituents. For the perfect fluid, the fluid particle traces out integral curves of the velocity field, whereas for a real fluid the fluid particle provides integral curves of the *barycentric velocity* (Section 16.1), which is the center of mass velocity.

Note that some books define fluid particles as finite sized fluid regions, much like the fluid parcel in Section 13.2.2 or fluid element in Section 13.2.4. We instead find it conceptually useful to define a fluid particle as a point particle with zero spatial extent, thus serving solely as a probe for the flow and a means to specify the Lagrangian reference frame. The nonzero size of fluid parcels and fluid elements serves further needs that go beyond that of a point fluid particle.

¹This conceptual formulation of fluid mechanics, namely as a continuous collection of infinitesimally small fluid elements, originates from the work of Leonard Euler (1707-1783). Indeed, as detailed by [Truesdell \(1953\)](#), both the Eulerian and Lagrangian formulation of fluid kinematics originates with Euler!

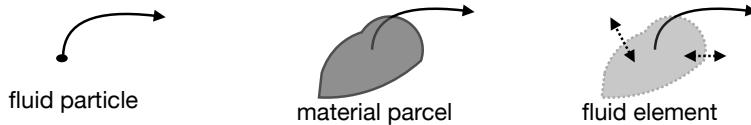


FIGURE 13.1: Schematic of the conceptual systems used in our considerations of fluid kinematics. Left panel: fluid particle that has zero extent and follows the fluid flow field, \mathbf{v} , without altering it. The trajectories of fluid particles are integral curves that define the Lagrangian reference frame. Middle panel: a material fluid parcel comprised of a fixed material content (and thus a fixed mass) and fixed thermodynamic properties. Fluid parcels are tiny deformable regions of a perfect fluid that move with the fluid velocity. Right panel: a fluid element is comprised of a fixed mass but with matter and thermodynamic properties exchanged across its boundary. The fluid element moves with the barycentric velocity (see Section 16.1), which is the center of mass velocity.

13.2.2 Material fluid parcel in perfect fluids

For many purposes we find it useful to study fluid mechanics in the absence of irreversible processes such as friction, heat exchange, and diffusive mixing. In this case the fluid is referred to as an *ideal fluid* or a *perfect fluid*. More precisely, perfect fluid mechanics is concerned with motion of a homogeneous fluid (e.g., pure water) with zero viscosity (no friction), and in the absence of any heat exchange (adiabatic).²

In describing perfect fluids we follow Euler by making use of *material fluid parcels*, which are infinitesimal deformable fluid regions that follow the fluid flow as specified by the velocity field (middle panel in Figure 13.1). A material fluid parcel maintains a fixed matter content so that it has fixed mass. Furthermore, it does not experience irreversible exchanges of momentum arising from friction since the perfect fluid has zero viscosity. Hence, its only interaction with the surrounding fluid environment is through reversible mechanical exchanges from pressure. The material fluid parcel is thus a closed thermodynamic system that is open to reversible mechanical interactions.

Just like a fluid particle, a material fluid parcel is uniquely specified by its material coordinate and time. However, a material parcel is not a point. Rather, it has an infinitesimal volume that deforms with the flow. Conceptually we can imagine the material fluid parcel as a tiny region of fluid surrounded by a perfectly slippery bag that is also perfectly insulating. This bag is closed to matter exchange but it deforms in response to the pressure of the surrounding fluid. Additionally, the bag expands or contracts according to the density of the fluid within the bag. This conceptual picture is qualified by noting that we never have occasion to precisely specify the boundary of a material fluid parcel. Rather, we make use of the conceptual framework provided by fluid parcels as a means to formulate the differential equations of perfect fluid mechanics.

13.2.3 Finite sized material objects in perfect fluids

Any extended region in a perfect fluid, either infinitesimal or finite, open or closed, remains exactly coherent (fixed matter content) as the region moves through the fluid. The reason for such coherency is that a perfect fluid supports no mixing or other irreversible processes that would act to diffuse the matter content. A closed material region in a perfect fluid is a finite volume generalization of a material fluid parcel. Conversely, a material fluid parcel is the infinitesimal limit of a closed material fluid region. Likewise, we can define finite sized material regions of any shape, each of which retains a fixed mass and fixed matter content as it moves through a perfect fluid. We study the kinematics of perfect fluid material lines, surfaces, and volumes in Chapter 14.

²We study frictional stresses arising from viscosity in Chapter 19, with viscosity the means for the irreversible transfer of momentum within a non-perfect fluid. We study enthalpy in the Chapter 21, with a perfect fluid maintaining constant enthalpy since it is entirely adiabatic.

13.2.4 Fluid element in real fluids

A fluid element is an infinitesimal and deformable fluid region of fixed mass yet non-fixed matter and non-fixed thermodynamic properties (right panel in Figure 13.1). For a homogeneous fluid comprised of a single matter constituent and no irreversible processes, then a fluid element reduces to a material fluid parcel. However, there is a distinction for real fluids such as the ocean and atmosphere, both of which have multiple constituents and support irreversible processes.

The exchange of matter across the boundary of a fluid element arises from the irreversible mixing of matter constituents within the fluid (Sections 16.1 and 48.3). As detailed in Section 16.1, diffusive matter exchange leaves the mass of the fluid element unchanged since the fluid element velocity is determined by its center of mass (*barycentric velocity*). Just as for a material fluid parcel, we have no need to experimentally specify the boundary of a fluid element. Instead, fluid elements are conceptual systems used to formulate the differential equations of a real fluid. Much of the kinematics in the current chapter holds for both material fluid parcels and fluid elements. However, in Chapter 15 and elsewhere, we make the distinction when studying the kinematics of multi-constituent fluids.

Many authors do not distinguish between material fluid parcels and fluid elements, choosing instead to retain a single overloaded term for both a perfect fluid and real fluid. We have found this approach can lead to confusion, thus motivating us to make the distinction between parcel (perfect fluid) and element (real fluid).³ The distinction offers an added signal for when the fluid under study is perfect or real.

13.2.5 Test fluid element in real fluids

A *test fluid element* is a fluid element that has no effect on the surrounding fluid environment. Rather, it is used as a conceptual probe of the fluid much like the fluid particle in Section 13.2.1. Unlike the fluid particle, the test fluid element has extent and it can exchange matter and energy with its surrounding environment. The test fluid element is of particular use when studying buoyancy in Chapter 25.

13.2.6 Finite sized fluid region in real fluids

A region within a real fluid is the most general subsystem we consider, with the region having boundaries that are open to the exchange of matter, mechanical forces, and thermodynamic properties with the surrounding environment. Here, we are often concerned with details of the region boundary and study the transport of properties across that boundary.

13.3 Lagrangian and Eulerian reference frames

There are two reference frames commonly used as the basis for describing motion of a fluid continuum. These reference frames retain a 1-to-1 relation that allows for the transformation of descriptions between the frames.

- **LAGRANGIAN OR MATERIAL REFERENCE FRAME:** This reference frame is defined by that of fluid particles (Section 13.2.1). A mechanical description in this reference frame aims to determine the trajectory for each fluid element. The material approach is commonly termed *Lagrangian*. We note that since fluid particles generally experience accelerations (i.e., they change their speed and/or direction), the Lagrangian frame is a non-inertial reference frame.

³We were inspired to introduce the term “fluid element” from page 3 of [Olbers et al. \(2012\)](#).

- **EULERIAN OR LABORATORY REFERENCE FRAME:** The second reference frame is based on observing the fluid from a fixed spatial position, \mathbf{x} , within a “laboratory”. This *Eulerian* approach measures fluid properties as the fluid streams by a fixed observer. It is not concerned with determining fluid particle trajectories. Instead, the focus of Eulerian kinematics is on fluid properties determined as a function of position \mathbf{x} and time t .

The Eulerian and Lagrangian approaches complement one another. For example, the Lagrangian approach lends itself to fruitful physical insights since we can borrow freely from the point particle mechanics of Part II. In contrast, the Eulerian approach is often more straightforward when developing numerical methods for simulations and it is typically simpler when making laboratory or field measurements. In general, we make use of both Eulerian and Lagrangian kinematics. A goal of this chapter is to provide the foundation for these two perspectives and to develop tools for transforming from one to the other.

In classical fluid mechanics, the laboratory frame is fixed in space and thus it represents an inertial reference frame, which contrasts to the non-inertial Lagrangian frame. However, for geophysical fluid mechanics we generally consider an Eulerian reference frame fixed with respect to the rotating planet (a rotating laboratory frame), and as such the laboratory frame is not inertial. However, the discussion in this chapter is not concerned with such non-inertial features that give rise to planetary centrifugal and Coriolis accelerations. Instead, since the constant rotation of the planet does not impart any new strain to the fluid, classical fluid kinematics is sufficient for geophysical fluid kinematics.

13.4 Material and position coordinates

A material description is suggested by the Lagrangian reference frame, whereby fluid particles are labeled with a material coordinate, \mathbf{a} , and sometimes referred to as the \mathbf{a} -space description. This description complements the Eulerian or \mathbf{x} -space description, whereby each point in space is labeled by its position. The trajectory of a fluid particle provides the mathematical transformation between the material coordinates and position coordinates. We explore this transformation in this section.

13.4.1 Trajectories of fluid particles

In describing the motion of a classical point particle (Chapter 10), we specify its spatial position according to a time dependent position vector, \mathbf{X} , that is a function of time, t . At a given time, t , the position vector is located at a space point denoted by the Cartesian position, \mathbf{x} , in which case we write

$$\mathbf{x} = \mathbf{X}(t) \quad \text{point particle.} \quad (13.1)$$

A sample trajectory is shown in Figure 13.2. We emphasize the notation convention used here, which may seem pedantic but in later discussions proves essential. Namely, the time dependent position of a particle is denoted with the capital, $\mathbf{X}(t)$, whose instantaneous space position is denoted by the lowercase, \mathbf{x} . This convention aims to distinguish time dependent functions, such as $\mathbf{X}(t)$, from the value of these functions evaluated at a point in time, \mathbf{x} .

When there are N discrete particles, we distinguish the various particle trajectories by introducing a discrete label. The position of particle n at time t is thus written

$$\mathbf{x} = \mathbf{X}(n, t). \quad (13.2)$$

When the matter is a continuum, such as for a rigid body, an elastic solid, or a fluid, then the discrete label becomes a continuous vector, \mathbf{a} , which is referred to as the *material* coordinate. At

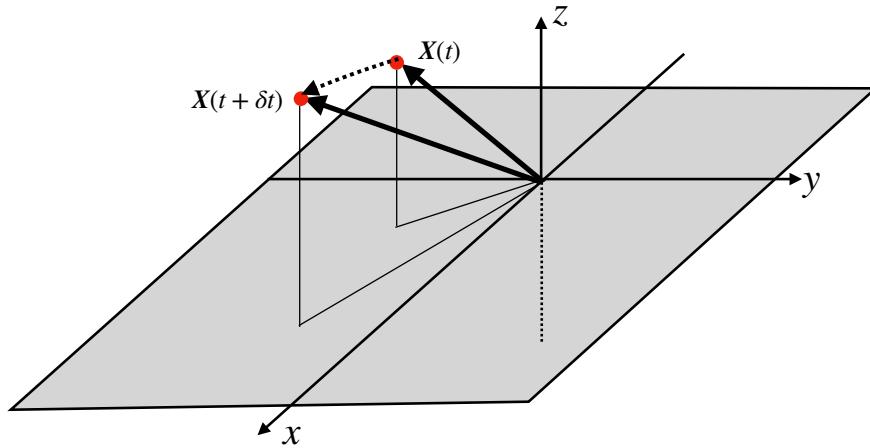


FIGURE 13.2: Sample trajectory of a fluid particle with endpoints $\mathbf{X}(t)$ and $\mathbf{X}(t + \delta t)$. The trajectory passes through the point $\mathbf{x} = \mathbf{X}(t)$ at time t and $\mathbf{x} + \delta\mathbf{x} = \mathbf{X}(t + \delta t)$ at time $t + \delta t$. Eulerian kinematics describes the fluid flow from the perspective of an observer fixed with respect to the laboratory frame. Lagrangian kinematics describes the fluid flow from the perspective of an observer moving in the frame of fluid particles.

time t , the position of a fluid element labelled by the material coordinate, \mathbf{a} , is written

$$\mathbf{x} = \mathbf{X}(\mathbf{a}, t) \quad \text{continuum of matter.} \quad (13.3)$$

The continuous vector, \mathbf{a} , labels a point of matter within the continuum fluid.

13.4.2 Material coordinates

In classical fluid mechanics, we ignore special relativistic effects of fluid particles. Consequently, both the material reference frame and the laboratory reference frame measure the same universal Newtonian time, t . In contrast, the spatial coordinates are distinct for the Eulerian and Lagrangian references frames. Again, the spatial coordinates for the Eulerian frame are given by the position relative to a fixed laboratory frame, whereas the three components of a material Lagrangian coordinate remains unchanged as the fluid particle moves. Additionally, the three coordinates for both the Eulerian and Lagrangian description must be linearly independent to allow for a unique specification of the fluid particle.

One common choice for material coordinate is to define it as the spatial position of a fluid particle at an arbitrary time

$$\mathbf{a} = \mathbf{X}(t = t_0). \quad (13.4)$$

As a slightly more refined example, consider a perfect fluid (single material component with no irreversible processes). For this fluid, the specific entropy of each fluid parcel remains fixed at its initial value. When the fluid is placed in a gravitational field, layers of constant specific entropy are generally found to be monotonically stacked, or *stratified*, in the vertical direction (Chapter 25). As a result, we can uniquely specify a fluid parcel by giving its horizontal coordinate position, (x, y) , as well as the specific entropy. The material coordinates for a parcel can thus be written as

$$\mathbf{a} = (X, Y, \theta)_{t=t_0}, \quad (13.5)$$

where we write θ as a measure of the specific entropy (or potential temperature as discussed in Section 21.8). As indicated by this example, the physical dimensions of material coordinates can generally differ. It is this generality that necessitates the use of general tensor methods when

developing the mechanical equations using arbitrary Lagrangian coordinates. The mathematics and physics of these *generalized vertical coordinates* are detailed in Chapters 43, 44, and 45. For the current chapter we assume the Lagrangian coordinates are Cartesian, as given by the initial Cartesian positions as in equation (13.4).

13.4.3 Transforming between material and position coordinates

Motion of a fluid continuum is described by a *point transformation*. A point transformation is a mathematical way of saying that the fluid motion causes a fluid particle labeled by \mathbf{a} to continuously move from an initial position, $\mathbf{x}_0 = \mathbf{X}(\mathbf{a}, t_0)$, to another position, $\mathbf{x} = \mathbf{X}(\mathbf{a}, t)$, at time $t > t_0$. The point transformation is defined by the vector relation

$$\mathbf{x} = \mathbf{X}(\mathbf{a}, t), \quad (13.6)$$

which is written in component form as⁴

$$x^m = X^m(a^i, t). \quad (13.7)$$

The relation (13.6) defines a single-valued and invertible transformation from material coordinates (\mathbf{a}, t) to position coordinates (\mathbf{x}, t) . That is, for each material coordinate \mathbf{a} and time t , there is a unique spatial point \mathbf{x} , with this point specified by the trajectory $\mathbf{X}(\mathbf{a}, t)$. Conversely, for each space-time point (\mathbf{x}, t) there exists a unique material coordinate found by inverting equation (13.6)⁵

$$\mathbf{a} = \mathbf{A}(\mathbf{x}, t). \quad (13.8)$$

In this equation, \mathbf{A} is the inverse function that specifies the material coordinate \mathbf{a} given (\mathbf{x}, t) . The single-valued property means that a fluid particle trajectory does not split, nor do two trajectories occupy the same point at the same time. This assumption is fundamental to the continuum approximation and the associated use of fluid particles to map out pathlines of the fluid flow. We acknowledge that fluid particle trajectories generally become increasingly complex in turbulent flow, thus making the Lagrangian description less convenient. However, so long as trajectories do not split or merge, the trajectories are well defined, as is the corresponding Lagrangian formulation.

13.4.4 Transformation matrix and its Jacobian

In our analysis of fluid motions, we make routine use of the position coordinates (\mathbf{x} – *space*) of an Eulerian description and material coordinates (\mathbf{a} – *space*) of a Lagrangian description. We here introduce the transformation matrix allowing us to transform from one description to the other. Recall that in Section 8.2.1 we encountered the transformation matrix between Cartesian and spherical coordinates. We will later consider a transformation matrix for moving between different references frames in Sections 13.6 and 13.7. Here, we develop the transformation matrix between \mathbf{x} -space and \mathbf{a} -space.

The fluid particle trajectories, $\mathbf{X}(\mathbf{a}, t)$, are fundamental to this transformation. Namely, the trajectories as given by equation (13.6) provide a transformation between position coordinates, (\mathbf{x}, t) , and material coordinates, (\mathbf{a}, t) . The transformation is invertible so long as the Jacobian (determinant) of the transformation matrix remains nonzero.

⁴We choose tensor labels m, n, p for spatial coordinates and trajectories, and i, j, k for material coordinates.

⁵The use of \mathbf{A} for the inverse function in equation (13.8) should not be confused with the acceleration, also written as \mathbf{A} elsewhere in this book. We will not have much use for equation (13.8), thus minimizing the opportunity for confusion.

The transformation matrix is given by the matrix of partial derivatives, and we organize this matrix according to the following convention⁶

$$F_i^m = \frac{\partial X^m}{\partial a^i} \equiv \begin{bmatrix} \partial X^1 / \partial a^1 & \partial X^1 / \partial a^2 & \partial X^1 / \partial a^3 \\ \partial X^2 / \partial a^1 & \partial X^2 / \partial a^2 & \partial X^2 / \partial a^3 \\ \partial X^3 / \partial a^1 & \partial X^3 / \partial a^2 & \partial X^3 / \partial a^3 \end{bmatrix}. \quad (13.9)$$

The upper label, m , denotes the row and the lower label, i , is the column. The transformation matrix F_i^m is also known as the *deformation tensor*, as it provides a means to measure how trajectories are deformed by the flow. That is, each element of the tensor measures how much the m 'th component of the trajectory is modified when altering the i 'th component of the material coordinate. We can more easily understand these partial derivatives when the material coordinate is the initial position, $\mathbf{a} = \mathbf{X}(\mathbf{z}, t = 0)$, in which case the partial derivatives measure how trajectories are modified when altering the initial position. If there the fluid has no deformation, then the trajectory remains unaffected if changing the material coordinate (i.e., the initial position), in which case the deformation tensor is the identity tensor

$$F_i^m = \delta_i^m \quad \text{if there is no flow deformation.} \quad (13.10)$$

Generally each component of the deformation tensor is nonzero so that trajectories are deformed by the flow and as such they are dependent on the initial position.⁷

The Jacobian of the transformation matrix is the determinant, which can be written in either of the following manners

$$\det(F_i^m) = \frac{\partial \mathbf{X}}{\partial \mathbf{a}} = \det \begin{bmatrix} \partial X^1 / \partial a^1 & \partial X^1 / \partial a^2 & \partial X^1 / \partial a^3 \\ \partial X^2 / \partial a^1 & \partial X^2 / \partial a^2 & \partial X^2 / \partial a^3 \\ \partial X^3 / \partial a^1 & \partial X^3 / \partial a^2 & \partial X^3 / \partial a^3 \end{bmatrix}. \quad (13.11)$$

We make use of the notation $\partial \mathbf{X} / \partial \mathbf{a}$ for the Jacobian as it offers a useful means to distinguish between the determinant of the transformation (13.9), and the determinant of the inverse transformation, written as $\partial \mathbf{a} / \partial \mathbf{X}$.

13.4.5 A discrete algorithm to compute the transformation matrix

To help further our understanding of the transformation matrix (13.9), we here sketch an algorithm for its discrete approximation. For this purpose, we illustrate an algorithm for two-dimensional flow and write the trajectory using Cartesian coordinates

$$\mathbf{X}(t) = X^1(t) \hat{\mathbf{x}} + X^2(t) \hat{\mathbf{y}}, \quad (13.12)$$

and use a Cartesian representation for the material coordinate

$$\mathbf{a} = a^1 \hat{\mathbf{x}} + a^2 \hat{\mathbf{y}}. \quad (13.13)$$

Now lay down a two-dimensional lattice with discrete indices (e, f) for each of the nodal points (grid points) on the lattice, and with corresponding spatial coordinates

$$\mathbf{x}(e, f) = x(e, f) \hat{\mathbf{x}} + y(e, f) \hat{\mathbf{y}}. \quad (13.14)$$

⁶The use of covariant and contravariant notation in equation (13.9), with the upstairs and downstairs tensor labels, is strictly not needed since we assume in this chapter that both the \mathbf{x} -space and \mathbf{a} -space coordinates are Cartesian. However, the notation is quite helpful in keeping track of the ordering of partial derivatives and the corresponding coordinate transformations.

⁷We have much more to say about the deformation tensor in Section 14.2.

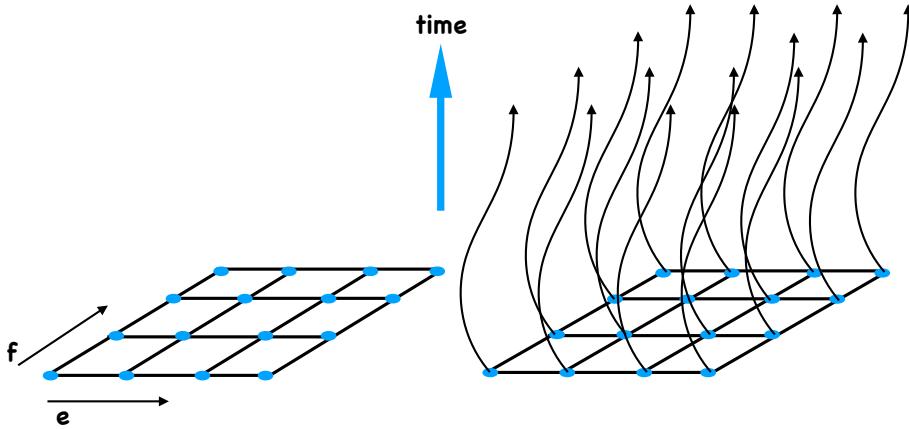


FIGURE 13.3: Illustrating the computational algorithm of Section 13.4.5 used to approximate the transformation matrix (also known as the deformation tensor) F_i^m . The left panel shows the two-dimensional grid with nodal points defining the initial positions for fluid particles. Each position on the grid is labeled by a unique integer (e, f) . The initial position of each particle is taken as the material coordinate, with the discrete label (e, f) maintained by the particles as they evolve. The right panel shows the pathlines for the fluid particles after time $t > 0$. When working on a discrete grid, the position of the fluid particles is not generally at a nodal point. Hence, the position must be found by interpolating between the node points.

Initialize fluid particles at each of the lattice grid points,

$$\mathbf{X}(e, f; t = 0) = \mathbf{x}(e, f) = \mathbf{a}(e, f), \quad (13.15)$$

with the discrete material coordinates defined by the initial positions. Then time step the trajectories using the velocity field to compute the particle pathlines $\mathbf{X}[\mathbf{a}(e, f); t]$ as illustrated in Figure 13.3. At any particular time, the position of a fluid particle is found by interpolating from the lattice grid points. Setting the material coordinates equal to the initial position then leads to the finite difference approximation to the transformation matrix

$$F_i^m = \begin{bmatrix} F_1^1 & F_2^1 \\ F_1^2 & F_2^2 \end{bmatrix} \approx \begin{bmatrix} \frac{X^1(e+1, f; t) - X^1(e-1, f; t)}{X^1(e+1, f; 0) - X^1(e-1, f; 0)} & \frac{X^1(e, f+1; t) - X^1(e, f-1; t)}{X^2(e, f+1; 0) - X^2(e, f-1; 0)} \\ \frac{X^2(e+1, f; t) - X^2(e-1, f; t)}{X^2(e+1, f; 0) - X^2(e-1, f; 0)} & \frac{X^2(e, f+1; t) - X^2(e, f-1; t)}{X^2(e, f+1; 0) - X^2(e, f-1; 0)} \end{bmatrix}. \quad (13.16)$$

If the grid is regular in both directions, then the initial positions have a separation, Δ , given by the grid spacing so that

$$F_i^m \approx \frac{1}{2\Delta} \begin{bmatrix} X^1(e+1, f; t) - X^1(e-1, f; t) & X^1(e, f+1; t) - X^1(e, f-1; t) \\ X^2(e+1, f; t) - X^2(e-1, f; t) & X^2(e, f+1; t) - X^2(e, f-1; t) \end{bmatrix}. \quad (13.17)$$

This expression illustrates how the transformation matrix provides a measure of trajectory spreading as fluid particles move away from their initial positions. As a check on the formulation, consider the case without any deformation. In this case $X^1(e, f; t) = X^1(e; t)$ and $X^2(e, f; t) = X^2(f; t)$ so that the deformation tensor is diagonal, and furthermore, $X^1(e+1; t) - X^1(e-1; t) = 2\Delta$ and $X^2(f+1; t) - X^2(f-1; t) = 2\Delta$, so that the deformation tensor is the identity.

13.4.6 Jacobian as the ratio of volumes

We here establish the Jacobian as the ratio of the coordinate volume of fluid parcels as written using position space coordinates and material space coordinates. This property holds at each point within the continuum fluid so that it also holds on fluid parcels.

Volume of an infinitesimal region of space within the fluid continuum

Consider the volume of an infinitesimal region of space. For simplicity, write this volume in terms of Cartesian coordinates

$$dV(\mathbf{x}) = dx dy dz. \quad (13.18)$$

The \mathbf{x} argument is introduced on the left hand side to distinguish this volume, which has dimensions L^3 , from the volume written in material coordinates

$$dV(\mathbf{a}) = da db dc, \quad (13.19)$$

where $\mathbf{a} = (a^1, a^2, a^3) = (a, b, c)$ are coordinates in material space. Note that $dV(\mathbf{a})$ does not generally have dimension L^3 , since the dimension for each component of the material coordinates is not necessarily length (e.g., see the example in equation (13.5)).

The two volumes are related by the Jacobian of transformation between the two coordinate systems

$$dV(\mathbf{x}) = \frac{\partial \mathbf{X}}{\partial \mathbf{a}} dV(\mathbf{a}). \quad (13.20)$$

This relation indicates that the Jacobian measures the ratio of the volume written in terms of position coordinates to the volume written in terms of material coordinates

$$\frac{\partial \mathbf{X}}{\partial \mathbf{a}} = \frac{dV(\mathbf{x})}{dV(\mathbf{a})}. \quad (13.21)$$

Hence, the Jacobian is a function of space according to the ratio of the volume of a fluid element in \mathbf{x} -space to the volume in \mathbf{a} -space. This is a key result of great value for transforming between Eulerian and Lagrangian coordinates.

Volume of an infinitesimal fluid element

The above results have been formulated for an arbitrary region of the fluid continuum. Hence, the expressions also hold when evaluated on an arbitrary material fluid parcel or fluid element. We use the notation δV for the parcel/element volume, in terms of which the above relations take the form

$$\delta V(\mathbf{x}) = \frac{\partial \mathbf{X}}{\partial \mathbf{a}} \delta V(\mathbf{a}) \Rightarrow \frac{\partial \mathbf{X}}{\partial \mathbf{a}} = \frac{\delta V(\mathbf{x})}{\delta V(\mathbf{a})}. \quad (13.22)$$

Hence, when evaluated on a moving material parcel, the Jacobian measures the ratio of the parcel volume written in terms of position coordinates to the parcel volume written in terms of material coordinates. In the particular case where the material coordinates are the initial fluid particle positions, then the Jacobian measures the ratio of the instantaneous volume of a fluid element to its initial volume

$$\frac{\partial \mathbf{X}}{\partial \mathbf{X}(0)} = \frac{\delta V(\mathbf{x})}{\delta V(0)}. \quad (13.23)$$

13.4.7 Further study

[This video](#) from the *National Committee for Fluid Mechanics Films* offers insightful visualizations to help understand Eulerian and Lagrangian fluid descriptions.

13.5 Lagrangian and Eulerian time derivatives

As noted in Section 13.4.2, we assume non-relativistic motion so that the material reference frame and the laboratory reference frame both measure the same universal time, t . However, when computing time derivatives, the laboratory frame does so by fixing the space coordinate, \mathbf{x} , whereas the material frame computes time derivatives by fixing the material coordinate, \mathbf{a} . These two time derivatives generally measure distinct changes in the fluid since one is computed in the laboratory frame and the other from the material frame. Relating their changes constitutes a key result of fluid kinematics.

13.5.1 Infinitesimal space-time increment of a function

Consider a fluid property as represented by a space-time dependent field, Π . For example, Π could be the temperature, mass density, or velocity component. When measured at a fixed point in space this fluid property is written mathematically as

$$\Pi = \Pi(\mathbf{x}, t). \quad (13.24)$$

The difference between $\Pi(\mathbf{x}, t)$ and $\Pi(\mathbf{x} + d\mathbf{x}, t + dt)$ delivers the differential increment, computed to leading order via a Taylor series expansion

$$d\Pi = \Pi(\mathbf{x} + d\mathbf{x}, t + dt) - \Pi(\mathbf{x}, t) \quad (13.25a)$$

$$= dt \partial_t \Pi + d\mathbf{x} \cdot \nabla \Pi. \quad (13.25b)$$

In this equation, dt is the infinitesimal time increment, and $d\mathbf{x}$ is the vector of infinitesimal space increments. For example, making use of Cartesian coordinates leads to the increment

$$d\mathbf{x} = \hat{\mathbf{x}} dx + \hat{\mathbf{y}} dy + \hat{\mathbf{z}} dz. \quad (13.26)$$

We ignore higher order terms in equation (13.25b) since the space and time increments are infinitesimal.

13.5.2 Total time derivative of a function

In fluid mechanics, it is common to sample properties of the fluid from moving reference frames. In this case, the sampling position is a function of time. Consequently, the total time derivative of Π is determined by dividing both sides of equation (13.25b) by the infinitesimal time increment

$$\frac{d\Pi}{dt} = \frac{\partial \Pi}{\partial t} + \frac{d\mathbf{x}}{dt} \cdot \nabla \Pi. \quad (13.27)$$

The first term measures the time derivative of Π at the space point \mathbf{x} . The second term accounts for changes in Π arising from movement of the reference frame relative to a point \mathbf{x} according to the velocity, $d\mathbf{x}/dt$. Expression (13.27) holds in general since the velocity of the moving frame is arbitrary. We next specialize to the two common cases in fluid mechanics.

13.5.3 Eulerian: evolution measured in the laboratory frame

The Eulerian time derivative considers the evolution of a fluid property when sampled at a fixed space point

$$\text{Eulerian time derivative} = \frac{\partial \Pi(\mathbf{x}, t)}{\partial t}. \quad (13.28)$$

This result follows from specializing the total time derivative in equation (13.27) to the case of fixed spatial points, so that $d\mathbf{x}/dt = 0$. In the geophysical fluids literature, the Eulerian time derivative is often termed the *time tendency* and flows with a nonzero time tendency are said to be *developing flows* or *evolving flows*. When the Eulerian time derivative vanishes everywhere the flow is said to be in a *steady state* or in a *steady flow* condition, with all points in the laboratory (inertial) frame measuring a zero time change for fluid properties. Steady fluid flows are the analog of static configurations in classical particle mechanics.

13.5.4 Lagrangian: evolution measured in the material frame

The Lagrangian or material time derivative measures the evolution of a fluid property sampled along the trajectory of a moving fluid particle. The Lagrangian time derivative for a field is thus written

$$\text{Lagrangian time derivative} = \frac{D\Pi}{Dt} = \frac{\partial\Pi}{\partial t} + \mathbf{v} \cdot \nabla\Pi. \quad (13.29)$$

The second equality follows by setting $d\mathbf{x}/dt = \mathbf{v}$ in equation (13.27) since we are sampling points along the fluid particle trajectory $\mathbf{x} = \mathbf{X}(\mathbf{a}, t)$. The operator $\partial/\partial t$ is the Eulerian time derivative from equation (13.28), whereas $\mathbf{v} \cdot \nabla$ is referred to as the *advection* operator. Use of the capital D for the material time operator

$$\frac{D}{Dt} = \frac{\partial}{\partial t} + \mathbf{v} \cdot \nabla \quad (13.30)$$

signals that the time derivative is computed along a fluid particle trajectory. This notation distinguishes the material time derivative from the more generic total time derivative of equation (13.27). In some texts the material time derivative is referred to as the *convective time derivative*, since the term “convection” is often used rather than our preferred term “advection”.⁸ It is also sometimes referred to as the *substantial time derivative* since it refers to the time changes following a material substance.

Equation (13.30) provides an Eulerian expression (right hand side) to the material time derivative D/Dt . There are two Eulerian contributions: the local (fixed space point) time tendency $\partial/\partial t$ and advection, $\mathbf{v} \cdot \nabla$. Advection arises in the Eulerian reference frame due to the fluid passing by the fixed laboratory observer, whereas it is absent from the material reference frame since the material frame moves with the fluid particles. Figure 13.4 illustrates the differences between the Eulerian and Lagrangian perspectives.

A *steady flow* is one with zero Eulerian time derivatives so that a steady flow does not imply a vanishing Lagrangian time derivative. Rather, a steady flow is a statement that the flow is static when viewed from the Eulerian (laboratory) reference frame. Hence, a steady flow generally has changing properties when sampled along a fluid particle trajectory. That is, there can be a nonzero Lagrangian evolution (via advection) even when the Eulerian time tendency vanishes.

13.5.5 Lagrangian time derivative formulated from the material frame

Rather than start from the total time derivative (13.27), it is instructive to work in the moving material frame *a priori*. For this purpose, we measure the function, Π , on a fluid particle, in which case it is useful to introduce the shorthand notation

$$\Pi[\mathbf{X}(\mathbf{a}, t), t] \equiv \Pi^L(\mathbf{a}, t) \quad \Longleftarrow \text{sampling } \Pi \text{ on a trajectory } \mathbf{x} = \mathbf{X}(\mathbf{a}, t) \text{ at time } t. \quad (13.31)$$

⁸In the geophysical fluids literature, “convection” generally refers to gravitationally unstable vertical motion initiated by buoyancy forces.

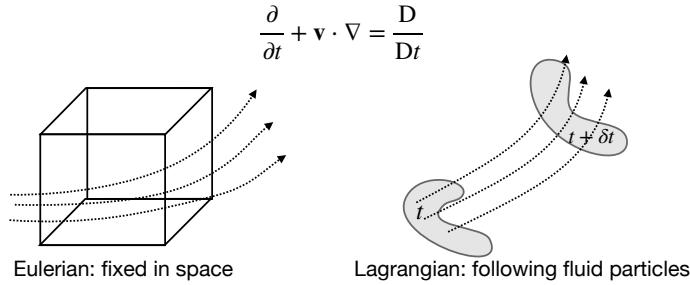


FIGURE 13.4: Illustrating the distinctions between the Eulerian (laboratory) and Lagrangian (material) reference frames for describing fluid motion. For the Eulerian description we consider a fixed control volume in the laboratory frame and measure properties as the fluid moves through the volume. For the Lagrangian description we tag fluid particles and measure fluid properties as sampled along the particle trajectories. The Eulerian representation of the material time derivative has two terms, one due to time changes local to the fixed laboratory point, and one due to the advection of properties that are swept by the local position.

In words, $\Pi^L(\mathbf{a}, t)$ is the function, Π , evaluated on a fluid particle trajectory $\mathbf{x} = \mathbf{X}(\mathbf{a}, t)$. That is, Π^L is the Lagrangian version of the function. For example, the Lagrangian velocity is given by

$$\mathbf{v}^L(\mathbf{a}, t) \equiv \mathbf{v}[\mathbf{X}(\mathbf{a}, t), t] = \left[\frac{\partial \mathbf{X}(\mathbf{a}, t)}{\partial t} \right]_{\mathbf{a}}. \quad (13.32)$$

The \mathbf{a} subscript emphasizes that the partial time derivative is computed while holding the material coordinate, \mathbf{a} , fixed.

The time derivative along a fluid particle trajectory is the material derivative, and we introduce a finite difference along the trajectory to estimate the derivative

$$\left[\frac{\partial \Pi[\mathbf{X}(\mathbf{a}, t), t]}{\partial t} \right]_{\mathbf{a}} = \lim_{\Delta t \rightarrow 0} \left[\frac{\Pi[\mathbf{X}(\mathbf{a}, t + \Delta t/2), t + \Delta t/2] - \Pi[\mathbf{X}(\mathbf{a}, t - \Delta t/2), t - \Delta t/2]}{\Delta t} \right]. \quad (13.33)$$

Expanding the numerator in a Taylor series, and keeping just the leading terms, yields

$$\left[\frac{\partial \Pi[\mathbf{X}(\mathbf{a}, t), t]}{\partial t} \right]_{\mathbf{a}} = \lim_{\Delta t \rightarrow 0} \left[\frac{\Pi[\mathbf{X}(\mathbf{a}, t + \Delta t/2), t + \Delta t/2] - \Pi[\mathbf{X}(\mathbf{a}, t - \Delta t/2), t - \Delta t/2]}{\Delta t} \right] \quad (13.34a)$$

$$= \left[\left(\frac{\partial}{\partial t} \right)_{\mathbf{X}} + \left(\frac{\partial \mathbf{X}(\mathbf{a}, t)}{\partial t} \right)_{\mathbf{a}} \cdot \nabla \right] \Pi[\mathbf{X}(\mathbf{a}, t), t] \quad (13.34b)$$

$$= \left[\left(\frac{\partial}{\partial t} \right)_{\mathbf{X}} + \mathbf{v}[\mathbf{X}(\mathbf{a}, t), t] \cdot \nabla \right] \Pi[\mathbf{X}(\mathbf{a}, t), t]. \quad (13.34c)$$

We included a subscript on the derivative operators on the right hand side to be explicit about what variables are held fixed during differentiation. This extra notation can generally be dropped, since a partial derivative operation is based on holding all variables fixed except for the variable being differentiated. Evaluating the trajectory at the spatial point, $\mathbf{X}(\mathbf{a}, t) = \mathbf{x}$, allows us to dispense with the trajectory notation to recover the more succinct expression (13.29). Even so, it is useful to keep in mind the underlying trajectory basis for the material time derivative.

13.5.6 Sample material time derivative operations

The material time derivative operator is perhaps the most important operator in fluid mechanics, and its relation to the Eulerian time derivative plus advection is a key result of fluid kinematics. Therefore, it is critical to develop experience with this operator and its generalizations. The examples here offer a starting point.

Linear wave characteristics

Consider the phase function, $\Gamma(\mathbf{x}, t) = \mathbf{k} \cdot \mathbf{x} - \omega t$, where ω is a constant angular frequency and \mathbf{k} is a constant wavevector. In the study of linear waves, surfaces of constant Γ are surfaces on which the wave phase remains constant. In the theory of partial differential equations, the direction normal to these phase surfaces define characteristic lines for linear hyperbolic equations, such as the advection equation (Section 3.2) and wave equation (Section 3.7). Now consider the material time derivative of the phase function

$$\frac{D\Gamma(\mathbf{x}, t)}{Dt} = \frac{D(\mathbf{k} \cdot \mathbf{x} - \omega t)}{Dt} \quad (13.35a)$$

$$= \left[\frac{\partial}{\partial t} + \mathbf{v} \cdot \nabla \right] (\mathbf{k} \cdot \mathbf{x} - \omega t) \quad (13.35b)$$

$$= \mathbf{v} \cdot \nabla (\mathbf{k} \cdot \mathbf{x}) - \frac{\partial(\omega t)}{\partial t} \quad (13.35c)$$

$$= \mathbf{v} \cdot \mathbf{k} - \omega. \quad (13.35d)$$

The first term says that the velocity of a point following a fluid particle is none other than the fluid velocity. This result, which follows by definition, serves as a useful means to verify internal consistency of the formalism. The second term arises from the constant angular frequency.

Material invariant

Consider a scalar function that remains constant on a material trajectory so that its material time derivative vanishes

$$\frac{D\Pi}{Dt} = 0. \quad (13.36)$$

Material constancy is generally referred to as *material invariance* in this book. We may also say that the property Π is materially conserved. At a fixed point in space, a materially invariant property has its Eulerian time derivative arising only via advection

$$\frac{\partial\Pi}{\partial t} = -\mathbf{v} \cdot \nabla\Pi. \quad (13.37)$$

Recall the case of a wave characteristic whereby $\partial\Pi/\partial t = \omega$. Wave characteristics are material if the fluid particle velocity satisfies $\mathbf{k} \cdot \mathbf{v} = \omega$. In one space dimension and for linear non-rotating gravity waves (Section 32.8.3), the wave phase speed is given by ω/k so that the wave characteristics are material if the fluid particle speed matches the wave speed, $|\mathbf{v}| = \omega/k$. The ratio of the particle speed to the wave speed is known as the Froude number. The Froude number is typically much less than unity, through highly nonlinear behavior (e.g., hydraulic jumps as discussed in [Pratt and Whitehead \(2008\)](#)) occurs when the Froude number reaches unity or larger (see Exercise 31.2).

We garner geometric insight into relation (13.37) by introducing the unit normal vector to the surface of constant Π

$$\hat{\mathbf{n}} = \frac{\nabla\Pi}{|\nabla\Pi|}. \quad (13.38)$$

Material invariance of Π thus means that the normalized Eulerian time tendency equals to the negative of the projection of the velocity into the direction normal to constant Π surfaces

$$\frac{\partial\Pi/\partial t}{|\nabla\Pi|} = -\mathbf{v} \cdot \hat{\mathbf{n}}. \quad (13.39)$$

That is, the fluid particle velocity, \mathbf{v} , is matched precisely to the velocity of the moving surface of constant Π . As detailed in Section 15.4.2, this result means there are no fluid parcels that cross surfaces of constant Π . That is, constant Π surfaces are material.

Time derivative measured in a general moving frame

Now consider a reference frame moving at an arbitrary velocity, $\mathbf{v}^{(s)}$. Examples include the quasi-Lagrangian reference frames of a float in the ocean or balloon in the atmosphere. Due to their finite size and associated drag effects, these objects only approximate material particle motion, so that $\mathbf{v}^{(s)} \neq \mathbf{v}$. Returning to the general expression (13.27) for the total time derivative, we have the time derivative operator as measured in this non-material moving reference frame

$$\frac{D^{(s)}}{Dt} = \frac{\partial}{\partial t} + \mathbf{v}^{(s)} \cdot \nabla. \quad (13.40)$$

A function that remains constant within this general moving frame thus satisfies

$$\frac{D^{(s)}\Pi}{Dt} = 0 \implies \frac{\partial\Pi}{\partial t} = -\mathbf{v}^{(s)} \cdot \nabla\Pi. \quad (13.41)$$

Introducing the normal direction $\hat{\mathbf{n}} = |\nabla\Pi|^{-1} \nabla\Pi$ leads to

$$\frac{\partial\Pi/\partial t}{|\nabla\Pi|} = -\mathbf{v}^{(s)} \cdot \hat{\mathbf{n}}, \quad (13.42)$$

which is an analog to the material constancy condition (13.39).

13.5.7 Worked example: velocity and acceleration from a trajectory

Following example 3.2 from [Kundu et al. \(2016\)](#), we here consider a one-dimensional fluid motion whereby the trajectory of a fluid particle is given by

$$\mathbf{X}(t) = \hat{\mathbf{x}} X(t) = \hat{\mathbf{x}} [K(t - t_0) + x_0^3]^{1/3}, \quad (13.43)$$

where K is a constant with dimensions volume per time and x_0 is the particle position at time t_0 . The particle velocity and particle acceleration are determined through time differentiation

$$\frac{d\mathbf{X}}{dt} = \hat{\mathbf{x}} \frac{K}{3X^2} \quad \text{and} \quad \frac{d^2\mathbf{X}}{dt^2} = -\hat{\mathbf{x}} \frac{2K^2}{9X^5}. \quad (13.44)$$

The Eulerian velocity field is then determined by

$$\mathbf{v}(\mathbf{x}, t) \equiv \left[\frac{d\mathbf{X}}{dt} \right]_{\mathbf{x}=\mathbf{X}(t)} = \hat{\mathbf{x}} \frac{K}{3x^2}, \quad (13.45)$$

which reveals that the flow is steady since there is no time dependence to the Eulerian velocity field. The Eulerian acceleration is given by the material time derivative of the Eulerian velocity, which is equal to the second time derivative of the trajectory evaluated at the field point

$$\frac{Du}{Dt} = \frac{\partial u}{\partial t} + \mathbf{v} \cdot \nabla u = 0 + u \partial_x u = -[K/(3x^2)] [(2K)/(3x^3)] = -\frac{2K^2}{9x^5} = \left[\frac{d^2X(t)}{dt^2} \right]_{\mathbf{x}=\mathbf{X}(t)}. \quad (13.46)$$

13.5.8 Summarizing some terminology

We here summarize certain terminology used in this book in reference to the variety of equations in geophysical fluid mechanics. Some of this terminology was introduced in this chapter, whereas others will be encountered in subsequent development.

- PROGNOSTIC: This is an equation that determines the time tendency (Eulerian evolution) of a quantity such as the temperature or velocity.
- DIAGNOSTIC: This is an equation that determines the value of a field at a particular instance in time. An example is the non-divergence condition satisfied by flow in an incompressible fluid as well in Boussinesq fluids (see Chapters 17 and 24). There are no time derivatives appearing in diagnostic equations.
- STEADY STATE: All Eulerian time derivatives vanish, so that all fluid properties are time independent when measured in the laboratory frame.
- MATERIAL INVARIANCE: The Lagrangian time derivative vanishes for a property that is a material invariant.

13.6 Galilean invariance

Galilean invariance means that the laws of motion are the same in all inertial reference frames. Furthermore, two inertial reference frames can only be moving with a constant velocity relative to one another. We discussed Galilean invariance for a point particle in Section 10.1.2. Here we consider its expression for a fluid. As for the particle, Galilean invariance means that the material acceleration of a fluid particle remains the same when viewed in an arbitrary inertial reference frame. Some care is required when translating this invariance into a mathematical statement when decomposing the material acceleration into its Eulerian components. Our considerations here provide a useful warmup to the more general discussion in Section 13.7, where we transform space and time derivative operators between an inertial frame and a rotating frame.

The notions of Galilean invariance hold in free space. In the presence of boundaries we do not generally expect the physical system to respect Galilean invariance since boundaries generally distinguish between reference frames. Our discussion here ignores boundaries, which is generally a useful starting point when considering the equations of motion and their kinematic properties.

13.6.1 Galilean transformation

A Galilean transformation is given by the linear space-time transformation

$$\bar{t} = t \quad \text{and} \quad \bar{\mathbf{x}} = \mathbf{x} + \mathbf{U}t \quad \text{and} \quad \bar{\mathbf{v}} = \mathbf{v} + \mathbf{U}. \quad (13.47)$$

By convention, we say that the barred coordinates are those measured in the moving (boosted) reference frame and the unbarred are measured in the rest frame. However, since both reference frames are inertial, there is no experiment on a Galilean invariant physical system that can determine which frame is at rest or which is moving. Instead, what is relevant is that the two inertial frames are moving relative to one another. Furthermore, note that time remains unchanged (universal Newtonian time), whereas the position of a point in the new frame equals to that in the original reference frame plus a contribution from the constant velocity \mathbf{U} . The inverse transformation is trivially given by

$$t = \bar{t} \quad \text{and} \quad \mathbf{x} = \bar{\mathbf{x}} - \mathbf{U}\bar{t} \quad \text{and} \quad \mathbf{v} = \bar{\mathbf{v}} - \mathbf{U}. \quad (13.48)$$

13.6.2 Transformation matrix

We take this opportunity to make use of the transformation matrix formalism to derive relations between the partial differential operators. Details of this formalism were presented in Section 7.9. However, the reader skipping that section should still be able to understand the gist of the following. For simplicity we work in the 1+1 dimensional case with time along with one space dimension.

Writing the space and time coordinates as $(t, x) = (x^0, x^1)$ and $(\bar{t}, \bar{x}) = (x^{\bar{0}}, x^{\bar{1}})$ renders the transformation of partial derivatives (following the chain rule)

$$\frac{\partial}{\partial x^{\bar{\alpha}}} = \frac{\partial x^\alpha}{\partial x^{\bar{\alpha}}} \frac{\partial}{\partial x^\alpha}, \quad (13.49)$$

where $\alpha = 0, 1$ is a tensor index that has $\alpha = 0$ for the time coordinate. The transformation matrix for the Galilean transformation is thus given by the 2×2 matrix

$$\frac{\partial x^{\bar{\alpha}}}{\partial x^\alpha} = \begin{bmatrix} \partial x^{\bar{0}}/\partial x^0 & \partial x^{\bar{0}}/\partial x^1 \\ \partial x^{\bar{1}}/\partial x^0 & \partial x^{\bar{1}}/\partial x^1 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ U & 1 \end{bmatrix}, \quad (13.50)$$

and the inverse is

$$\frac{\partial x^\alpha}{\partial x^{\bar{\alpha}}} = \begin{bmatrix} 1 & 0 \\ -U & 1 \end{bmatrix}. \quad (13.51)$$

The Jacobian determinant of the transformation matrix is unity, so that the Galilean transformation always has an inverse.

13.6.3 Transforming the differential operators

Given the transformation matrix, we can compute the Eulerian time derivative as measured in the moving frame as

$$\frac{\partial}{\partial x^{\bar{0}}} = \frac{\partial x^0}{\partial x^{\bar{0}}} \frac{\partial}{\partial x^0} + \frac{\partial x^1}{\partial x^{\bar{0}}} \frac{\partial}{\partial x^1} = \frac{\partial}{\partial x^0} - U \frac{\partial}{\partial x^1} = \frac{\partial}{\partial t} - U \frac{\partial}{\partial x}. \quad (13.52)$$

In words, this identity says that the time derivative computed between two inertial reference frames differs due to an advective term (with the constant Galilean boost velocity) arising from the relative motion of the two inertial observers. The space derivatives are related by

$$\frac{\partial}{\partial x^{\bar{1}}} = \frac{\partial x^0}{\partial x^{\bar{1}}} \frac{\partial}{\partial x^0} + \frac{\partial x^1}{\partial x^{\bar{1}}} \frac{\partial}{\partial x^1} = \frac{\partial}{\partial x^1}, \quad (13.53)$$

so that the space derivative operator remains form invariant under a Galilean transformation. This result holds also for the other two space dimensions. We thus prove that the material time derivative operator is form invariant under a Galilean transformation

$$\frac{D}{Dt} = \frac{\partial}{\partial t} + \mathbf{v} \cdot \nabla \quad (13.54a)$$

$$= \frac{\partial}{\partial \bar{t}} + \mathbf{U} \cdot \bar{\nabla} + (\bar{\mathbf{v}} - \mathbf{U}) \cdot \bar{\nabla} \quad (13.54b)$$

$$= \frac{\partial}{\partial \bar{t}} + \bar{\mathbf{v}} \cdot \bar{\nabla} \quad (13.54c)$$

$$= \frac{D}{D\bar{t}}, \quad (13.54d)$$

where we used the shorthand

$$\bar{\mathbf{v}} \cdot \bar{\nabla} = \bar{u} \frac{\partial}{\partial x^1} + \bar{v} \frac{\partial}{\partial x^2} + \bar{w} \frac{\partial}{\partial x^3}. \quad (13.55)$$

So although the individual pieces to the material time operator are modified by a Galilean transformation, the material time derivative operator is form invariant. Hence, if a function has a material time derivative $D\mathbf{u}/dt$ in one inertial reference frame, it has the same material time operator in any other inertial reference frame. That is, all properties sampled from a boosted fluid particle reference frame have the same material time evolution.

13.6.4 Comments

There are many features of geophysical fluid flows that break Galilean invariance. For example, a solid boundary breaks Galilean invariance since it establishes a special reference frame and thus breaks symmetry of the space. Additionally, a rotating planet distinguishes between longitude and latitude even if the planet is perfectly smooth. Nonetheless, as a starting point in our study of the equations of fluid mechanics, it is useful to establish their properties under a Galilean transformation. In general, if space is Galilean invariant and yet the equations of motion are not, then we question the physical relevance of the equations. Exercise 34.6 provides an example of this reasoning.

There are further symmetries of the equations of fluid mechanics, especially when there is no dissipation (inviscid). Section 2.2 of [Frisch \(1995\)](#) and Section 1.4 of [Badin and Crisciani \(2018\)](#) provide a discussion of these symmetries.

13.7 Invariance of the material time derivative

In our discussion of Galilean invariance in Section 13.6, we showed that the material time derivative operator remains form invariant under changes to the inertial reference frame. Consequently, the acceleration of a fluid particle is identical when measured in any inertial reference frame. We here consider the more general case of non-inertial reference frames that differ by both rotations and translations. We already know that the form for fluid particle accelerations differs between an inertial frame and a non-inertial frame. Nonetheless, we show here that the material time derivative operator remains form invariant. This is a result of great practical relevance, as it means that the scalar equations of fluid mechanics (e.g., mass continuity and tracer equations) remain form-invariant when changing reference frames.

13.7.1 Invariance based on definition of the material time derivative

We first determine invariance by focusing on the conceptual definition of the material time derivative. Namely, the material time derivative measures time changes of a fluid property in the reference frame of a moving fluid particle. The Lagrangian reference frame follows fluid particles, so it is the natural reference frame for measuring material time changes. In contrast, the Eulerian reference frame is fixed in a laboratory. The material time derivative computed from the laboratory reference frame is composed of an Eulerian time tendency plus an advection operator

$$\frac{D}{Dt} = \frac{\partial}{\partial t} + \mathbf{v} \cdot \nabla. \quad (13.56)$$

Importantly, this expression holds regardless the choice of laboratory reference frames, either inertial or non-inertial. Our choice of laboratory frames only impacts on the form of the Eulerian time

derivative and on the advection operator. The sum of the two terms returns the same material time derivative operator, no matter what laboratory frame is chosen.

Again for emphasis, the reason for the form invariance is that the material time derivative operator is, by definition, computed in the fluid particle reference frame. The particle reference frame is unconcerned with the subjective choice made by the observer in the laboratory reference frame. In the following, we exhibit how the mathematics respects this invariance. Namely, we show how the Eulerian expression for the material time derivative remains form invariant when changing reference frames.

13.7.2 Invariance for a rotating reference frame

Consider two reference frames. The first is at rest and so serves as an inertial frame, whereas the second is rotating with rotational axis aligned with the vertical direction as in Figure 8.1. Introduce Cartesian coordinates for the inertial frame, with corresponding basis vectors $(\hat{\bar{x}}, \hat{\bar{y}}, \hat{\bar{z}})$. Let these inertial frame unit vectors be related to rotating frame unit vectors according to

$$\hat{\bar{x}} = \hat{x} \cos \vartheta - \hat{y} \sin \vartheta \quad (13.57a)$$

$$\hat{\bar{y}} = \hat{x} \sin \vartheta + \hat{y} \cos \vartheta \quad (13.57b)$$

$$\hat{\bar{z}} = \hat{z}, \quad (13.57c)$$

and let time be the same in the two reference frames. The angle ϑ measures the counter-clockwise angle between the inertial frame direction $\hat{\bar{x}}$ and the moving frame direction \hat{x} , with this angle a linear function of time

$$\vartheta = \Omega t. \quad (13.58)$$

The above relations between the two sets of basis vectors translates into the same relations between the corresponding coordinate representations for an arbitrary vector. Including time, we have the relation between inertial coordinates (the barred frame) and rotating coordinates (unbarred frame)

$$\bar{t} = t \quad (13.59a)$$

$$\bar{x} = x \cos \vartheta - y \sin \vartheta \quad (13.59b)$$

$$\bar{y} = x \sin \vartheta + y \cos \vartheta \quad (13.59c)$$

$$\bar{z} = z. \quad (13.59d)$$

The inverse transformation can be easily found

$$t = \bar{t} \quad (13.60a)$$

$$x = \bar{x} \cos \vartheta + \bar{y} \sin \vartheta \quad (13.60b)$$

$$y = -\bar{x} \sin \vartheta + \bar{y} \cos \vartheta \quad (13.60c)$$

$$z = \bar{z}. \quad (13.60d)$$

We are now prepared to make use of the transformation formalism generally considered in Section 7.9, and specifically applied for the Galilean transformation in Section 13.6. As in these other cases, we include time as part of the formalism by introducing the Greek label $\alpha = 0, 1, 2, 3$ so that the transformation matrix between the inertial frame and rotating frame is given by

$$\frac{\partial x^{\bar{\alpha}}}{\partial x^{\alpha}} = \begin{bmatrix} \partial x^{\bar{0}}/\partial x^0 & \partial x^{\bar{0}}/\partial x^1 & \partial x^{\bar{0}}/\partial x^2 & \partial x^{\bar{0}}/\partial x^3 \\ \partial x^{\bar{1}}/\partial x^0 & \partial x^{\bar{1}}/\partial x^1 & \partial x^{\bar{1}}/\partial x^2 & \partial x^{\bar{1}}/\partial x^3 \\ \partial x^{\bar{2}}/\partial x^0 & \partial x^{\bar{2}}/\partial x^1 & \partial x^{\bar{2}}/\partial x^2 & \partial x^{\bar{2}}/\partial x^3 \\ \partial x^{\bar{3}}/\partial x^0 & \partial x^{\bar{3}}/\partial x^1 & \partial x^{\bar{3}}/\partial x^2 & \partial x^{\bar{3}}/\partial x^3 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ -\Omega \bar{y} & \cos \vartheta & -\sin \vartheta & 0 \\ \Omega \bar{x} & \sin \vartheta & \cos \vartheta & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}. \quad (13.61)$$

Similarly, the inverse transformation is given by

$$\frac{\partial x^\alpha}{\partial \bar{x}^{\bar{\alpha}}} = \begin{bmatrix} \partial x^0/\partial \bar{x}^0 & \partial x^0/\partial \bar{x}^1 & \partial x^0/\partial \bar{x}^2 & \partial x^0/\partial \bar{x}^3 \\ \partial x^1/\partial \bar{x}^0 & \partial x^1/\partial \bar{x}^1 & \partial x^1/\partial \bar{x}^2 & \partial x^1/\partial \bar{x}^3 \\ \partial x^2/\partial \bar{x}^0 & \partial x^2/\partial \bar{x}^1 & \partial x^2/\partial \bar{x}^2 & \partial x^2/\partial \bar{x}^3 \\ \partial x^3/\partial \bar{x}^0 & \partial x^3/\partial \bar{x}^1 & \partial x^3/\partial \bar{x}^2 & \partial x^3/\partial \bar{x}^3 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ \Omega y & \cos \vartheta & \sin \vartheta & 0 \\ -\Omega x & -\sin \vartheta & \cos \vartheta & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}. \quad (13.62)$$

The derivative operators transform according to

$$\frac{\partial}{\partial x^\alpha} = \frac{\partial \bar{x}^{\bar{\alpha}}}{\partial x^\alpha} \frac{\partial}{\partial \bar{x}^{\bar{\alpha}}}, \quad (13.63)$$

in which case

$$\frac{\partial}{\partial t} = \frac{\partial}{\partial \bar{t}} + (\boldsymbol{\Omega} \wedge \bar{\boldsymbol{x}}) \cdot \bar{\nabla} \quad (13.64a)$$

$$\frac{\partial}{\partial x} = \cos \vartheta \frac{\partial}{\partial \bar{x}} + \sin \vartheta \frac{\partial}{\partial \bar{y}} \quad (13.64b)$$

$$\frac{\partial}{\partial y} = -\sin \vartheta \frac{\partial}{\partial \bar{x}} + \cos \vartheta \frac{\partial}{\partial \bar{y}} \quad (13.64c)$$

$$\frac{\partial}{\partial z} = \frac{\partial}{\partial \bar{z}}. \quad (13.64d)$$

The velocity vector components transform according to

$$v^\alpha = \frac{\partial x^\alpha}{\partial \bar{x}^{\bar{\alpha}}} v^{\bar{\alpha}}, \quad (13.65)$$

so that

$$v^0 = v^{\bar{0}} \quad (13.66a)$$

$$u = \Omega y + \bar{u} \cos \vartheta + \bar{v} \sin \vartheta \quad (13.66b)$$

$$v = -\Omega x - \bar{u} \sin \vartheta + \bar{v} \cos \vartheta \quad (13.66c)$$

$$w = \bar{w}, \quad (13.66d)$$

where

$$v^0 = v^{\bar{0}} = 1. \quad (13.67)$$

Bringing these result together leads to the transformation of the horizontal advection operator

$$u \frac{\partial}{\partial x} + v \frac{\partial}{\partial y} = (\bar{\boldsymbol{u}} - \boldsymbol{\Omega} \wedge \bar{\boldsymbol{x}}) \cdot \bar{\nabla}. \quad (13.68)$$

Combining this result with the transformed Eulerian time derivative leads to the material time derivative

$$\frac{D}{Dt} = \frac{\partial}{\partial t} + \boldsymbol{v} \cdot \nabla \quad (13.69a)$$

$$= \frac{\partial}{\partial t} + u \frac{\partial}{\partial x} + v \frac{\partial}{\partial y} + w \frac{\partial}{\partial z} \quad (13.69b)$$

$$= \frac{\partial}{\partial \bar{t}} + (\boldsymbol{\Omega} \wedge \bar{\boldsymbol{x}}) \cdot \bar{\nabla} + (\bar{\boldsymbol{u}} - \boldsymbol{\Omega} \wedge \bar{\boldsymbol{x}}) \cdot \bar{\nabla} + \bar{w} \frac{\partial}{\partial \bar{z}} \quad (13.69c)$$

$$= \frac{\partial}{\partial \bar{t}} + \bar{\boldsymbol{v}} \cdot \bar{\nabla}. \quad (13.69d)$$

As advertised, the operator is form invariant under time dependent transformations to a non-inertial reference frame.

13.7.3 Invariance using space-time tensors

We can generalize the previous result by writing the material time derivative operator using space-time tensor notation, in which case

$$\frac{D}{Dt} = \frac{\partial}{\partial t} + \mathbf{v} \cdot \nabla \quad (13.70a)$$

$$= \frac{\partial}{\partial x^0} + v^m \frac{\partial}{\partial x^m} \quad (13.70b)$$

$$= v^\alpha \frac{\partial}{\partial x^\alpha}, \quad (13.70c)$$

where we introduced the velocity 4-vector

$$(v^0, v^1, v^2, v^3) = (1, v^1, v^2, v^3). \quad (13.71)$$

All space-time indices are contracted in equation (13.70c), which means the material time derivative is a space-time scalar. Consequently, we can change coordinates or change reference frames without changing the material time operator.

13.7.4 Comments

As argued at the start of this section, the invariance of the material time derivative to changes in the laboratory reference frame is rather obvious: why would a time derivative computed in a material frame be concerned with the nature of the laboratory frame? Even so, it is satisfying to see the tools of coordinate transformations put to use verifying this result. It is this sort of exercise that nurtures trust in tensor analysis, at which point it becomes a reliable tool for exploration where the answer is not *a priori* known.

13.8 Fluid flow lines

There are three types of flow lines commonly used to visualize fluid motion: *pathlines*, *streamlines*, and *streaklines*. These flow lines are identical for time independent (steady) flow, where steady flow means that all fields are constant in time when observed in the Eulerian reference frame. However, these flow lines differ for unsteady flow. They each offer complementary information about the flow field, and have uses in both theoretical and experimental contexts. We have use for pathlines and streamlines in this book, yet also introduce streaklines for completeness.

13.8.1 Material pathlines from fluid particle trajectories

As introduced in Section 13.4.1, a fluid particle traces out a *trajectory* as it moves through space and time (Figure 13.2). We use the term material *pathline* for a fluid particle trajectory, with a collection of pathlines providing a means to visualize fluid particle motion throughout the flow. In this book we are only concerned with *smooth velocity fields*, which allow for an unambiguous specification of the particle trajectory at each point of the fluid.

Mathematically, a fluid particle trajectory is a curve, $\mathbf{X}(\mathbf{a}, t)$, in space that is traced by fixing the material coordinate, \mathbf{a} , and letting time advance. Trajectories are computed by time integrating the ordinary differential equation

$$\frac{\partial \mathbf{X}(\mathbf{a}, t)}{\partial t} = \mathbf{v}[\mathbf{X}(\mathbf{a}, t), t] \quad (13.72a)$$

$$\mathbf{X}(\mathbf{a}, t = 0) = \mathbf{a}, \quad (13.72b)$$

where

$$\mathbf{v}[\mathbf{X}(\mathbf{a}, t), t] = \mathbf{v}^L(\mathbf{a}, t) \quad (13.73)$$

is Lagrangian velocity of the fluid particle (see equation (13.32)), and we have assumed the material coordinates are determined by the initial position. Again, the partial time derivative is computed with the material coordinate held fixed, so that the material coordinate distinguishes between particle trajectories. In a mathematical context we refer to a particle trajectory as the *integral curve* for the velocity vector.

In the laboratory, we can insert tiny particles into the fluid to offer a means for visualizing the flow, with a time exposed photograph providing an estimate of fluid particle pathlines. Experimental particles provide an accurate estimate of fluid particle pathlines if the particles do not disperse through diffusion (see Chapter 48). As another example, consider cars moving at night with a time exposed photograph revealing car pathlines formed by car head and tail lights. Like cars, the material pathlines in a fluid can intersect, cross, and become quite complex, particularly when the flow is turbulent.

13.8.2 Fluid streamlines and streamtubes

Streamlines are curves whose tangent is parallel to the instantaneous fluid velocity field. Streamlines can intersect only at a stagnation point; i.e., a point where the fluid is not moving. Let

$$d\mathbf{x} = \hat{\mathbf{x}} dx + \hat{\mathbf{y}} dy + \hat{\mathbf{z}} dz \quad (13.74)$$

be an infinitesimal increment along a streamline written using Cartesian coordinates. The family of streamlines at a given time, t , satisfy the tangent constraint

$$\mathbf{v} \wedge d\mathbf{x} = 0, \quad (13.75)$$

which is equivalent to

$$\frac{dx}{u(\mathbf{x}, t)} = \frac{dy}{v(\mathbf{x}, t)} = \frac{dz}{w(\mathbf{x}, t)}. \quad (13.76)$$

Alternatively, we can introduce a pseudo-time parameter, s , that determines a position along a streamline. Streamlines are the curves $\mathbf{x} = \mathbf{X}(s; \mathbf{a}, t)$ computed with (\mathbf{a}, t) held fixed, but with the pseudo-time varied

$$\frac{\partial \mathbf{X}(s; \mathbf{a}, t)}{\partial s} = \mathbf{v}[\mathbf{X}(s; \mathbf{a}, t), t] \quad (13.77a)$$

$$\mathbf{X}(s = 0; \mathbf{a}, t) = \mathbf{a}. \quad (13.77b)$$

Again, both the material coordinate \mathbf{a} and time t are held fixed when determining streamlines, so that (\mathbf{a}, t) act as parameters to distinguish streamlines. Streamlines thus do not know about the time evolution of unsteady flow. Instead, streamlines only sample a snapshot of the velocity field; they are freshly computed at each time instance.

A *streamtube* is a bundle of streamlines crossing through an arbitrary closed curve (see Figure 13.5). Hence, at each time instance, streamtube sides are parallel to the velocity vector. Furthermore, when the flow is steady then streamlines are identical to material particle pathlines. Hence, a streamtube is a material tube for steady flow, in which case no fluid particles cross the streamtube boundary.

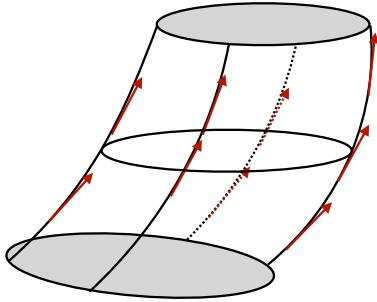


FIGURE 13.5: This image shows an example streamtube. The side boundaries of a streamtube consist of streamlines. At each point of a streamline, the local tangent vector equals to the velocity field (see equation (13.75)). Streamlines are identical to pathlines only for steady flow; they differ for unsteady flows. Hence, for unsteady flows, particle trajectories generally cross through the streamtube boundary.

13.8.3 Distinguishing streamlines and pathlines

The tangent to a streamline gives the velocity at a single point in time, whereas the tangent to a material pathline (i.e., a trajectory) gives the velocity at subsequent times. These tangents are identical when the flow is steady. However, if the flow is time dependent (unsteady), then streamlines differ from material pathlines. Furthermore, for unsteady flow, the pseudo-time parameter, s , determining the streamlines in equation (13.77a) is not equal to the time, t , used to compute fluid particle trajectories in equation (13.72a). Consequently, the condition $\mathbf{v} \cdot \hat{\mathbf{n}} = 0$ satisfied at each instance by a streamline does not ensure that fluid particles do not cross streamlines. The reason is that a material pathline moves with the fluid in such a way that

$$(\mathbf{v} - \mathbf{v}^{\text{line}}) \cdot \hat{\mathbf{n}} = 0 \implies \mathbf{v} \cdot \hat{\mathbf{n}} = \mathbf{v}^{\text{line}} \cdot \hat{\mathbf{n}} \quad \text{material lines,} \quad (13.78)$$

where \mathbf{v}^{line} is the velocity of a point on the material pathline. The material pathline thus moves so that no fluid particles cross it. Only when the flow is steady, so that $\mathbf{v}^{\text{line}} \cdot \hat{\mathbf{n}} = 0$, will material pathlines and streamlines be equal. That is, the streamline constraint $\mathbf{v} \cdot \hat{\mathbf{n}} = 0$ is not a material constraint when $\mathbf{v}^{\text{line}} \cdot \hat{\mathbf{n}} \neq 0$. The key point is that streamlines do not probe the time behavior of the flow, so they do not know whether the velocity is steady or unsteady.

13.8.4 Fluid streaklines

A *streakline* is a curve obtained by connecting the positions for all fluid particles that emanate from a fixed point in space (see Figure 13.6). Streaklines are simple to define conceptually and to realize experimentally. However, they are a bit convoluted to specify mathematically. We thus present two formulations.

At any time t , the streakline passing through a fixed point \mathbf{y} is a curve going from \mathbf{y} to $\mathbf{X}(\mathbf{y}, t)$, the position reached by the particle initialized at $t = 0$ at the point \mathbf{y} . A particle is on the streakline if it passed the fixed point \mathbf{y} at some time between 0 and t . If this time was s , then the material coordinate of the particle would be given by $\mathbf{a}(\mathbf{y}, s)$ (see equation (13.8) relating the material coordinate to its corresponding laboratory position). Furthermore, at time t , this particle is at \mathbf{x} , so that the equation of the streakline at time t is

$$\mathbf{x} = \mathbf{X}[\mathbf{a}(\mathbf{y}, s), t] \quad 0 \leq s \leq t. \quad (13.79)$$

We can connect the streakline specification to that given for a pathline and streamline through the following. A streakline at some time instance \tilde{t} is a curve defined by fixing \tilde{t} and varying s over

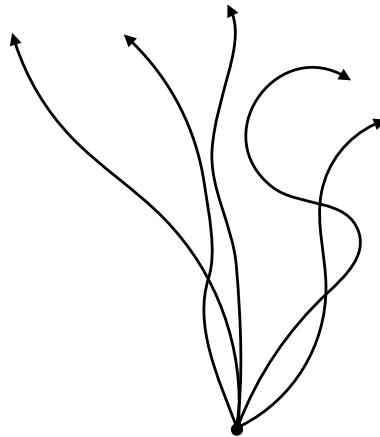


FIGURE 13.6: A suite of trajectories emanating from a single point. Common realizations include the paths of fluid particles that leave from a chimney, or the smoke from a point source. A streakline is defined as the accumulation of positions at time t of particles that passed through the common point at some earlier time $s < t$.

$s \leq \tilde{t}$ in the function $\mathbf{X}(s; \mathbf{a}, \tilde{t})$. We determine the curves $\mathbf{x} = \mathbf{X}(s; \mathbf{a}, \tilde{t})$ by solving the following set of initial value problems for trajectories with initial conditions imposed at $t = s$ rather than $t = 0$

$$\frac{\partial \mathbf{X}(s; \mathbf{a}, t)}{\partial t} = \mathbf{v}[\mathbf{X}(s; \mathbf{a}, t), t] \quad (13.80a)$$

$$\mathbf{X}(t = s; \mathbf{a}, t) = \mathbf{a}. \quad (13.80b)$$

Note that \mathbf{a} remains fixed, as we start all trajectories determining a streakline from the same initial point (e.g., the chimney does not move). A streakline can thus be generated by emitting a dye from a point over a time interval equal to the range of s , with the dye following fluid particle trajectories.

13.8.5 An analytic example of flow lines

Consider the following two-dimensional example as taken from Section 4.13 of [Aris \(1962\)](#). Let the Eulerian velocity field be given by

$$u = \frac{x}{\tau + t} \quad (13.81a)$$

$$v = \frac{y}{\tau} \quad (13.81b)$$

$$w = 0, \quad (13.81c)$$

where $\tau > 0$ is a constant with the dimensions of time.

Pathlines

Pathlines are determined by solving the trajectory equations

$$\frac{dX(t)}{dt} = \frac{X(t)}{\tau + t} \quad (13.82a)$$

$$\frac{dY(t)}{dt} = \frac{Y(t)}{\tau} \quad (13.82b)$$

$$\frac{dZ(t)}{dt} = 0, \quad (13.82c)$$

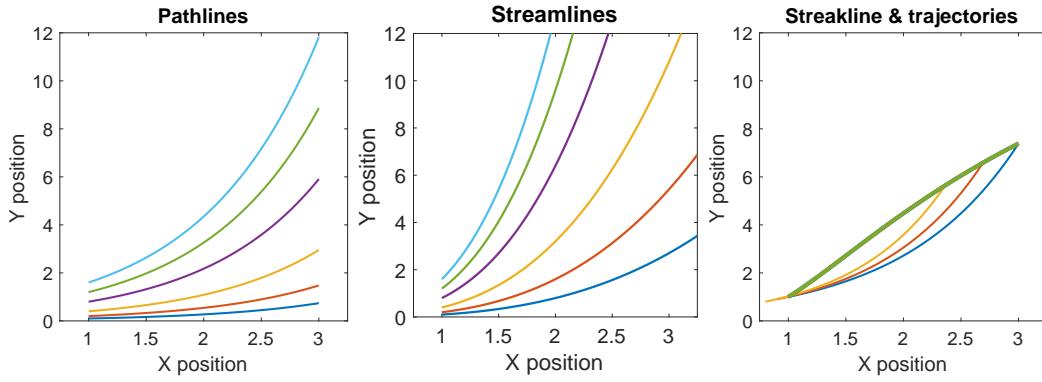


FIGURE 13.7: Left panel: sample pathlines $X(t) = X_0(1 + t/\tau)$ and $Y(t) = Y_0 e^{t/\tau}$ (see equations (13.83a) and (13.83b)) during times $t \in [0, 2]$. The trajectories drawn here all start at $X_0 = 1$ and set the parameter $\tau = 1$. Note that those pathlines with $X_0 = 0$ remain on the y-axis, and those with $Y_0 = 0$ remain on the x-axis. Middle panel: Sample streamlines $X(s; t) = X_0 e^{s/(\tau+t)}$ and $Y(s; t) = Y_0 e^{s/\tau}$ (see equations (13.86a) and (13.86b)). We set $t = 2$ and let the pseudo-time run from $s \in [0, 4]$. All streamlines shown here start at $X_0 = 1$. Note that those that start with $X_0 = 0$ remain on the y-axis, and those that start with $Y_0 = 0$ remain on the x-axis. Right panel: sample analytic streakline (dark bold line) at $t = 2$ according to equations (13.89a) and (13.89b). This streakline is determined by the position of particles at $t = 2$ that pass through $(X, Y) = (1, 1)$ during times $t \in (-\infty, 2)$. We show three sample trajectories that fall onto the streakline. The longest trajectory starts at $(X, Y) = (1, 1)$ at $t = 0$, whereas the two shorter trajectories pass through $(X, Y) = (1, 1)$ at some time $0 < t < 2$. Notice the distinction between all three flow lines, which is to be expected since the flow field is unsteady.

which are found to be

$$X(t) = X_0(1 + t/\tau) \quad (13.83a)$$

$$Y(t) = Y_0 e^{t/\tau} \quad (13.83b)$$

$$Z(t) = Z_0, \quad (13.83c)$$

where $\mathbf{X}(t = 0) = \mathbf{X}_0$. Sample trajectories are shown in Figure 13.7 over time $t \in [0, 2]$. We can eliminate time to yield a curve in the horizontal (x, y) plane

$$y = Y_0 e^{(x-X_0)/X_0}. \quad (13.84)$$

Streamlines

Streamlines are determined by solving the differential equations

$$\frac{dX(s; t)}{ds} = \frac{X(s; t)}{\tau + t} \quad (13.85a)$$

$$\frac{dY(s; t)}{ds} = \frac{Y(s; t)}{\tau} \quad (13.85b)$$

$$\frac{dZ(s; t)}{ds} = 0, \quad (13.85c)$$

where time, t , is a fixed parameter whereas the pseudo-time, s , is varied. Integration renders the streamlines

$$X(s; t) = X_0 e^{s/(\tau+t)} \quad (13.86a)$$

$$Y(s; t) = Y_0 e^{s/\tau} \quad (13.86b)$$

$$Z(s; t) = Z_0. \quad (13.86c)$$

Sample streamlines are shown in Figure 13.7. Note that we can eliminate the pseudo-time s to render a curve in the horizontal (x, y) plane

$$y = Y_0 \left[\frac{x}{X_0} \right]^{(\tau+t)/\tau} \quad (13.87a)$$

$$z = Z_0. \quad (13.87b)$$

Streaklines

For streaklines, invert the trajectory expressions (13.83a)-(13.83b) to find the material coordinates $\mathbf{a}(y, s)$ in the form

$$a_1 = \frac{y_1}{1 + s/\tau} \quad (13.88a)$$

$$a_2 = y_2 e^{-s/\tau} \quad (13.88b)$$

$$a_3 = y_3. \quad (13.88c)$$

We next evaluate the trajectory expressions (13.83a)-(13.83b) with \mathbf{a} as the initial positions to find the streaklines

$$X(s; \mathbf{a}, t) = \frac{y_1 (1 + t/\tau)}{1 + s/\tau} \quad (13.89a)$$

$$Y(s; \mathbf{a}, t) = y_2 e^{(t-s)/\tau} \quad (13.89b)$$

$$Z(s; \mathbf{a}, t) = y_3. \quad (13.89c)$$

Figure 13.7 illustrates the streakline for a particular point $(X, Y) = (1, 1)$.

13.8.6 Further study

A discussion of flow lines can be found in most books on fluid mechanics. The presentation here borrows from Sections 4.11-4.13 of [Aris \(1962\)](#), Section 3.3 of [Kundu et al. \(2016\)](#), and online lecture notes on fluid kinematics from Professor McIntyre of Cambridge University.

13.9 Exercises

EXERCISE 13.1: FLUID VELOCITY AND ACCELERATION DERIVED FROM A TRAJECTORY

Following the example in Section 13.5.7, consider the one-dimensional fluid particle trajectory

$$\mathbf{X}(t) = \hat{\mathbf{x}} X(t) = \hat{\mathbf{x}} [k(t - t_0)^2 + x_0^3]^{1/3}, \quad (13.90)$$

where k is a constant with dimensions $L^3 T^{-2}$ and x_0 is the particle position at time t_0 .

- (a) Determine the velocity of the fluid particle.
- (b) Determine the acceleration of the fluid particle.
- (c) Determine the Eulerian velocity field.
- (d) Determine the Eulerian acceleration field and show that it equals to the particle acceleration when evaluated at the field point, $\mathbf{x} = \mathbf{X}(t)$.

EXERCISE 13.2: MATERIAL EVOLUTION OF THE PARTIAL DERIVATIVE OF A FUNCTION

In this exercise we establish some properties of the material time derivative operator when acting on spatial derivatives of a scalar field.

- (a) If a scalar field Π is materially constant, prove that the material evolution of its spatial derivative is given by

$$\frac{D(\partial_m \Pi)}{Dt} = -\partial_m \mathbf{v} \cdot \nabla \Pi. \quad (13.91)$$

For example, if $D\Pi/Dt = 0$, then the zonal partial derivative $\partial_x \Pi$ has a material time derivative given by

$$\frac{D(\partial \Pi / \partial x)}{Dt} = -\frac{\partial \mathbf{v}}{\partial x} \cdot \nabla \Pi. \quad (13.92)$$

- (b) What is the material time derivative of $\nabla \Pi$ for the case that Π is not materially constant?

Hint: Some might find it more suitable to first solve the general case in part (a) and then specialize to $D\Pi/Dt = 0$ in part (b). Also, use Cartesian tensors for convenience.



Material fluid objects

Draw an imaginary line, area, or volume within a homogeneous perfect fluid. Since the fluid is perfect, all fluid particles that start on the geometric object remain on the object, even as the shape of the geometric object is deformed by the moving fluid. We say that the geometric object is a perfectly *coherent structure*. We study the kinematics of such objects in this chapter by using both Lagrangian and Eulerian descriptions, with the two offering complementary insights. Some extra attention is given to the kinematics of two-dimensional flow due to the relative mathematical ease and the associated intuition that proves useful for more general three dimensional flows.

The kinematic ideas developed in this chapter for a perfect fluid have use in the study of mixing and transport within real geophysical fluids. That is, for a real fluid we also find it useful to study geometric objects that follow the flow. However, in a real fluid such objects are not perfectly coherent since matter generally crosses all surfaces through diffusive mixing. The question then becomes whether there are special surfaces where such mixing is suppressed, with this question forming a key part of the study of coherent structures in real fluids (e.g., [Haller \(2015\)](#)).

READER'S GUIDE FOR THIS CHAPTER

This chapter builds from the kinematics of Chapter 13 and makes some use of Section 3.1 of [Batchelor \(1967\)](#), whereas a more thorough development is given by [Ottino \(1989\)](#). These ideas and methods are of interest to a variety of kinematic aspects of fluid motion, with particular application to the kinematics of mixing, stirring, and coherency in eddying geophysical fluids. The discussion is restricted to Cartesian tensors to reduce the math overhead. Consequently, all tensor labels are downstairs with no distinction between covariant and contravariant. Extensions to arbitrary coordinates follow the methods of Chapter 7.

14.1	Increments in material and position space	270
14.1.1	Eulerian/position space differential	270
14.1.2	Lagrangian/material space differential	270
14.1.3	Duality between Eulerian and Lagrangian perspectives	271
14.2	Evolution of material curves	271
14.2.1	Deformation gradient tensor	272
14.2.2	Cauchy-Green strain tensor	272
14.2.3	Kinematic evolution equation	273
14.2.4	Cauchy's solution	273
14.2.5	Velocity gradient tensor	275
14.2.6	Stretching and tilting of a material lines	276
14.2.7	Rate of strain tensor	276

14.2.8	Rotation tensor	277
14.2.9	Cauchy-Stokes decomposition	278
14.3	Evolution of material surfaces	278
14.3.1	Surfaces in three-dimensional flow	278
14.3.2	Evolution of the area	279
14.3.3	Evolution of the normal vector	280
14.3.4	Material area in two-dimensional flow	281
14.4	Volume and the Jacobian of transformation	282
14.4.1	Material parcel volume	282
14.4.2	Evolution of the Jacobian of transformation	283
14.5	Kinematics of two-dimensional flow	283
14.5.1	Diverging flow	284
14.5.2	Rotational flow with nonzero vorticity	285
14.5.3	Flow with nonzero deformation	285
14.5.4	Further study	286
14.6	Exercises	286

14.1 Increments in material and position space

In this section we summarize some of the maths associated with the differential increment of a function. In particular, we expose how the increment appears in both position/Eulerian space (\mathbf{x} -space or \mathbf{x} -coordinates) and in material/Lagrangian space (\mathbf{a} -space or \mathbf{a} -coordinates).

14.1.1 Eulerian/position space differential

In Section 13.5.1, we considered the space-time increment of a function. Here we consider just the space increment, as defined by the differential increment of a function evaluated at the same time but at two infinitesimally close points in space

$$d\Pi(\mathbf{x}, t) = \Pi(\mathbf{x} + d\mathbf{x}, t) - \Pi(\mathbf{x}, t) \quad (14.1a)$$

$$= (d\mathbf{x} \cdot \nabla)\Pi. \quad (14.1b)$$

The dimensionless operator

$$d\mathbf{x} \cdot \nabla = dx_m \frac{\partial}{\partial x_m} \quad (14.2)$$

is a scalar since it remains form invariant when switching to another set of Cartesian position coordinates.¹

14.1.2 Lagrangian/material space differential

Consider the same function Π evaluated on a material particle trajectory, and write this “Lagrangian” function as

$$\Pi^L(\mathbf{a}, t) = \Pi[\mathbf{X}(\mathbf{a}, t), t]. \quad (14.3)$$

In words, the Lagrangian version of a function is obtained by evaluating that function on a fluid particle trajectory. We use the notation $\Pi^L(\mathbf{a}, t)$ as a shorthand, which is defined by this equality.²

¹This form invariance also holds when using curvilinear coordinates if we make use of the general tensor analysis formalism of Chapter 7.

²We introduced this notation in Section 13.5.5.

Consider an infinitesimal increment of $\Pi^L(\mathbf{a}, t)$ within material coordinate space. This increment represents the difference of Π when evaluated on two separate fluid particles labelled by \mathbf{a} and $\mathbf{a} + \delta\mathbf{a}$. Note that we use the δ symbol to signal material increments. Taking a Taylor series and truncating to leading order yields

$$\delta\Pi^L(\mathbf{a}, t) = \Pi[\mathbf{X}(\mathbf{a} + \delta\mathbf{a}, t), t] - \Pi[\mathbf{X}(\mathbf{a}, t), t] \quad (14.4a)$$

$$= \Pi^L(\mathbf{a} + \delta\mathbf{a}, t) - \Pi^L(\mathbf{a}, t) \quad (14.4b)$$

$$= (\delta\mathbf{a} \cdot \nabla_{\mathbf{a}})\Pi^L(\mathbf{a}, t). \quad (14.4c)$$

The dimensionless operator

$$\delta\mathbf{a} \cdot \nabla_{\mathbf{a}} = \delta a_j \frac{\partial}{\partial a_j} \quad (14.5)$$

is a scalar since it remains form invariant when switching to another set of Cartesian material coordinates.³ We use the notation $\nabla_{\mathbf{a}}$ to emphasize that the gradient operator is in material space rather than position space.

14.1.3 Duality between Eulerian and Lagrangian perspectives

By construction, the value of a function at a postion \mathbf{x} (Eulerian perspective) equals to the function evaluated on a moving fluid particle (Lagrangian perspective) when the particle trajectory passes through \mathbf{x} . Mathematically, this identity takes the form

$$\Pi^L(\mathbf{a}, t) = \Pi(\mathbf{x}, t) \quad \text{if } \mathbf{X}(\mathbf{a}, t) = \mathbf{x}. \quad (14.6)$$

Likewise, if the infinitesimal increment in space, $\delta\mathbf{x}$, equals to the vector increment of the two fluid particles,

$$\delta\mathbf{X}(\mathbf{a}, t) = \mathbf{X}(\mathbf{a} + \delta\mathbf{a}, t) - \mathbf{X}(\mathbf{a}, t), \quad (14.7)$$

then the functional increments are identical

$$\delta\Pi^L(\mathbf{a}, t) = \delta\Pi(\mathbf{x}, t) \quad \text{if } \delta\mathbf{X}(\mathbf{a}, t) = \delta\mathbf{x}, \quad (14.8)$$

where

$$\delta\Pi(\mathbf{x}, t) = \Pi(\mathbf{x} + \delta\mathbf{x}, t) - \Pi(\mathbf{x}, t). \quad (14.9)$$

These identities allow us to develop relations using either a Lagrangian or an Eulerian perspective, and then to interpret them in the complementary perspective. We make routine use of this Eulerian/Lagrangian duality throughout this book.

14.2 Evolution of material curves

Material curves, and their generalizations to areas and volumes, are one-dimensional geometric objects that follow fluid particles in a perfect fluid. We initialize a material curve by drawing a line in the fluid and then following the fluid particles contained on the initial line. The material curve is stretched and folded by the fluid flow (see Figure 14.1). But since the fluid is perfect, the same fluid particles remain on the curve so that it retains exact coherency. We here develop the rudimentary kinematics of material curves.

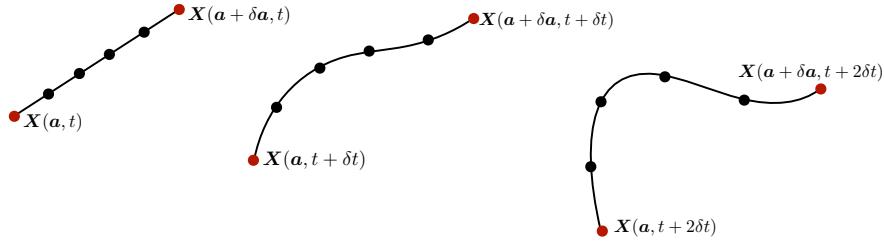


FIGURE 14.1: Three time instances of a material curve with endpoints determined by the trajectories of two fluid particles, $\mathbf{X}(\mathbf{a}, t)$ and $\mathbf{X}(\mathbf{a} + \delta\mathbf{a}, t)$. All points between the endpoints are part of the material curve, and the curve moves through the fluid by following the trajectories of the fluid particles. Kinematics of the curve are determined by properties of the velocity gradient tensor discussed in Section 14.2.5.

14.2.1 Deformation gradient tensor

Assume the material curve is initially infinitesimal with endpoints given by fluid particles with trajectories $\mathbf{X}(\mathbf{a}, t)$ and $\mathbf{X}(\mathbf{a} + \delta\mathbf{a}, t)$ (see Figure 14.1). At time t , the vector displacement between these two particles is written

$$\delta\mathbf{X}(\mathbf{a}, t) = \mathbf{X}(\mathbf{a} + \delta\mathbf{a}, t) - \mathbf{X}(\mathbf{a}, t). \quad (14.10)$$

Expanding this expression to leading order yields an expression of Cauchy's solution (see Section 14.2.4)

$$\delta\mathbf{X}(\mathbf{a}, t) = \mathbf{X}(\mathbf{a} + \delta\mathbf{a}, t) - \mathbf{X}(\mathbf{a}, t) \approx (\delta\mathbf{a} \cdot \nabla_{\mathbf{a}}) \mathbf{X}(\mathbf{a}, t), \quad (14.11)$$

where again $\nabla_{\mathbf{a}}$ is the gradient operator acting on the material coordinates. Writing this equation in component form leads to

$$\delta X_m = \delta a_j \frac{\partial X_m}{\partial a_j}. \quad (14.12)$$

As in Chapter 13, we assign the labels for m, n, p position/Eulerian coordinates, \mathbf{x} , and the labels i, j, k for material/Lagrangian coordinates, \mathbf{a} .

The components

$$F_{mj} \equiv \frac{\partial X_m}{\partial a_j} \quad (14.13)$$

appearing in equation (14.12) form elements of the transformation matrix linking position space (\mathbf{x} -space) to material space (\mathbf{a} -space). We already encountered this tensor in Section 13.4.3 (see equation (13.9)). In the continuum mechanics literature, the tensor (14.13) is known as the *displacement gradient tensor* or the *deformation gradient tensor*.

14.2.2 Cauchy-Green strain tensor

The deformation gradient tensor plays a role in measuring the length of material curves. We see this role by considering the squared length of an infinitesimal portion of a material curve

$$\delta\mathbf{X} \cdot \delta\mathbf{X} = \frac{\partial\mathbf{X}}{\partial a_i} \cdot \frac{\partial\mathbf{X}}{\partial a_j} \delta a_i \delta a_j \equiv C_{ij} \delta a_i \delta a_j. \quad (14.14)$$

Algebraically, this expression is a quadratic form, and the symmetric tensor

$$C_{ij} \equiv \frac{\partial\mathbf{X}}{\partial a_i} \cdot \frac{\partial\mathbf{X}}{\partial a_j} = F_{mi} F_{mj} \quad (14.15)$$

³ As for the position space (\mathbf{x} -space), this invariance also holds when using curvilinear material coordinates if we make use of the general tensor analysis formalism of Chapter 7.

is the metric tensor that provides the means to measure distance along an infinitesimal material curve. This metric tensor is called the *Cauchy-Green strain* tensor.

14.2.3 Kinematic evolution equation

Now consider the material time derivative of the material increment

$$\frac{\partial[\delta\mathbf{X}(\mathbf{a}, t)]}{\partial t} = \frac{\partial\mathbf{X}(\mathbf{a} + \delta\mathbf{a}, t)}{\partial t} - \frac{\partial\mathbf{X}(\mathbf{a}, t)}{\partial t} \quad (14.16a)$$

$$= \mathbf{v}^L(\mathbf{a} + \delta\mathbf{a}, t) - \mathbf{v}^L(\mathbf{a}, t) \quad (14.16b)$$

$$\equiv \delta\mathbf{v}^L(\mathbf{a}, t). \quad (14.16c)$$

In these equations, we introduced the Lagrangian velocity

$$\mathbf{v}^L(\mathbf{a}, t) = \mathbf{v}[\mathbf{X}(\mathbf{a}, t), t] \quad (14.17)$$

as per equation (13.32) and the discussion in Section 14.1.2. As for the manipulations in Section 14.2.1, we can massage the expression (14.16c) by performing a Taylor series expansion and truncating to leading order

$$\frac{\partial[\delta\mathbf{X}(\mathbf{a}, t)]}{\partial t} = \delta\mathbf{v}^L(\mathbf{a}, t) = (\delta\mathbf{a} \cdot \nabla_{\mathbf{a}}) \mathbf{v}^L(\mathbf{a}, t). \quad (14.18)$$

Alternatively, we can choose to evaluate this expression using an Eulerian perspective (see Section 14.1.3), in which case

$$\frac{D(\delta\mathbf{x})}{Dt} = \delta\mathbf{v}(\mathbf{x}, t) = (\delta\mathbf{x} \cdot \nabla_{\mathbf{x}}) \mathbf{v}(\mathbf{x}, t). \quad (14.19)$$

14.2.4 Cauchy's solution

We can determine a general solution to the kinematic evolution equation (14.19) following the method used by Cauchy to solve the perfect fluid barotropic vorticity equation discussed in Section 36.6.⁴ Indeed, we already encountered the Cauchy solution in Section 14.2.1. We rederive it here, following a more laborious route, to expose some of the formalism that can be of use when working with Lagrangian coordinates, and to offer some further insight into the meaning of the solution.

⁴See [Frisch and Villone \(2014\)](#) for an insightful discussion of the enduring impact of Cauchy's solution.

Derivation

The derivation that follows involves strategic uses of the transformation between \mathbf{a} -coordinates and \mathbf{x} -coordinates, which we start by writing

$$\frac{D(\delta x_n)}{Dt} = (\delta \mathbf{x} \cdot \nabla_{\mathbf{x}}) v_n \quad \text{kinematic evolution} \quad (14.20a)$$

$$= \delta x_m \frac{\partial}{\partial x_m} \left[\frac{Dx_n}{Dt} \right] \quad v_n = Dx_n/Dt \quad (14.20b)$$

$$= \delta x_m \frac{\partial a_j}{\partial x_m} \frac{\partial}{\partial a_j} \left[\frac{Dx_n}{Dt} \right] \quad \text{chain rule} \quad (14.20c)$$

$$= \delta x_m \frac{\partial a_j}{\partial x_m} \frac{D}{Dt} \frac{\partial x_n}{\partial a_j} \quad \text{commute } \partial/\partial a_j \text{ and } D/Dt \quad (14.20d)$$

$$= \delta x_m \frac{D}{Dt} \left[\frac{\partial a_j}{\partial x_m} \frac{\partial x_n}{\partial a_j} \right] - \delta x_m \frac{D}{Dt} \left[\frac{\partial a_j}{\partial x_m} \right] \frac{\partial x_n}{\partial a_j} \quad \text{move } D/Dt \quad (14.20e)$$

$$= -\delta x_m \frac{D}{Dt} \left[\frac{\partial a_j}{\partial x_m} \right] \frac{\partial x_n}{\partial a_j}. \quad \frac{\partial a_j}{\partial x_m} \frac{\partial x_n}{\partial a_j} = \delta_{mn} = \text{constant.} \quad (14.20f)$$

The commutation in equation (14.20d) follows since the material coordinates are fixed on a fluid particle so that

$$\frac{\partial}{\partial a_j} \frac{D}{Dt} = \frac{D}{Dt} \frac{\partial}{\partial a_j}. \quad (14.21)$$

Contracting the identity (14.20f) with $\partial a_k / \partial x_n$ yields

$$\frac{\partial a_k}{\partial x_n} \frac{D(\delta x_n)}{Dt} = -\delta x_m \frac{D}{Dt} \left[\frac{\partial a_j}{\partial x_m} \right] \frac{\partial x_n}{\partial a_j} \frac{\partial a_k}{\partial x_n} \quad \text{contraction (inner product)} \quad (14.22a)$$

$$= -\delta x_m \frac{D}{Dt} \left[\frac{\partial a_k}{\partial x_m} \right] \quad \frac{\partial x_n}{\partial a_j} \frac{\partial a_k}{\partial x_n} = \delta_{jk} \quad (14.22b)$$

$$= -\delta x_n \frac{D}{Dt} \left[\frac{\partial a_k}{\partial x_n} \right] \quad \text{change indices from } m \text{ to } n. \quad (14.22c)$$

We can thus identify the material invariant

$$\frac{D}{Dt} \left[\delta x_n \frac{\partial a_k}{\partial x_n} \right] = 0 \implies \delta \mathbf{x} \cdot \nabla_{\mathbf{x}} a_k = \text{material invariant.} \quad (14.23)$$

This result follows rather trivially by recognizing the coordinate transformation

$$\delta \mathbf{x} \cdot \nabla_{\mathbf{x}} a_k = \delta x_n \frac{\partial a_k}{\partial x_n} = \delta a_k, \quad (14.24)$$

and noting that the δa_k is a material invariant, in which case $\delta \mathbf{x} \cdot \nabla_{\mathbf{x}} a_k$ must also be materially invariant.

If we choose the material coordinates as the initial positions of fluid particles,

$$\mathbf{a} = \mathbf{x}(t = 0), \quad (14.25)$$

then the Jacobian at the initial time is given by the unit tensor

$$\frac{\partial a_k}{\partial x_m} = \delta_{km}, \quad (14.26)$$

so that the material invariance in equation (14.23) leads to

$$\delta\mathbf{x}(t) \cdot \nabla_{\mathbf{x}} a_k = \delta x_k(t=0). \quad (14.27)$$

Inverting this equation by contracting with $\partial x_n / \partial a_k$ leads to *Cauchy's solution*

$$\delta x_n(t) = \delta x_k(t=0) \frac{\partial x_n}{\partial a_k}. \quad (14.28)$$

Discussion and extension to vorticity

The Cauchy solution (14.28) says that the increment along a line segment defined by fluid particles, $\delta\mathbf{x}(t)$, expands or contracts according to the time and space dependent Jacobian, $\partial x_n / \partial a_k$. This result follows from our assumption that the increment is defined by fixed \mathbf{a} -space coordinates (fixed fluid particles), which then constrains the increment's \mathbf{x} -space evolution. It is also an expression that we wrote down, rather trivially, in equation (14.11) when performing a truncated Taylor series expansion of the spatial increment between two material fluid particles. The ability to derive certain kinematic results using either simple (indeed trivial) methods versus more long-winded methods is somewhat exemplary of fluid kinematics. Both approaches offer distinct insights as well as confidence in the validity of the result.⁵

As seen in Section 36.4, the relative vorticity, $\boldsymbol{\omega} = \nabla \wedge \mathbf{v}$, for a homogeneous inviscid, barotropic fluid satisfies the same kinematic equation (14.19) as for a material line increment

$$\frac{D\boldsymbol{\omega}}{Dt} = (\boldsymbol{\omega} \cdot \nabla_{\mathbf{x}}) \mathbf{v}(\mathbf{x}, t). \quad (14.29)$$

We can thus invoke the Cauchy solution to write the vorticity

$$\omega_n(t) = \omega_k(t=0) \frac{\partial x_n}{\partial a_k}, \quad (14.30)$$

which is one form of Kelvin's circulation theorem detailed in Section 36.3.

14.2.5 Velocity gradient tensor

Writing the Eulerian result (14.19) in component form leads to

$$\frac{D(\delta x_m)}{Dt} = \delta x_n \frac{\partial v_m}{\partial x_n}. \quad (14.31)$$

The derivatives $\partial v_m / \partial x_n$ form components to the second-order *velocity gradient* tensor whose dimensions are inverse time (i.e., a rate). The velocity gradient tensor determines how an infinitesimal material line is deformed as it moves through the fluid.

As with any matrix, a second-order tensor can be decomposed into its symmetric and anti-symmetric components

$$\frac{\partial v_m}{\partial x_n} = \frac{1}{2} \left[\frac{\partial v_m}{\partial x_n} + \frac{\partial v_n}{\partial x_m} \right] + \frac{1}{2} \left[\frac{\partial v_m}{\partial x_n} - \frac{\partial v_n}{\partial x_m} \right] \equiv \mathbb{S}_{mn} + \mathbb{A}_{mn}, \quad (14.32)$$

⁵If there is more than one way to solve a problem, then do so!

where⁶

$$\mathbb{S}_{mn} = \frac{1}{2} \left[\frac{\partial v_m}{\partial x_n} + \frac{\partial v_n}{\partial x_m} \right] = \mathbb{S}_{nm} \quad \text{rate of strain tensor} \quad (14.33a)$$

$$\mathbb{A}_{mn} = \frac{1}{2} \left[\frac{\partial v_m}{\partial x_n} - \frac{\partial v_n}{\partial x_m} \right] = -\mathbb{A}_{nm} \quad \text{rotation tensor.} \quad (14.33b)$$

As seen in the following, these tensors affect the motion of a line element in very distinct manners.

14.2.6 Stretching and tilting of a material lines

Consider a material line that is initially aligned with the vertical axis

$$\delta \mathbf{x}_{t=0} = \hat{\mathbf{z}} \delta Z_0. \quad (14.34)$$

Consequently, the initial evolution of this material line takes on the form

$$\underbrace{\frac{D(\delta x)}{Dt}}_{\text{tilting}} = \delta Z_0 \underbrace{\left[\frac{\partial u}{\partial z} \right]}_{\text{tilting}} \quad \text{and} \quad \underbrace{\frac{D(\delta y)}{Dt}}_{\text{tilting}} = \delta Z_0 \underbrace{\left[\frac{\partial v}{\partial z} \right]}_{\text{tilting}} \quad \text{and} \quad \underbrace{\frac{D(\delta z)}{Dt}}_{\text{stretching}} = \delta Z_0 \underbrace{\left[\frac{\partial w}{\partial z} \right]}_{\text{stretching}}. \quad (14.35)$$

In the presence of a vertical derivative in the horizontal velocity field (vertical shear), the first and second terms create a non-zero projection of the material line onto the horizontal plane. That is, these terms *tilt* the material line. Additionally, in the presence of a vertical derivative in the vertical velocity, the material line is expanded or compressed along its initial axis. This term is called *stretching*. We return to the tilting and stretching mechanisms when discussing the dynamics of vorticity in Chapter 36. There, we see that vortex lines in a perfect fluid flow are material lines. Consequently, vortex lines are also affected by tilting and stretching just like a material line.

14.2.7 Rate of strain tensor

Recall the expression (14.14) for the squared length of a line element

$$\delta \mathbf{X} \cdot \delta \mathbf{X} = \frac{\partial \mathbf{X}}{\partial a_i} \cdot \frac{\partial \mathbf{X}}{\partial a_j} \delta a_i \delta a_j. \quad (14.36)$$

Its material time derivative is given by

$$\left[\frac{\partial(\delta \mathbf{X} \cdot \delta \mathbf{X})}{\partial t} \right]_a = 2 \frac{\partial \mathbf{v}^L}{\partial a_i} \cdot \frac{\partial \mathbf{X}}{\partial a_j} \delta a_i \delta a_j. \quad (14.37)$$

We can express this result using Eulerian \mathbf{x} -coordinates by making use of the duality described in Section 14.1.3, which leads to

$$\frac{\partial \mathbf{v}^L}{\partial a_i} \delta a_i = \frac{\partial \mathbf{v}}{\partial x_n} \delta x_n \quad (14.38a)$$

$$\frac{\partial \mathbf{X}}{\partial a_j} \delta a_j = \delta \mathbf{x}, \quad (14.38b)$$

⁶The rate of strain tensor, \mathbb{S}_{mn} , is sometimes called the *deformation* tensor in the literature. We do not use this nomenclature to avoid confusion with the *deformation gradient* tensor defined by equation (14.13).

so that

$$\frac{D(\delta\mathbf{x} \cdot \delta\mathbf{x})}{Dt} = 2 \frac{\partial v_m}{\partial x_n} \delta x_n \delta x_m. \quad (14.39)$$

Since the product $\delta x_n \delta x_m$ is symmetric on the labels m, n , it projects out the symmetric portion of the velocity gradient tensor, thus yielding

$$\frac{1}{2} \frac{D(\delta\mathbf{x} \cdot \delta\mathbf{x})}{Dt} = \mathbb{S}_{mn} \delta x_n \delta x_m. \quad (14.40)$$

Consequently, the rate of strain tensor, \mathbb{S}_{mn} , determines the rate at which a material curve changes its length. When the rate of strain tensor vanishes, then the curve retains a constant length. We can understand the result (14.40) by considering two fluid particles initialized very close together. The distance between the two particles will be modified so long as there are nonzero gradients in the velocity field. The distance between the particles evolves according to the rate of strain tensor as given by equation (14.40).

As a symmetric matrix, the rate of strain tensor can be diagonalized, with the diagonal elements equal to the eigenvalues. Each eigenvalue measures the rate that material lines oriented according to the principle axes (eigenvectors) expand/contract under the impacts of straining motion in the fluid. According to equation (14.40), the expansion/contraction is exponential when aligned along the principle axes, with the exponential rate determined by the strain tensor's eigenvalues. Furthermore, as shown in Section 14.4, the sum of these eigenvectors (trace of the rate of strain tensor) measures the rate that a material volume changes through the divergence of the velocity

$$\mathbb{S}_{mm} = \nabla \cdot \mathbf{v}. \quad (14.41)$$

14.2.8 Rotation tensor

The rotation tensor is given by

$$\mathbb{A}_{mn} = (1/2)(\partial_n v_m - \partial_m v_n) = (1/2)(v_{m,n} - v_{n,m}), \quad (14.42)$$

with the second equality introducing the commonly used comma notation for the partial derivative; a notation that is helpful to keep the signs properly sorted with the definition. Notably, the rotation tensor is anti-symmetric:

$$\mathbb{A}_{mn} = -\mathbb{A}_{nm}. \quad (14.43)$$

Its components are related to the vorticity vector $\boldsymbol{\omega} = \nabla \wedge \mathbf{v}$ according to

$$\mathbb{A}_{mn} = -\epsilon_{mnp} \omega_p / 2 \iff \mathbb{A} = \frac{1}{2} \begin{bmatrix} 0 & -\omega_3 & \omega_2 \\ \omega_3 & 0 & -\omega_1 \\ -\omega_2 & \omega_1 & 0 \end{bmatrix} \iff \omega_p = -\epsilon_{pmn} \mathbb{A}_{mn}, \quad (14.44)$$

where the final expression made use of the identity (1.38) in the form

$$\epsilon_{mns} \epsilon_{mnp} = 2 \delta_{sp}. \quad (14.45)$$

The contribution of the rotation matrix to evolution of the material curve is given by

$$\left[\frac{D(\delta x_m)}{Dt} \right]_{\text{rot}} = \mathbb{A}_{mn} \delta x_n = -(\epsilon_{mnp} \omega_p / 2) \delta x_n \Rightarrow \left[\frac{D(\delta\mathbf{x})}{Dt} \right]_{\text{rot}} = \frac{1}{2} (\boldsymbol{\omega} \wedge \delta\mathbf{x}). \quad (14.46)$$

This relation is in the form of a pure rotation of the vector $\delta\mathbf{x}$ generated by the vector $\boldsymbol{\omega}/2$ (recall the discussion of rotations in Section 10.4). We thus conclude that the anti-symmetric rotation tensor, \mathbb{A}_{mn} , provides a rigid body rotation to a material curve (or any infinitesimal fluid region). It rotates the object without altering the size (length, area, volume). We return to this interpretation when discussing vorticity in Section 33.1.

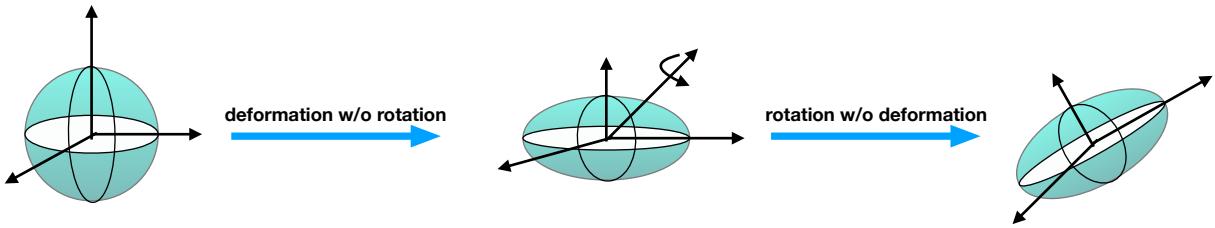


FIGURE 14.2: Schematic illustrating the Cauchy-Stokes decomposition of how fluid flow can modify a material region according to equation (14.47). First the sphere can be deformed without rotation, with this process encompassed by the rate of strain tensor, \mathbb{S}_{mn} . Next it can be rigidly rotated without changing its shape, as encompassed by the rotation tensor, \mathbb{A}_{mn} . The axes shown represent the principle axes so that deformation corresponds to expansion or contraction along the principle axes directions.

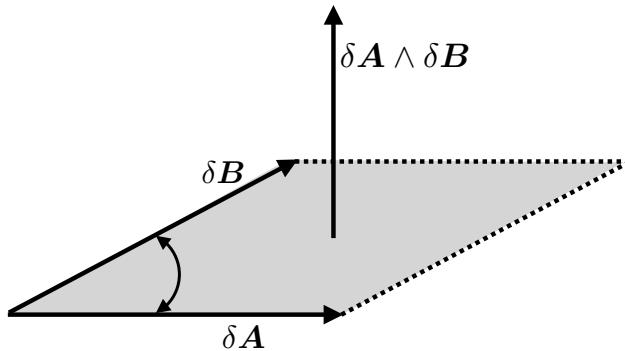


FIGURE 14.3: A material surface as defined by the cross product of two material line elements, $\delta \mathbf{S} = \delta \mathbf{A} \wedge \delta \mathbf{B}$. In the special case of $\delta \mathbf{A} = \hat{\mathbf{x}} \delta x$ and $\delta \mathbf{B} = \hat{\mathbf{y}} \delta y$, then $\delta \mathbf{S} = \delta x \delta y \hat{\mathbf{z}}$.

14.2.9 Cauchy-Stokes decomposition

The above discussion of how fluid motion impacts on a material curve falls under the more general insights from the *Cauchy-Stokes decomposition theorem*. This theorem says that the arbitrary motion of a body can be decomposed into a uniform translation, dilation along three perpendicular axes, plus a rigid body rotation. Mathematically, this decomposition can be written by expanding equation (14.31) to read

$$v_m(\mathbf{x}, t) = v_m(\mathbf{x}_0, t) + \mathbb{S}_{mn} \delta x_n + \mathbb{A}_{mn} \delta x_n \iff \mathbf{v}(\mathbf{x}, t) = \mathbf{v}(\mathbf{x}_0, t) + \mathbb{S} \cdot \delta \mathbf{x} + \mathbb{A} \cdot \delta \mathbf{x}. \quad (14.47)$$

Figure 14.2 illustrates the deformation and rotation portion of this decomposition. A more thorough discussion of these fundamental kinematic notions can be found in Chapter 4 of [Aris \(1962\)](#) as well as a brief summary in Section 1.1 of [Olbers et al. \(2012\)](#).

14.3 Evolution of material surfaces

We extend the discussion of material curves in Section 14.2 to a material surface such as that shown in Figure 14.3, with consideration given to both three-dimensional and two-dimensional flows.

14.3.1 Surfaces in three-dimensional flow

Following from the geometric interpretation of the vector product in Section 14.5, we here define a material surface by (see Figure 14.3)

$$\delta \mathbf{S} = \delta \mathbf{A} \wedge \delta \mathbf{B} \Rightarrow \delta \mathcal{S}_m = \epsilon_{mnp} \delta A_n \delta B_p \quad (14.48)$$

where $\delta\mathbf{A}$ and $\delta\mathbf{B}$ are non-parallel infinitesimal material lines. The surface projected onto the direction $\hat{\mathbf{n}}$ is given by

$$\hat{\mathbf{n}} \cdot \delta\mathbf{S} = \hat{\mathbf{n}} \cdot (\delta\mathbf{A} \wedge \delta\mathbf{B}). \quad (14.49)$$

The evolution of the material surface is given by

$$\frac{D(\delta\mathbf{S})}{Dt} = \frac{D(\delta\mathbf{A})}{Dt} \wedge \delta\mathbf{B} + \delta\mathbf{A} \wedge \frac{D(\delta\mathbf{B})}{Dt} \quad (14.50a)$$

$$= [(\delta\mathbf{A} \cdot \nabla) \mathbf{v}] \wedge \delta\mathbf{B} + \delta\mathbf{A} \wedge [(\delta\mathbf{B} \cdot \nabla) \mathbf{v}], \quad (14.50b)$$

where the second equality made use of the material line evolution equation (14.19). To proceed we expose indices and make use of some tensor identities

$$\frac{D(\delta\mathcal{S}_m)}{Dt} = \epsilon_{mnp} [(\delta A_q \partial_q) v_n] \delta B_p + \epsilon_{mnp} \delta A_n [(\delta B_q \partial_q) v_p] \quad (14.51a)$$

$$= \epsilon_{mnp} [\delta A_q \delta B_p \partial_q v_n + \delta A_n \delta B_q \partial_q v_p] \quad (14.51b)$$

$$= \epsilon_{mnp} \partial_q v_n [\delta A_q \delta B_p - \delta A_p \delta B_q] \quad (14.51c)$$

$$= \epsilon_{mnp} \partial_q v_n \epsilon_{rqp} \delta \mathcal{S}_r \quad (14.51d)$$

$$= (\delta_{mr} \delta_{nq} - \delta_{mq} \delta_{nr}) \partial_q v_n \delta \mathcal{S}_r \quad (14.51e)$$

$$= (\nabla \cdot \mathbf{v}) \delta \mathcal{S}_m - (\partial_m \mathbf{v}) \cdot \delta \mathbf{S}. \quad (14.51f)$$

To reach this result we made use of the following identities available for the permutation symbol from Chapter 1

$$\delta A_q \delta B_p - \delta A_p \delta B_q = \epsilon_{rqp} \delta \mathcal{S}_r \quad (14.52a)$$

$$\epsilon_{mnp} \epsilon_{rqp} = \delta_{mr} \delta_{nq} - \delta_{mq} \delta_{nr} \quad (14.52b)$$

where δ_{mn} is the Kronecker (or identity) tensor.

14.3.2 Evolution of the area

Now orient the material surface area according to its outward normal vector

$$\delta\mathbf{S} = \hat{\mathbf{n}} \delta\mathcal{S} \implies \delta\mathcal{S}_m = \hat{n}_m \delta\mathcal{S}. \quad (14.53)$$

Doing so brings equation (14.51f) to the form

$$\frac{1}{\delta\mathcal{S}} \frac{D(\hat{n}_m \delta\mathcal{S})}{Dt} = (\nabla \cdot \mathbf{v}) \hat{n}_m - (\partial_m \mathbf{v}) \cdot \hat{\mathbf{n}}. \quad (14.54)$$

We can thus develop evolution equations for the surface area, $\delta\mathcal{S}$, and the normal vector, $\hat{\mathbf{n}}$.

For the surface area evolution we take the inner product with \hat{n}_m to yield

$$\frac{1}{\delta\mathcal{S}} \frac{D\delta\mathcal{S}}{Dt} = \nabla \cdot \mathbf{v} - [(\hat{\mathbf{n}} \cdot \nabla) \mathbf{v}] \cdot \hat{\mathbf{n}} \quad (14.55)$$

where we set $\hat{\mathbf{n}} \cdot \hat{\mathbf{n}} = 1$ and followed the discussion in Section 2.1.5 to find that

$$\hat{\mathbf{n}} \cdot \hat{\mathbf{n}} = 1 \implies \hat{\mathbf{n}} \cdot \frac{D\hat{\mathbf{n}}}{Dt} = 0, \quad (14.56)$$

so that the normal vector is always perpendicular to its material time derivative. Rearrangement of equation (14.55) then leads to the kinematic evolution equation for the area

$$\frac{1}{\delta\mathcal{S}} \frac{D\delta\mathcal{S}}{Dt} = [\partial_m - \hat{n}_m (\hat{\mathbf{n}} \cdot \nabla)] v_m. \quad (14.57)$$

We next provide some interpretation of this result.

Surface derivative operator

The derivative operator on the right hand side of equation (14.57),

$$\partial_m^{\text{surf}} \equiv \partial_m - \hat{n}_m (\hat{n} \cdot \nabla), \quad (14.58)$$

is a *surface derivative operator* since it subtracts from the gradient operator the projection onto the local normal, thus leaving a gradient operator just in the tangent plane of the surface. The area evolution equation (14.57) can thus be written in the tidy form

$$\frac{1}{\delta S} \frac{D\delta\mathcal{S}}{Dt} = \nabla^{\text{surf}} \cdot \mathbf{v}, \quad (14.59)$$

so that the relative area of the material surface changes according to the surface divergence of the velocity field.

Special case of a horizontal surface

To help understand the kinematic equation (14.59), consider the special case of a horizontal surface with a vertical normal vector, $\hat{n} = \hat{z}$, so that

$$\frac{1}{\delta S} \frac{D\delta\mathcal{S}}{Dt} = \nabla \cdot \mathbf{v} - \hat{z} \cdot \partial_z \mathbf{v} = \nabla_z \cdot \mathbf{v}. \quad (14.60)$$

In this case we see that the area of a horizontal surface increases when the horizontal velocity diverges, and the area decreases when the velocity converges. We expect this behavior since the surface is material and thus is moving with the flow. We encounter this result again in Section 14.3.4 for two-dimensional flow, in which the area is always horizontal.

As another special case, consider a three dimensional flow that is non-divergent, $\nabla \cdot \mathbf{v} = 0$ (see Chapter 17). In this case the area changes are only due to the projection of the normal gradient onto the normal direction. So again considering a horizontal area with $\hat{n} = \hat{z}$, the area evolution in a non-divergent flow is given by

$$\frac{1}{\delta S} \frac{D\delta\mathcal{S}}{Dt} = -\hat{z} \cdot (\hat{z} \cdot \nabla) \mathbf{v} = -\partial_z w, \quad (14.61)$$

which follows from equation (14.60) with $\nabla_x u + \partial_y v + \partial_z w = \nabla_z \mathbf{u} + \partial_z w = 0$.

14.3.3 Evolution of the normal vector

We make use of the area evolution equation (14.59) within equation (14.54) to derive an evolution equation for the normal vector

$$\frac{D\hat{n}_m}{Dt} = -\hat{n}_m \frac{1}{\delta S} \frac{D\delta\mathcal{S}}{Dt} + (\partial_p v_p) \hat{n}_m - (\partial_m v_p) \hat{n}_p \quad (14.62a)$$

$$= -\hat{n}_m [\partial_p v_p - \hat{n}_p (\hat{n}_q \partial_q) v_p] + (\partial_p v_p) \hat{n}_m - (\partial_m v_p) \hat{n}_p \quad (14.62b)$$

$$= -\hat{n}_p [\partial_m - \hat{n}_m \hat{n}_q \partial_q] v_p \quad (14.62c)$$

$$= -\hat{n}_p \partial_m^{\text{surf}} v_p \quad (14.62d)$$

$$= -\hat{n} \cdot \partial_m^{\text{surf}} \mathbf{v}. \quad (14.62e)$$

We can readily verify that this expression satisfies the kinematic constraint (14.56) in which $\hat{n} \cdot D\hat{n}/Dt = 0$.

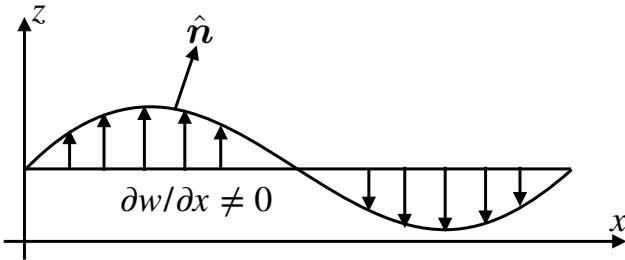


FIGURE 14.4: Horizontal shear in the vertical velocity, $\nabla_z w$, creates undulations in an initially horizontal material surface that leads to a horizontal component in the normal vector. We here show the case where $\partial w / \partial x \neq 0$, thus leading to a zonal component to the normal vector according to $D\hat{n}_x/Dt = -\partial_x w$.

To help understand the evolution equation (14.62e) consider again a horizontal area with its normal vector initially in the vertical direction. The evolution of this normal vector is thus given by

$$\frac{D\hat{n}_m}{Dt} = -(\partial_m - \hat{z}_m \partial_z) w, \quad (14.63)$$

with each component evolving according to

$$\frac{D\hat{n}_1}{Dt} = -\partial_x w \quad (14.64a)$$

$$\frac{D\hat{n}_2}{Dt} = -\partial_y w \quad (14.64b)$$

$$\frac{D\hat{n}_3}{Dt} = 0. \quad (14.64c)$$

Hence, an initially vertical normal vector tilts into the horizontal direction according to minus the horizontal shear in the vertical velocity. We understand this result by noting that such a shear creates undulations in the initially horizontal surface that render a horizontal component to the normal vector (Figure 14.4).

14.3.4 Material area in two-dimensional flow

Now consider a material area for two-dimensional fluid flow with velocity, $\mathbf{v} = (u, v, 0)$, and $\delta\mathbf{A} = \hat{x}\delta x, \delta\mathbf{B} = \hat{y}\delta y$, with zero dependence on z . In this case, the area of an infinitesimal material region is

$$\delta\mathcal{S} = (\delta\mathbf{A} \wedge \delta\mathbf{B}) \cdot \hat{z} = \delta x \delta y, \quad (14.65)$$

and its evolution is given by

$$\frac{D(\delta\mathcal{S})}{Dt} = (\delta\mathbf{B} \wedge \hat{z}) \cdot (\delta\mathbf{A} \cdot \nabla) \mathbf{u} + (\hat{z} \wedge \delta\mathbf{A}) \cdot (\delta\mathbf{B} \cdot \nabla) \mathbf{u} \quad (14.66a)$$

$$= \delta x \delta y \nabla \cdot \mathbf{u}, \quad (14.66b)$$

so that

$$\frac{1}{\delta\mathcal{S}} \frac{D(\delta\mathcal{S})}{Dt} = \nabla \cdot \mathbf{u}. \quad (14.67)$$

Hence, the area of the material region evolves according to the divergence of the horizontal velocity. Correspondingly, the area remains constant in a horizontally non-divergent flow. This result follows from specializing the three-dimensional result (14.51f) to the case of two-dimensional flow by assuming no dependence on the vertical direction.

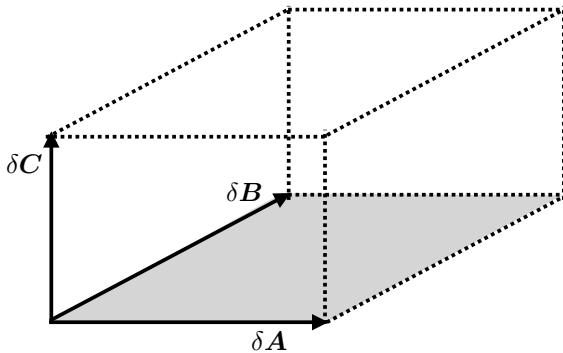


FIGURE 14.5: A parallelepiped defined by three material lines with volume (to within a sign) given by $\delta V = (\delta \mathbf{A} \wedge \delta \mathbf{B}) \cdot \delta \mathbf{C}$. See also the discussion surrounding Figure 14.4.

14.4 Volume and the Jacobian of transformation

The mass of a material parcel is constant. However, the volume is not generally constant since the fluid density is not generally uniform. We here derive the expression for how volume evolves for a material parcel. We also derive the material evolution equation for the Jacobian of transformation between position space and material space. We will see that the relative change for both the parcel volume and the Jacobian are determined by the divergence of the velocity field.

14.4.1 Material parcel volume

Consider a material region with a volume δV spanned by the infinitesimal material lines $\delta \mathbf{A}$, $\delta \mathbf{B}$, and $\delta \mathbf{C}$ (see Figure 14.5). To within a sign the volume is given by

$$\delta V = (\delta \mathbf{A} \wedge \delta \mathbf{B}) \cdot \delta \mathbf{C}. \quad (14.68)$$

Making use of the material line evolution equation (14.19) renders

$$\frac{D(\delta V)}{Dt} = (\delta \mathbf{B} \wedge \delta \mathbf{C}) \cdot (\delta \mathbf{A} \cdot \nabla) \mathbf{v} + (\delta \mathbf{C} \wedge \delta \mathbf{A}) \cdot (\delta \mathbf{B} \cdot \nabla) \mathbf{v} + (\delta \mathbf{A} \wedge \delta \mathbf{C}) \cdot (\delta \mathbf{C} \cdot \nabla) \mathbf{v}. \quad (14.69)$$

Now specialize to the case where the parcel is a parallelepiped oriented according to the coordinate axes

$$\delta \mathbf{A} = \hat{\mathbf{x}} \delta x \quad \text{and} \quad \delta \mathbf{B} = \hat{\mathbf{y}} \delta y \quad \text{and} \quad \delta \mathbf{C} = \hat{\mathbf{z}} \delta z, \quad (14.70)$$

so that

$$\delta V = \delta x \delta y \delta z. \quad (14.71)$$

Plugging into equation (14.69) leads to

$$\frac{1}{\delta V} \frac{D(\delta V)}{Dt} = \nabla \cdot \mathbf{v}. \quad (14.72)$$

This result is a three-dimensional generalization of the material area equation derived in Section 14.3.4.

We offer an alternative derivation of equation (14.72) in Section 15.2, where no assumptions are made concerning the shape of the material region. That derivation leads us to conclude that the relative volume of a material parcel increases when the parcel moves through a region where the velocity diverges ($\nabla \cdot \mathbf{v} > 0$). We think of a diverging velocity field as “spreading out” the material

parcel boundary, thus increasing its volume. In contrast, the volume of a material parcel decreases where the fluid velocity converges ($\nabla \cdot \mathbf{v} < 0$)

$$\nabla \cdot \mathbf{v} > 0 \Rightarrow \text{material volume increases in diverging flow} \Rightarrow \text{parcel expands} \quad (14.73\text{a})$$

$$\nabla \cdot \mathbf{v} < 0 \Rightarrow \text{material volume decreases in converging flow} \Rightarrow \text{parcel contracts.} \quad (14.73\text{b})$$

14.4.2 Evolution of the Jacobian of transformation

Recall the discussion in Section 13.4.6, where we showed that the Jacobian of transformation between material space (\mathbf{a} -space) and position space (\mathbf{x} -space) is related to the ratio of the volume elements written in the two coordinate systems. In particular, equation (13.22) is given by

$$\delta V(\mathbf{x}) = \frac{\partial \mathbf{X}}{\partial \mathbf{a}} \delta V(\mathbf{a}) \Rightarrow \frac{\partial \mathbf{X}}{\partial \mathbf{a}} = \frac{\delta V(\mathbf{x})}{\delta V(\mathbf{a})}. \quad (14.74)$$

The material coordinate volume $\delta V(\mathbf{a})$ is time independent when following the flow. Consequently, the material evolution of the Jacobian is given by

$$\frac{D}{Dt} \frac{\partial \mathbf{X}}{\partial \mathbf{a}} = \frac{1}{\delta V(\mathbf{a})} \frac{D(\delta V(\mathbf{x}))}{Dt} \quad (14.75\text{a})$$

$$= \frac{\delta V(\mathbf{x})}{\delta V(\mathbf{a})} \nabla \cdot \mathbf{v} \quad (14.75\text{b})$$

$$= \frac{\partial \mathbf{X}}{\partial \mathbf{a}} \nabla \cdot \mathbf{v}. \quad (14.75\text{c})$$

The second equality made use of the equation (14.72), which expresses the material time change for the volume of a material fluid parcel, as measured in position space, in terms of the velocity divergence. We thus see that the relative change of the Jacobian is determined by the divergence of the velocity

$$\left[\frac{\partial \mathbf{X}}{\partial \mathbf{a}} \right]^{-1} \frac{D}{Dt} \left[\frac{\partial \mathbf{X}}{\partial \mathbf{a}} \right] = \nabla \cdot \mathbf{v}. \quad (14.76)$$

This equation is identical to the parcel volume equation (14.72), which is expected given the relation between the Jacobian and the parcel volume. In Exercise 14.4, we derive this result using the explicit expression for the Jacobian in terms of the permutation symbol ϵ .

14.5 Kinematics of two-dimensional flow

In this section we consider the rudiments of two-dimensional flow as a venue to illustrate certain topics presented earlier in this chapter such as dilation, rotation, and strains. In so doing we expose kinematic properties commonly used to characterize two-dimensional flow, with generalizations to three-dimensions available with a bit more mathematics.

Our starting point is Figure 14.6, which shows a square region of fluid exposed to a variety of flow regimes. We can kinematically describe these changes by making use of the velocity gradient tensor introduced in Section 14.2.5, here written for the two-dimensional flow with horizontal

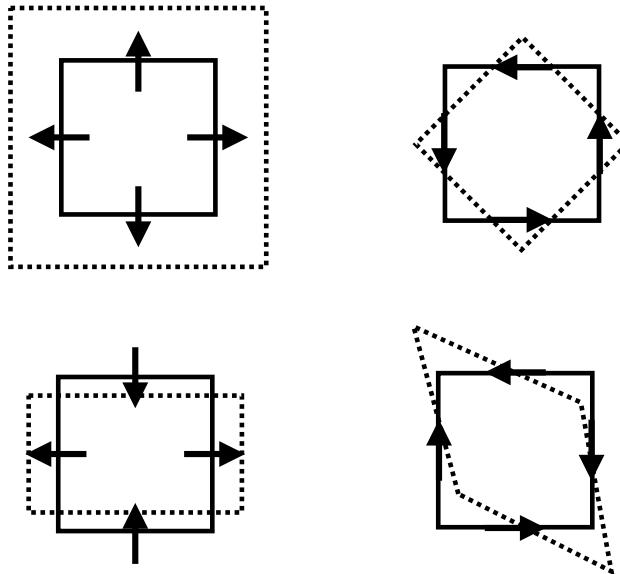


FIGURE 14.6: Illustrating the varieties of shape changes for a square material fluid region. Upper left: purely divergent flow, whereby $\nabla \cdot \mathbf{u} > 0$ yet with zero vorticity, thus leading to a uniform increase in the area. Upper right: rotational flow with nonzero vorticity, $\zeta = \hat{\mathbf{z}} \cdot (\nabla \wedge \mathbf{u})$, yet zero divergence, thus leading to a pure rotation of the square patch. Lower left: result of a pure tension/compression straining flow with zero divergence and zero vorticity, leading to compression in one direction and dilation in the orthogonal direction. This flow is realized by the velocity field $\mathbf{u} = \hat{\mathbf{z}} \wedge \nabla \psi$, which is generated by the streamfunction $\psi = -\gamma xy$. Lower right: pure shearing strain flow with zero divergence and zero vorticity, as as realized by the flow $\mathbf{u} = \hat{\mathbf{z}} \wedge \nabla \psi$ generated by the streamfunction $\psi = -(\gamma/2)(x^2 - y^2)$. This figure is adapted from Figure 2.4 of [Hoskins and James \(2014\)](#).

velocity components, (u, v)

$$\frac{\partial v_m}{\partial x_n} = \begin{bmatrix} \partial u / \partial x & \partial u / \partial y \\ \partial v / \partial x & \partial v / \partial y \end{bmatrix} \quad (14.77a)$$

$$= \begin{bmatrix} \partial u / \partial x & (1/2)(\partial u / \partial y + \partial v / \partial x) \\ (1/2)(\partial u / \partial y + \partial v / \partial x) & \partial v / \partial y \end{bmatrix} + \frac{\zeta}{2} \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \quad (14.77b)$$

$$= \mathbb{S}_{mn} + \mathbb{A}_{mn}, \quad (14.77c)$$

where

$$\zeta = \hat{\mathbf{z}} \cdot \nabla \wedge \mathbf{v} = \partial v / \partial x - \partial u / \partial y \quad (14.78)$$

is the vertical component to the vorticity.

14.5.1 Diverging flow

Recall from Section 14.3.4 that a material surface in two-dimensional flow changes its area according to the divergence. The upper left panel of Figure 14.6 thus illustrates equation (14.67)

$$\frac{1}{\delta \mathcal{S}} \frac{D(\delta \mathcal{S})}{Dt} = \nabla \cdot \mathbf{u} = \mathbb{S}_{mm}, \quad (14.79)$$

where $\delta \mathcal{S}$ is the area and \mathbb{S}_{mm} is the trace of the rate of strain tensor. That is, a diverging flow as depicted by this figure, with $\nabla \cdot \mathbf{u} > 0$, leads to an expansion of the area. The opposite occurs for a converging flow, where the area compresses.

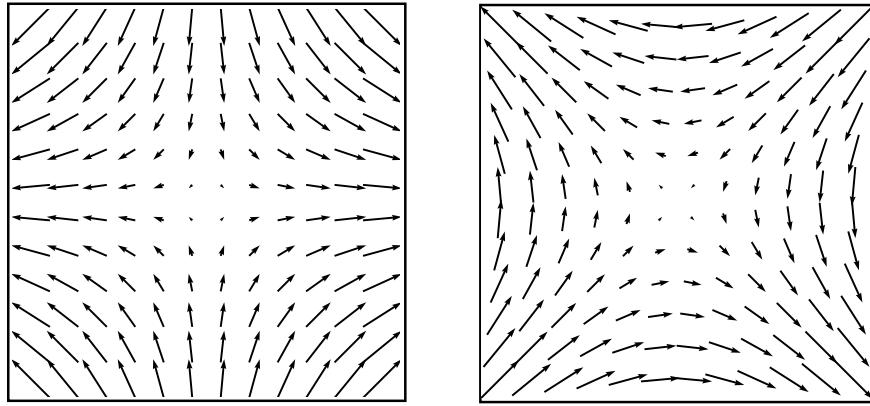


FIGURE 14.7: Two-dimensional non-divergent and irrotational flow with nonzero deformation/strain. Left panel: pure tension strain as determined by the streamfunction, $\psi = -\gamma xy$, so that the velocity components are $u = -\partial\psi/\partial y = \gamma x$ and $v = \partial\psi/\partial x = -\gamma y$. The vertical axis orients the direction along which flow contracts (compression) whereas the horizontal axis is the dilation axis (tension). Right panel: pure shearing flow as determined by the streamfunction $\psi = -(\gamma/2)(x^2 - y^2)$ so that the velocity components are $u = -\gamma y$ and $v = -\gamma x$. We set $\gamma = 1$ for both examples.

14.5.2 Rotational flow with nonzero vorticity

The upper right panel of Figure 14.6 illustrates the effects from a flow with a non-zero vorticity $\zeta = \partial v/\partial x - \partial u/\partial y$, or a nonzero circulation. We provide extensive discussion of vorticity and circulation in Part VI of this book. For now we remain satisfied by noting that vorticity measures the spin at a point within a fluid. The nonzero spin imparts a rotation to an area element, in this case bringing about a counter-clockwise rotation. All components of the strain tensor vanish for a purely rotational flow, so that there is no deformation of the square as it rotates.

14.5.3 Flow with nonzero deformation

The lower left panel of Figure 14.6 shows the square within a deformational flow whereby it contracts along the y -axis and dilates along the x -axis. This flow is non-divergent, $\nabla \cdot \mathbf{u} = 0$, and has zero vorticity, $\zeta = 0$, so that the area remains constant and the orientation is fixed. However, it has shear that acts to deform the area. This particular non-divergent deformational flow is determined by

$$\mathbf{u} = \hat{\mathbf{z}} \wedge \nabla \psi, \quad (14.80)$$

with the streamfunction, $\psi = -\gamma xy$ where γ is a constant inverse time scale and hence the strength of the strain. The resulting velocity components are $u = -\partial\psi/\partial y = \gamma x$ and $v = \partial\psi/\partial x = -\gamma y$.

There are two combinations of the rate of strain tensor elements that are useful in describing deformational flows:

$$\text{tension strain} = S_T = \mathbb{S}_{11} - \mathbb{S}_{22} = \partial u / \partial x - \partial v / \partial y \quad (14.81a)$$

$$\text{shearing strain} = S_S = 2 \mathbb{S}_{12} = \partial u / \partial y + \partial v / \partial x. \quad (14.81b)$$

The tension strain and shearing strain are also known as tension and shearing *deformation rates*. Note that negative tension is known as *compression*. For the deformation flow with streamfunction $\psi = -\gamma xy$, we have

$$S_T = 2\gamma \quad \text{and} \quad S_S = 0, \quad (14.82)$$

so that this velocity leads to a purely tension straining flow. In contrast, the following non-divergent irrotational flow is a purely shearing strain flow

$$\psi = -(\gamma/2)(x^2 - y^2) \quad u = -\gamma y \quad v = -\gamma x \quad S_T = 0 \quad S_S = -2\gamma, \quad (14.83)$$

as depicted by the right panel of Figure 14.7. This pure shearing flow leads to the deformation of the fluid square shown in the lower right panel of Figure 14.6.

14.5.4 Further study

Elements of this section can be found in Section 2.3 of [Hoskins and James \(2014\)](#). More detailed examinations of two-dimensional flow kinematics are offered by [Weiss \(1991\)](#) and [Lilly \(2018\)](#).

14.6 Exercises

EXERCISE 14.1: VELOCITY FIELD WITH ZERO STRAIN ([Aris \(1962\)](#) EXERCISE 4.41.1)

If the rate of strain tensor vanishes, show that the velocity field can be written

$$\mathbf{v} = \mathbf{U} + \boldsymbol{\Omega} \wedge \mathbf{x}, \quad (14.84)$$

where $\boldsymbol{\Omega}$ is a constant angular rotation rate and \mathbf{U} is a constant velocity. That is, a fluid velocity equal to a constant rotation plus translation renders zero strain. Hint: if $S_{mn} = 0$, what does that imply about the velocity field? You may also wish to make use of the general decomposition (14.47).

EXERCISE 14.2: RATE OF STRAIN AND ROTATION TENSORS ([Aris \(1962\)](#) EXERCISE 4.43.3)

Consider a two-dimensional flow with horizontal velocity

$$\mathbf{u} = (f/r)(\hat{\mathbf{x}}y - \hat{\mathbf{y}}x), \quad (14.85)$$

where $f = f(r)$ is an arbitrary function of the radial distance $r = \sqrt{x^2 + y^2}$. Note: f is here *not* the Coriolis parameter found elsewhere in this book, but it is instead just an arbitrary function of the radial distance.

- (a) Show that the velocity field is non-divergent.
- (b) Determine the streamlines.
- (c) Determine the elements to the rate of strain tensor, \mathbb{S} , given by equation (14.77c). Write the expression using polar coordinates $x = r \cos \varphi$ and $y = r \sin \varphi$ (see Section 8.3) and the structure function $F(r) = [f' - f/r]/2$, where $f' = df/dr$.
- (d) Determine elements to the rotation tensor, \mathbb{A} , given by equation (14.77c), also written in polar coordinates.

EXERCISE 14.3: RATE OF STRAIN AND ROTATION FOR PARALLEL SHEAR FLOW

Consider a two-dimensional parallel shear flow with horizontal velocity

$$\mathbf{u} = a x \hat{\mathbf{y}}, \quad (14.86)$$

where a is a constant with dimension inverse time.

- (a) Compute the rate of strain tensor, \mathbb{S} (equation (14.77c)) for this velocity field.
- (b) Compute the rotation tensor, \mathbb{A} (equation (14.77c)) for this velocity field.
- (c) Decompose the velocity field according to equation (14.47), and show that each of the velocity components is non-divergent. That is, write

$$\mathbf{u} = \mathbf{u}^0 + \mathbf{u}^{(S)} + \mathbf{u}^{(A)} \quad \text{with} \quad \nabla \cdot \mathbf{u}^{(S)} = \nabla \cdot \mathbf{u}^{(A)} = 0, \quad (14.87)$$

with u_m^0 the velocity at the point where $\delta x_m = x_m - x_m^0 = 0$. The velocity $\mathbf{u}^{(S)}$ has a constant strain but no vorticity whereas $\mathbf{u}^{(A)}$ has a constant vorticity but no strain. Hint: $\mathbf{u}^{(S)}$ and $\mathbf{u}^{(A)}$ have nonzero \hat{x} and \hat{y} components.

- (d) Determine the streamfunctions for $\mathbf{u}^{(S)}$ and $\mathbf{u}^{(A)}$.
- (e) Sketch the velocity vectors $\mathbf{u}^{(S)}$ and $\mathbf{u}^{(A)}$.

EXERCISE 14.4: EVOLUTION OF THE JACOBIAN USING ϵ -TENSOR GYMNASTICS

There is another way to derive the identity (14.76) for the evolution of the Jacobian. This other method is somewhat more tedious. However, it exercises some useful methods of index gymnastics involving the ϵ -tensor. It also has a natural generalization to curved spaces. This exercise is only for aficionados of tensor analysis.

An explicit expression for the Jacobian of transformation is given by

$$\frac{\partial \mathbf{X}}{\partial \mathbf{a}} = \frac{1}{3!} \epsilon_{mnp} \epsilon_{ijk} \frac{\partial X_m}{\partial a_i} \frac{\partial X_n}{\partial a_j} \frac{\partial X_p}{\partial a_k}. \quad (14.88)$$

Take the material derivative of this expression and show that we get the same expression as equation (14.76). Hint: make use of the identity

$$\frac{D}{Dt} \frac{\partial X_m}{\partial a_i} = \frac{\partial V_m}{\partial a_i}, \quad (14.89)$$

which holds since the material time derivative is taken with the material coordinates, \mathbf{a} , held fixed.



Mass conservation

We assume that mass is neither created nor destroyed within the fluid domain. Since this assumption holds independently of the forces acting on the fluid, mass conservation is a topic within fluid kinematics. In this chapter, we derive the mathematical expressions for mass conservation in a homogeneous fluid and the associated kinematic constraints placed on fluid motion. These constraints are examined both in the interior of the fluid as well as at fluid boundaries.

READER'S GUIDE TO THIS CHAPTER

In this chapter we are concerned with homogeneous fluids, with generalizations to multiple-component fluids given in Chapter 16. Spatial positions and trajectories are represented using Cartesian coordinates to simplify the maths, with generalizations following the tensor analysis methods of Chapter 7. We build on our understanding of the Eulerian and Lagrangian kinematic descriptions developed in Chapters 13 and 14.

15.1	Eulerian fluid regions	290
15.1.1	Finite volume expression	290
15.1.2	Arbitrary Eulerian region	291
15.2	Material fluid parcels	292
15.2.1	Lagrangian expression for mass conservation	292
15.2.2	Alternative derivation based on the Jacobian	293
15.2.3	Summary of material kinematic equations	293
15.3	Material fluid regions	293
15.3.1	Evolution of volume	293
15.3.2	Mass conservation	294
15.3.3	Mass conservation using Lagrangian methods	295
15.3.4	Reynolds transport theorem	296
15.3.5	Comments on notation for the time derivative	297
15.4	Kinematic boundary conditions	298
15.4.1	Static material surface	298
15.4.2	Moving material surface	299
15.4.3	Dynamic and permeable surface	302
15.5	Volume and mass budgets for an ocean column	305
15.5.1	Kinematic free surface equation	305
15.5.2	Budget for mass per horizontal area	306
15.6	Exercises	307

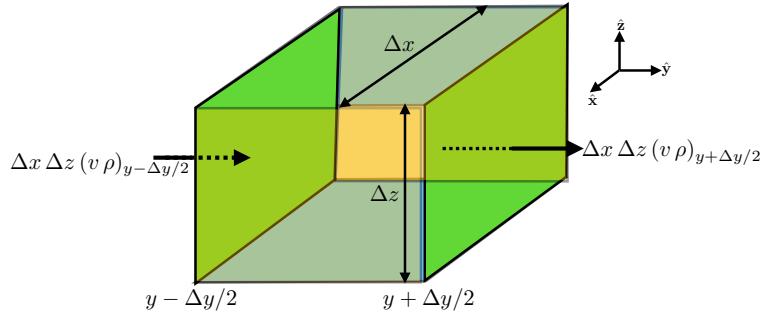


FIGURE 15.1: A finite sized cube (also known as a cell) of fixed dimensions and position (an Eulerian region) used to formulate the Eulerian form of mass conservation. We highlight two cell faces with area $\Delta x \Delta z$ and with meridional mass transport crossing the faces given by $\Delta x \Delta z (v \rho)_{y-\Delta y/2}$ and $\Delta x \Delta z (v \rho)_{y+\Delta y/2}$. To establish signs we assume the meridional velocity is positive, $v > 0$, so that mass enters the face at $y - \Delta y/2$ and leaves the face at $y + \Delta y/2$. Differences between these two transports leads to an accumulation of mass within the cell. Note that the resulting mass budget holds regardless the direction of the flow velocity.

15.1 Eulerian fluid regions

We here develop expressions for the mass budget within an Eulerian region, both infinitesimal and finite. Recall that Eulerian regions are fixed in space and thus have constant volumes.

15.1.1 Finite volume expression

Consider a finite sized cube region that is fixed in space; i.e., an Eulerian region such as shown in Figure 15.1. The mass contained within the cube is given by

$$\Delta M = \rho \Delta V = \rho \Delta x \Delta y \Delta z, \quad (15.1)$$

where the cube volume,

$$\Delta V = \Delta x \Delta y \Delta z, \quad (15.2)$$

is constant in time. As we will be taking the limit as the size of the cube gets smaller, it is sufficient to approximate the density as that at the cube center, $\rho = \rho(x, y, z, t)$. In the absence of mass sources within the fluid, the mass within the cube changes only through the transport of mass across the six cube faces.

Focusing on the mass transported in the meridional direction as illustrated in Figure 15.1, the accumulation of mass within the cube through this transport is determined by the difference in mass transport crossing the two adjacent cell faces

$$\text{mass change from meridional transport} = (\Delta x \Delta z) [(v \rho)_{y-\Delta y/2} - (v \rho)_{y+\Delta y/2}]. \quad (15.3)$$

Expanding the difference into a Taylor series leads to

$$\text{mass change from meridional transport} \approx -(\Delta x \Delta y \Delta z) \frac{\partial(v \rho)}{\partial y}. \quad (15.4)$$

The same analysis for the zonal and vertical directions leads to the mass budget for the cube

$$\frac{\partial(\rho \Delta V)}{\partial t} = -\Delta V \left[\frac{\partial(u \rho)}{\partial x} + \frac{\partial(v \rho)}{\partial y} + \frac{\partial(w \rho)}{\partial z} \right]. \quad (15.5)$$

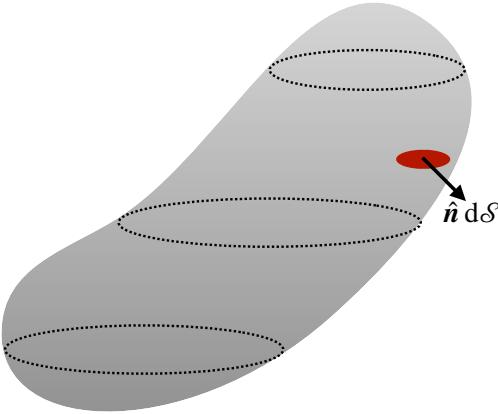


FIGURE 15.2: An arbitrarily shaped simply closed region, \mathcal{R} , within the fluid. If the region is fixed in space, then it represents a general Eulerian region for considering mass budgets. A surface area element, $d\mathcal{S}$, is oriented according to the outward normal, $\hat{\mathbf{n}}$.

Hence, the cube mass changes according to the convergence of mass across the cube boundaries. Cancelling the constant volume ΔV (again, the volume is assumed fixed as per an Eulerian region) renders the *flux-form* Eulerian mass continuity equation

$$\frac{\partial \rho}{\partial t} + \nabla \cdot (\rho \mathbf{v}) = 0. \quad (15.6)$$

The mass continuity equation (15.6) is in the form of a *flux-form conservation law*, in which the local time tendency of a field is determined by the convergence of a flux

$$\frac{\partial \rho}{\partial t} = -\nabla \cdot (\rho \mathbf{v}). \quad (15.7)$$

The *mass flux*, $\rho \mathbf{v}$, with dimensions $M L^{-2} T^{-1}$, measures the mass per time of matter crossing a unit area oriented with an outward normal in each of the three Cartesian directions. If more mass flux comes into a point than leaves, then the density increases (mass converges), and vice versa for a mass flux that diverges from a point. It is notable that the mass flux also serves as a measure of the *momentum flux*, which we discuss in Chapter 18.

15.1.2 Arbitrary Eulerian region

The discussion for the infinitesimal cube can be generalized by making use of the divergence theorem. For that purpose, consider an arbitrary static and simply closed region within the fluid such as in Figure 15.2. Integrating the continuity equation (15.6) over that region leads to

$$\int_{\mathcal{R}} \frac{\partial \rho}{\partial t} dV = - \int_{\mathcal{R}} \nabla \cdot (\rho \mathbf{v}) dV. \quad (15.8)$$

Since the region is static we can move the partial time derivative outside on the left hand side. Furthermore, the divergence theorem can be applied to the right hand side to convert the volume integral to a surface integral over the boundaries of the static domain. The resulting mass budget is given by

$$\frac{d}{dt} \int_{\mathcal{R}} \rho dV = - \oint_{\partial \mathcal{R}} \rho \mathbf{v} \cdot \hat{\mathbf{n}} d\mathcal{S}, \quad (15.9)$$

where $\hat{\mathbf{n}}$ is the outward normal vector along the closed boundary of the region, and $d\mathcal{S}$ is the surface area element along that boundary. This equation says that the mass within a fixed region of the fluid changes in time (left hand side) according to the accumulation of mass crossing the region boundary (right hand side). The minus sign means that the mass decreases in the region if there is a flux of mass leaving the domain in the direction of the outward normal.

15.2 Material fluid parcels

We here derive the differential expressions for mass conservation of a constant mass fluid parcel within a Lagragian reference frame. The differential expressions for volume and density arise as a corollary. This discussion complements the Eulerian discussion from Section 15.1. To motivate the derivations we expand the Eulerian flux-form mass continuity equation (15.6) to have

$$\frac{1}{\rho} \frac{D\rho}{Dt} = -\nabla \cdot \mathbf{v}, \quad (15.10)$$

where $D/Dt = \partial_t + \mathbf{v} \cdot \nabla$ is the material time derivative operator. We now derive this form of the mass continuity using Lagrangian methods.

15.2.1 Lagrangian expression for mass conservation

The mass of an infinitesimal fluid parcel is written

$$\delta M = \rho \delta V, \quad (15.11)$$

where δV is the volume and

$$\rho = \frac{\delta M}{\delta V} \quad (15.12)$$

is the mass density of the parcel. By definition, the fluid parcel has a constant mass so that its material time derivative vanishes

$$\frac{D(\delta M)}{Dt} = 0. \quad (15.13)$$

Equation (15.13) is the most basic form of mass conservation for a material parcel. However, one often has need to express this result in terms of parcel density and parcel volume

$$\frac{D(\delta M)}{Dt} = \frac{D(\rho \delta V)}{Dt} \quad (15.14a)$$

$$= \delta M \left[\frac{1}{\rho} \frac{D\rho}{Dt} + \frac{1}{\delta V} \frac{D(\delta V)}{Dt} \right]. \quad (15.14b)$$

Comparing to the mass continuity equation (15.10) leads to¹

$$\frac{1}{\delta M} \frac{D(\delta M)}{Dt} = \frac{1}{\rho} \frac{D\rho}{Dt} + \nabla \cdot \mathbf{v}. \quad (15.15)$$

Setting $D(\delta M)/Dt = 0$ renders the continuity equation (15.10) derived from the Eulerian expression

$$\frac{1}{\rho} \frac{D\rho}{Dt} = -\nabla \cdot \mathbf{v}. \quad (15.16)$$

The parcel volume contracts in regions where the velocity converges (we prove that property in Sections 14.4.1 and 15.3.1). The continuity equation (15.16) then says that regions of volume contraction are where the parcel density increases whereas the opposite occurs for regions where the velocity diverges.

¹In Section 14.4.1 we derived the material evolution of volume, and will again see this result in equation (15.26).

15.2.2 Alternative derivation based on the Jacobian

An alternative approach to deriving the mass conservation equation makes use of the material time evolution of the Jacobian. We detail this approach for integrals in Section 15.3.3, and discussed it for fluid parcels in Section 14.4.2. Making use of those results, in particular equation (14.76), allows us to write the material evolution for the fluid parcel mass in the form

$$\frac{D}{Dt} [\rho \delta V(\mathbf{x})] = \frac{D}{Dt} \left[\rho \frac{\partial \mathbf{X}}{\partial \mathbf{a}} \delta V(\mathbf{a}) \right] \quad (15.17a)$$

$$= \left[\frac{D\rho}{Dt} + \rho \nabla \cdot \mathbf{v} \right] \left[\frac{\partial \mathbf{X}}{\partial \mathbf{a}} \right] \delta V(\mathbf{a}) \quad (15.17b)$$

$$= \left[\frac{D\rho}{Dt} + \rho \nabla \cdot \mathbf{v} \right] \delta V(\mathbf{x}). \quad (15.17c)$$

We recover the mass conservation equation (15.16) when noting that the mass of a material parcel is constant so that $D\rho/Dt = -\rho \nabla \cdot \mathbf{v}$.

15.2.3 Summary of material kinematic equations

Let us now summarize the variety of differential evolution equations for mass, volume, and density as viewed from a material reference frame

$$\frac{D(\delta M)}{Dt} = 0 \quad \text{parcel mass is constant} \quad (15.18)$$

$$\frac{1}{\delta V} \frac{D(\delta V)}{Dt} = \nabla \cdot \mathbf{v} \quad \text{parcel volume increases in divergent flow} \quad (15.19)$$

$$\frac{1}{\rho} \frac{D\rho}{Dt} = -\nabla \cdot \mathbf{v} \quad \text{parcel density increases in convergent flow.} \quad (15.20)$$

To help remember the signs on the right hand side, recall that as the fluid diverges from a point ($\nabla \cdot \mathbf{v} > 0$), it expands the boundaries of the material parcel and so increases the parcel volume as per equation (15.19). Since the parcel has a fixed mass, the diverging velocity field causes the material parcel density to decrease ($-\nabla \cdot \mathbf{v} < 0$) as per equation (15.20).

15.3 Material fluid regions

We now extend the kinematics of material fluid parcels in Section 15.2 to finite sized material fluid regions. Just as for material fluid parcels, the finite sized material fluid region retains the same matter content, and thus maintains a constant mass. We contrast the discussion here with that for the Eulerian regions (fixed in space) considered in Section 15.1. One key operational distinction between the Eulerian and material domains is that partial time derivative operators commute with integration over a fixed Eulerian domain, whereas material time derivative operators commute with integration over a material domain as per Reynolds transport theorem derived in Section 15.3.4.

15.3.1 Evolution of volume

Consider a finite material region, $\mathcal{R}(\mathbf{v})$, whose volume is given by the integral

$$V = \int_{\mathcal{R}(\mathbf{v})} dV, \quad (15.21)$$

with dV the volume element. The region changes its shape according to the motion of the fluid particles fixed to the boundary of the material region. We designate this volume as

$$\mathcal{R}(\mathbf{v}) = \text{region following flow}, \quad (15.22)$$

with the \mathbf{v} argument emphasizing that the volume moves with the flow. The material region expands when the flow moves outward and contracts when the flow moves inward. These statements take on the following mathematical expression

$$\frac{d}{dt} \int_{\mathcal{R}(\mathbf{v})} dV = \oint_{\partial\mathcal{R}(\mathbf{v})} \mathbf{v} \cdot \hat{\mathbf{n}} d\delta, \quad (15.23)$$

where $\hat{\mathbf{n}}$ is the outward normal on the region's closed boundary, $d\delta$ is the area element on the boundary, and

$$\mathbf{v} \cdot \hat{\mathbf{n}} d\delta = \text{volume transport (volume per time) at the boundary } \partial\mathcal{R}. \quad (15.24)$$

Use of the divergence theorem then leads to the equivalent expression

$$\frac{d}{dt} \int_{\mathcal{R}(\mathbf{v})} dV = \int_{\mathcal{R}(\mathbf{v})} \nabla \cdot \mathbf{v} dV. \quad (15.25)$$

We now take the limit as the material region becomes a material parcel, in which case we recover the differential expression

$$\frac{1}{\delta V} \frac{D(\delta V)}{Dt} = \nabla \cdot \mathbf{v}, \quad (15.26)$$

where we make use of D/Dt since the infinitesimal volume is moving with the fluid. This equation is also derived in Section 14.4.1 using different methods.

15.3.2 Mass conservation

The mass of fluid contained in a finite material region is given by

$$M = \int_{\mathcal{R}(\mathbf{v})} \rho dV. \quad (15.27)$$

As a material fluid region, it maintains a constant mass as it moves through the fluid

$$\frac{d}{dt} \int_{\mathcal{R}(\mathbf{v})} \rho dV = 0. \quad (15.28)$$

Just as for the volume in Section 15.3.1, taking the limit as the material region becomes infinitesimally small, the region mass conservation statement (15.28) becomes the parcel mass conservation statement (15.13)

$$\frac{D(\delta M)}{Dt} = \frac{D(\rho \delta V)}{Dt} = 0. \quad (15.29)$$

15.3.3 Mass conservation using Lagrangian methods

Rather than take the limit as the finite material region $\mathcal{R}(\mathbf{v})$ becomes infinitesimal, we develop some formalism that specifies how to move the time derivative across an integral of the mass within a finite sized material region as in equation (15.28). As part of this discussion we introduce two coordinate representations of the mass within a material region. The first is the Cartesian \mathbf{x} -space representation

$$\int_{\mathcal{R}(\mathbf{v})} \rho dV = \int_{x_1(t)}^{x_2(t)} \int_{y_1(t)}^{y_2(t)} \int_{z_1(t)}^{z_2(t)} \rho dx dy dz \equiv \int_{\mathcal{R}[\mathbf{v}(\mathbf{x}, t)]} \rho dV(\mathbf{x}). \quad (15.30)$$

The first equality expresses the integral with bounds that are functions of time, with the time dependence needed since the material region is moving with the fluid. The second equality uses a notation that emphasizes the region bounds and volume element are moving with the fluid velocity, \mathbf{v} , as specified using the \mathbf{x} -space representation, $\mathbf{v}(\mathbf{x}, t)$.

An alternative representation of mass in a material fluid region makes use of \mathbf{a} -space material coordinates. For this representation we perform a coordinate transformation from the \mathbf{x} -space representation to the \mathbf{a} -space representation, with this coordinate transformation requiring the Jacobian of transformation. To capture the gist of this transformation let us consider the one-dimensional case in which

$$\int_{x_1(t)}^{x_2(t)} \rho dx = \int_{a[x_1(t)]}^{a[x_2(t)]} \rho \frac{\partial X}{\partial a} da = \int_{a_1}^{a_2} \rho \frac{\partial X}{\partial a} da. \quad (15.31)$$

The first equality introduced the Jacobian, $\partial X / \partial a$, for the one-dimensional coordinate transformation from x -space to a -space, with corresponding changes to the limits of integration. We are allowed to make this coordinate transformation since there is a 1-to-1 relation between the a -space and x -space representation of a material fluid parcel. The second equality wrote the integral bounds in terms of the material coordinate. Since we are considering a material region that follows fluid particles, the integral bounds have fixed material coordinate values, $a[x_1(t)] = a_1$ and $a[x_2(t)] = a_2$. Generalizing to three dimensions then renders

$$\int_{\mathcal{R}[\mathbf{v}(\mathbf{x}, t)]} \rho dV(\mathbf{x}) = \int_{\mathcal{R}[\mathbf{v}(\mathbf{a})]} \rho \frac{\partial \mathbf{X}}{\partial \mathbf{a}} dV(\mathbf{a}). \quad (15.32)$$

Again, we distinguish the \mathbf{x} and \mathbf{a} arguments for the material region, with the \mathbf{x} representation also requiring a time argument whereas the material \mathbf{a} representation has no time dependence since it is attached to the moving fluid particles.

We now make use of the equality (15.32) to take the time derivative of the mass contained in the material region. Since the time derivative follows the material region, we can make use of the

material space coordinate representation as part of the manipulations

$$\frac{d}{dt} \left[\int_{\mathcal{R}[\mathbf{v}(\mathbf{x},t)]} \rho dV(\mathbf{x}) \right] = \frac{d}{dt} \left[\int_{\mathcal{R}[\mathbf{v}(\mathbf{a})]} \rho \frac{\partial \mathbf{X}}{\partial \mathbf{a}} dV(\mathbf{a}) \right] \quad (15.33a)$$

$$= \int_{\mathcal{R}[\mathbf{v}(\mathbf{a})]} \frac{D}{Dt} \left[\rho \frac{\partial \mathbf{X}}{\partial \mathbf{a}} \right] dV(\mathbf{a}) \quad (15.33b)$$

$$= \int_{\mathcal{R}[\mathbf{v}(\mathbf{x},t)]} \left[\frac{D\rho}{Dt} + \rho \nabla \cdot \mathbf{v} \right] dV(\mathbf{x}) \quad (15.33c)$$

$$= \int_{\mathcal{R}[\mathbf{v}(\mathbf{x},t)]} \left[\frac{1}{\rho} \frac{D\rho}{Dt} + \nabla \cdot \mathbf{v} \right] \rho dV(\mathbf{x}), \quad (15.33d)$$

where we used equation (14.76) for the equality (15.33c). When expressing the integral bounds using \mathbf{a} -space coordinates, the integral bounds have no time dependence since they are fixed on fluid particles. We can thus move the time derivative inside of the integral sign to reach the second equality. Upon entering the integral the time derivative is written as a material time derivative, D/Dt , since it is a time derivative computed by following the fluid particles that define the material region. The third equality made use of equation (15.17c) and converted back to \mathbf{x} -space. As the material region \mathcal{R} has a materially constant mass, we recover the mass continuity equation (15.16) by setting the integrand in equation (15.33d) to zero.

15.3.4 Reynolds transport theorem

On first encounter, the method from Section 15.3.3 that involves moving between Eulerian (\mathbf{x} -space) and Lagrangian (\mathbf{a} -space) representations is clumsy at best and a black box at worse. However, with some practice it becomes an elegant means to study the time evolution of fluid properties integrated over a material region. The method is formalized by the *Reynolds transport theorem* discussed in this section.

Manipulations leading to the mass conservation statement (15.33d) can be generalized by considering the material time derivative of a mass-weighted field ψ (e.g., a tracer concentration as in Section 16.1)

$$\frac{D(\psi \rho \delta V)}{Dt} = \frac{D\psi}{Dt} \rho \delta V + \psi \frac{D(\rho \delta V)}{Dt} \quad (15.34a)$$

$$= \rho \delta V \left[\frac{D\psi}{Dt} + \frac{\psi D\rho}{\rho} + \psi \nabla \cdot \mathbf{v} \right] \quad (15.34b)$$

$$= \delta V \left[\frac{\partial(\rho \psi)}{\partial t} + \nabla \cdot (\rho \psi \mathbf{v}) \right]. \quad (15.34c)$$

The first equality used the product rule, which holds for material time derivatives. Mass conservation means that the material derivative $D(\rho \delta V)/Dt = 0$. However, we choose to write mass conservation in the form of equation (15.17c), which allows us to introduce the flux-form Eulerian expression after replacing the material time derivative with its Eulerian form from equation (13.29).

Another, perhaps more direct means, to derive this result is to write

$$\rho \frac{D\psi}{Dt} = \rho \left[\frac{\partial \psi}{\partial t} + \mathbf{v} \cdot \nabla \psi \right] \quad (15.35a)$$

$$= \rho \left[\frac{\partial \psi}{\partial t} + \mathbf{v} \cdot \nabla \psi \right] + \psi \left[\frac{\partial \rho}{\partial t} + \nabla \cdot (\mathbf{v} \rho) \right] \quad (15.35b)$$

$$= \frac{\partial(\rho \psi)}{\partial t} + \nabla \cdot (\rho \mathbf{v} \psi). \quad (15.35c)$$

Following the discussion in Section 15.3.3, we can extend the material parcel result (15.34c) to a finite size material region. Again, each point in the material region is following a fluid particle. The result is known as the *Reynolds transport theorem*

$$\frac{d}{dt} \int_{\mathcal{R}(\mathbf{v})} \psi \rho dV = \int_{\mathcal{R}(\mathbf{v})} \frac{D\psi}{Dt} \rho dV \quad (15.36a)$$

$$= \int_{\mathcal{R}(\mathbf{v})} \left[\frac{\partial(\rho \psi)}{\partial t} + \nabla \cdot (\rho \psi \mathbf{v}) \right] dV \quad (15.36b)$$

$$= \int_{\mathcal{R}(\mathbf{v})} \frac{\partial(\rho \psi)}{\partial t} dV + \oint_{\partial \mathcal{R}(\mathbf{v})} \rho \psi \mathbf{v} \cdot \hat{\mathbf{n}} d\mathcal{S}, \quad (15.36c)$$

where the final equality follows from the divergence theorem. Additionally, we returned to the notation $\mathcal{R}(\mathbf{v})$ for material region as introduced in Section 15.3.1. This notation is sufficient to designate that the region is following fluid particles along with the fluid velocity, \mathbf{v} . Note that for the surface integral term, $\mathbf{v} \cdot \hat{\mathbf{n}}$ generally does not vanish. Rather, it is given by $\mathbf{v} \cdot \hat{\mathbf{n}} = \mathbf{v}^{(s)} \cdot \hat{\mathbf{n}}$, where $\mathbf{v}^{(s)}$ is the velocity of a point on the boundary of the material region. We detail this result in Section 15.4.2 when discussing the kinematics of a moving material surface.

15.3.5 Comments on notation for the time derivative

In this book we write d/dt for the time derivative operator acting on an integral. Furthermore, when the domain is specialized to follow fluid particles, we identify the special nature of such domains by introducing the fluid velocity argument to the domain name, $\mathcal{R}(\mathbf{v})$. This notation designates that all points in the domain, \mathcal{R} , move with the fluid velocity, \mathbf{v} , since all points have fluid particles attached. However, many authors choose an alternative notation by using the material time derivative, D/Dt , when acting on an integral over a material region. We thus have the following equality across the two notational conventions

$$\frac{D}{Dt} \int_{\mathcal{R}} \psi \rho dV = \frac{d}{dt} \int_{\mathcal{R}(\mathbf{v})} \psi \rho dV. \quad (15.37)$$

The use of one convention versus the other is a matter of taste. We follow Section 2.1 of [Batchelor \(1967\)](#) by restricting the D/Dt operator to act only on space-time fields, such as $\psi(\mathbf{x}, t)$. Hence, the D/Dt operator is not used when acting on integrals over spatial regions. Following this convention leads us to write $\mathcal{R}(\mathbf{v})$ for a region that moves with the fluid flow and to retain d/dt when acting on the integral over that region.

Admittedly, the $\mathcal{R}(\mathbf{v})$ notation is rarely used in the literature, with authors generally dropping the \mathbf{v} and thus letting words designate whether a region follows the flow or otherwise. As we have occasion in this book to consider a variety of fluid regions, we find it essential to introduce the somewhat more pedantic notation, $\mathcal{R}(\mathbf{v})$, to denote that a region moving with \mathbf{v} is being considered. This usage hopefully helps the reader freely swim along with the mathematical flow rather than struggling to swim in a turbulent sea driven by confused notation.

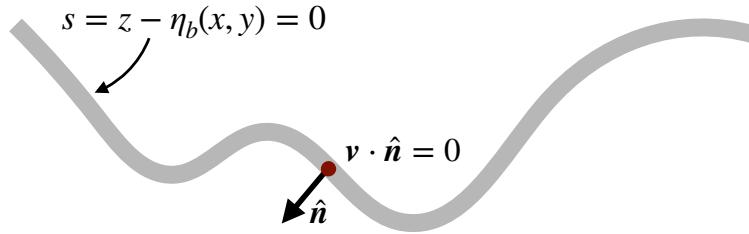


FIGURE 15.3: Illustrating the no-normal flow boundary condition maintained for a solid boundary, on which $\mathbf{v} \cdot \hat{\mathbf{n}} = 0$ (equation (15.38)). When the solid boundary denotes the ground or ocean bottom, then the position of the interface is written $s(x, y, z) = z - \eta_b(x, y) = 0$ (equation (15.39)). Correspondingly, the outward normal is given by $\hat{\mathbf{n}} = -\nabla s / |\nabla s| = -(\hat{\mathbf{z}} - \nabla \eta_b) / \sqrt{1 + |\nabla \eta_b|^2}$ (equation (15.40)).

15.4 Kinematic boundary conditions

When a fluid encounters a boundary, either at the edge of the fluid region or an imaginary boundary within the fluid itself, the fluid must accommodate the boundary. Conversely, the boundary must accommodate the fluid flow. Some boundaries are impermeable, so that they do not allow matter to cross. For material boundaries, any fluid originally in contact with the boundary stays in contact; at most this fluid can move tangential to the boundary without leaving it. Other boundaries are permeable, thus allowing matter to cross. Our goal in this section is to develop the various kinematic boundary conditions appropriate for the variety of cases encountered in fluid mechanics. These boundary conditions, and the considerations arising in their derivation, recur throughout fluid mechanics.

15.4.1 Static material surface

When a fluid encounters a static material surface, such as the solid-earth, the normal component of the fluid velocity must vanish since there is no fluid crossing the boundary (see Figure 15.3)

$$\mathbf{v} \cdot \hat{\mathbf{n}} = 0 \quad \text{no-normal flow condition on static material boundary.} \quad (15.38)$$

This relation constrains the flow so that any horizontal motion that is directed up or down a slope must have an associated vertical motion. Furthermore, recall our discussion of streamlines in Section 13.8.2, where $\mathbf{v} \cdot \hat{\mathbf{n}} = 0$ along a streamline. We thus see that the static material boundary is a flow streamline. That is, fluid that is in contact with the boundary will remain in contact. As the boundary is static, this result holds even in the case of a time dependent flow. Note that specification of the tangential velocity along a material boundary requires dynamical information unavailable from the purely kinematic analysis considered here.

For many cases in practice, the material surface is monotonic in the vertical, meaning there are no overturns. In this case, it is useful to introduce some differential geometry (at the level of introductory calculus) to unpack the boundary condition (15.38). Doing so helps to develop a geometric formalism especially useful for the more complicated moving boundary conditions in Sections 15.4.2 and 15.4.3. For this purpose, introduce a coordinate expression for the boundary according to

$$s(x, y, z) = z - \eta_b(x, y) = 0 \quad \text{static material boundary,} \quad (15.39)$$

with $z = \eta_b(x, y)$ the vertical position of the boundary. The outward normal vector at the boundary is thus given by

$$\hat{\mathbf{n}} = -\frac{\nabla s}{|\nabla s|} = -\frac{\nabla(z - \eta_b)}{|\nabla(z - \eta_b)|} = -\frac{\hat{\mathbf{z}} - \nabla \eta_b}{\sqrt{1 + |\nabla \eta_b|^2}}. \quad (15.40)$$

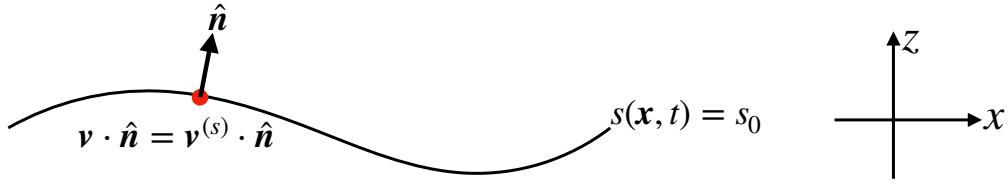


FIGURE 15.4: Illustrating the boundary condition for a moving material surface, on which $\hat{n} \cdot (\mathbf{v} - \mathbf{v}^{(s)}) = 0$ (equation (15.42)). For many cases, we can specify the surface by the value of a function that is a constant on the surface: $s(x, y, z, t) = s_0$ for some constant s_0 (equation (15.43)), in which case the normal direction is given by $\hat{n} = |\nabla s|^{-1} \nabla s$ (equation (15.44)). For example, s could represent a surface of constant buoyancy (Chapter 25) in a buoyancy stratified fluid.

Consequently, the no-flux boundary condition (15.38) takes the form

$$\nabla(z - \eta_b) \cdot \mathbf{v} = w - \mathbf{u} \cdot \nabla \eta_b = 0 \quad \text{at } z = \eta_b(x, y), \quad (15.41)$$

where the velocity is decomposed into its horizontal and vertical components, $\mathbf{v} = (\mathbf{u}, w)$. Hence, to maintain the no-flux boundary condition requires the vertical velocity component to precisely balance the projection of the horizontal velocity onto the slope of the material surface. If the material surface is flat, so that $\nabla \eta_b = 0$, then the kinematic boundary condition reduces to $w = 0$. Alternatively, if the flow is purely horizontal and thus moves along a constant η_b isoline, then $\mathbf{u} \cdot \eta_b = 0$ so that $w = 0$.

15.4.2 Moving material surface

Now consider the kinematic constraints imposed by a material surface moving with the fluid flow. Such material surfaces follow the flow as defined by the fluid particle velocity. Consequently, they do not allow net mass to cross the surface.²

General expression of the kinematic boundary condition

To ensure no matter crosses the material surface, the velocity of the surface must match that of fluid particles. More precisely, the normal component of the velocity for a point on the surface must match the normal component of the fluid at the surface. We are thus led to the kinematic boundary condition for a moving material surface

$$(\mathbf{v} - \mathbf{v}^{(s)}) \cdot \hat{n} = 0 \quad \text{moving material boundary condition.} \quad (15.42)$$

We illustrate this boundary condition in Figure 15.4, where $\mathbf{v}^{(s)}$ is the velocity of a point fixed on the moving material surface and \mathbf{v} is the velocity of the fluid particles.

The boundary condition (15.42) does not mean \mathbf{v} and $\mathbf{v}^{(s)}$ are identical. It only says that their normal components are the same when evaluated on the material surface. As a Corollary, we see that $\mathbf{v} \cdot \hat{n}$ is not generally zero so that a moving material boundary does *not* coincide with a flow streamline (see discussion in Sections 13.8.2 and 13.8.3).

²As we discuss in Chapter 16, with multiple components a surface that allows for zero net mass to cross it can still allow for the exchange of tracer mass in the presence of matter diffusion.

Specialized expression of the boundary condition

Now specialize the kinematic condition (15.42) to the case of a material surface specified by a function that takes on a constant value on the surface

$$s(\mathbf{x}, t) = s_0. \quad (15.43)$$

Correspondingly, the surface normal vector is given by

$$\hat{\mathbf{n}} = |\nabla s|^{-1} \nabla s. \quad (15.44)$$

From Section 13.5.6, we know that a point fixed on an arbitrary surface has a velocity that satisfies (see equation (13.42))

$$\frac{\partial s}{\partial t} + \mathbf{v}^{(s)} \cdot \nabla s = 0 \quad \text{on an iso-surface } s(\mathbf{x}, t) = s_0. \quad (15.45)$$

Use of the identity

$$\frac{\partial s}{\partial t} = \frac{Ds}{Dt} - \mathbf{v} \cdot \nabla s \quad (15.46)$$

renders

$$\frac{Ds}{Dt} - \mathbf{v} \cdot \nabla s + \mathbf{v}^{(s)} \cdot \nabla s = \frac{Ds}{Dt} + (\mathbf{v}^{(s)} - \mathbf{v}) \cdot \nabla s = 0. \quad (15.47a)$$

Since $(\mathbf{v}^{(s)} - \mathbf{v}) \cdot \nabla s = 0$ from the boundary condition (15.42), we are left with the material constancy condition

$$\frac{Ds}{Dt} = \frac{\partial s}{\partial t} + \mathbf{v} \cdot \nabla s = 0 \quad \text{on material surface } s(\mathbf{x}, t) = s_0. \quad (15.48)$$

Consequently, no matter crosses a surface of constant s as long as s is materially constant. This is an important kinematic property that reappears in many forms throughout fluid mechanics.

Boundary condition for a material interface

Consider the interface between two immiscible fluids. Assume this surface interface has an outward normal that has a nonzero vertical component, so that there are no breaking waves, for example. In this case we can express its vertical position of a point on the interface as

$$s(x, y, z, t) = z - \eta(x, y, t) = s_0. \quad (15.49)$$

The function $\eta(x, y, t)$ is the vertical deviation of the interface relative to the horizontal. The kinematic boundary condition (15.48) thus takes the form

$$\frac{Ds}{Dt} = \frac{D(z - \eta)}{Dt} = 0. \quad (15.50)$$

Hence, the vertical velocity component at the interface equals to the material time derivative of the interface displacement

$$\frac{Dz}{Dt} = \frac{D\eta}{Dt} \implies w = \frac{\partial \eta}{\partial t} + \mathbf{u} \cdot \nabla \eta \quad \text{material b.c. at interface } z = \eta(x, y, t). \quad (15.51)$$

This boundary condition can be equivalently written in the form

$$\mathbf{v} \cdot \hat{\mathbf{n}} = \frac{\partial \eta / \partial t}{\sqrt{1 + |\nabla \eta|^2}}, \quad (15.52)$$

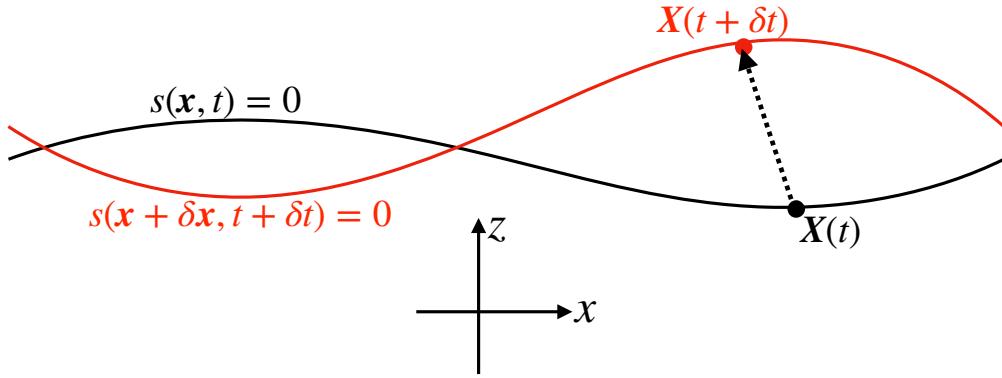


FIGURE 15.5: A surface that separates two fluid regions as realized at two time instants, along with the position of a sample point on the surface located at $\mathbf{X}(t)$ and $\mathbf{X}(t + \delta t)$. The equation $s(\mathbf{x}, t) = z - \eta(x, y, t) = 0$ specifies the vertical position for points on the surface as a function of horizontal position and time. At both time instances the vertical position is determined by $s(\mathbf{x}, t) = s(\mathbf{x} + \delta\mathbf{x}, t + \delta t) = 0$.

where

$$\hat{\mathbf{n}} = \frac{\nabla(z - \eta)}{|\nabla(z - \eta)|} = \frac{-\nabla\eta + \hat{\mathbf{z}}}{\sqrt{1 + |\nabla\eta|^2}} \quad (15.53)$$

is the outward normal at the material surface. These are expressions for the boundary condition placed on the ocean free surface when there is no water penetrating the surface (i.e., no rain or evaporation). In general, they provide kinematic expressions for the motion of an interface between two immiscible fluid layers.

A geometric derivation of the material boundary condition

The material invariance condition $Ds/Dt = 0$ is a key kinematic result. We thus offer yet another derivation to help solidify its meaning. As before, define the surface according to

$$s(\mathbf{x}, t) = z - \eta(x, y, t) = s_0, \quad (15.54)$$

which specifies the vertical position of a point on the surface at time t . Now consider the position of the surface after a small time interval, $t + \delta t$ (see Figure 15.5). The vertical position of the surface at the new time is determined by the same condition

$$s(\mathbf{x} + \delta\mathbf{x}, t + \delta t) = s_0, \quad (15.55)$$

where $\mathbf{X}(t + \delta t) = \mathbf{x} + \delta\mathbf{x}$ is the displaced position of a point on the surface that started at $\mathbf{X}(t) = \mathbf{x}$. Expanding equation (15.55) in a Taylor series to leading order yields

$$s(\mathbf{x}, t) + \delta\mathbf{x} \cdot \nabla s + \delta t \partial_t s = s_0. \quad (15.56)$$

Since $s(\mathbf{x}, t) = s_0$ from equation (15.54) we thus have

$$\frac{\partial s}{\partial t} + \frac{\delta\mathbf{x}}{\delta t} \cdot \nabla s = \frac{\partial s}{\partial t} + \frac{\delta\mathbf{x}}{\delta t} \cdot \hat{\mathbf{n}} |\nabla s| = 0, \quad (15.57)$$

where $\hat{\mathbf{n}} = |\nabla s|^{-1} \nabla s$ is the surface normal direction. This result means that when positioned at a fixed point in space, it is the normal component of the displacement that corresponds to a temporal modification of $s(\mathbf{x}, t)$

$$\frac{\partial s}{\partial t} = -\frac{\delta\mathbf{x}}{\delta t} \cdot \hat{\mathbf{n}} |\nabla s|. \quad (15.58)$$

In contrast, any tangential displacement along an s -isosurface leaves $s(\mathbf{x}, t)$ unchanged. Hence, when following motion of points on the surface, we are only concerned with motion along the direction set by the normal component of the velocity of that point, $\hat{\mathbf{n}}(\mathbf{v}^{(s)} \cdot \hat{\mathbf{n}})$. It is this velocity component that corresponds to movement of the surface normal to itself, thus leading to nonzero motion through space.

The above considerations mean that we are just concerned with the normal component of the velocity of a point on the surface, and so can write for the displacement

$$\delta\mathbf{x} = \delta t \hat{\mathbf{n}} (\mathbf{v}^{(s)} \cdot \hat{\mathbf{n}}). \quad (15.59)$$

Making use of this result in equation (15.57) leads to the differential equation satisfied by a point fixed on the moving surface

$$\frac{\partial s}{\partial t} + \mathbf{v}^{(s)} \cdot \nabla s = 0, \quad (15.60)$$

where we used the identity

$$\frac{\delta\mathbf{x}}{\delta t} \cdot \nabla s = \hat{\mathbf{n}} (\mathbf{v}^{(s)} \cdot \hat{\mathbf{n}}) \cdot \nabla s = \mathbf{v}^{(s)} \cdot \nabla s. \quad (15.61)$$

Assuming the surface is material means that

$$\mathbf{v}^{(s)} \cdot \hat{\mathbf{n}} = \mathbf{v} \cdot \hat{\mathbf{n}}, \quad (15.62)$$

so that motion of the surface normal to itself is identical to that of the fluid in the same direction. The identity in equation (15.60) renders the material invariance condition $Ds/Dt = 0$.

15.4.3 Dynamic and permeable surface

We now consider the kinematic boundary condition for a moving permeable surface that separates two fluid media (e.g., ocean and atmosphere) or two regions within a single media (e.g., surface of constant buoyancy within the ocean or within the atmosphere). As before, the kinematic boundary condition is a statement about the mass transport through the boundary. Whereas the previous conditions enforced a zero net mass transport through the boundary, here we allow for a generally non-zero net transport (mass per time). We write this transport condition as

$$\rho(\mathbf{v} - \mathbf{v}^{(s)}) \cdot \hat{\mathbf{n}} d\mathcal{S} = -\mathcal{Q}_m d\mathcal{S} \quad \text{moving non-material boundary condition.} \quad (15.63)$$

In this equation, $d\mathcal{S}$ is an infinitesimal area element on the surface, and \mathcal{Q}_m is the mass per time per surface area (mass flux) crossing the boundary. The minus sign is a convention that will be motivated in the following. We now massage this kinematic boundary condition into alternative forms of use for a variety of purposes.

Coordinate representation of the permeable surface

The expression (15.45) for $\mathbf{v}^{(s)} \cdot \hat{\mathbf{n}}$ holds for a point on an arbitrary surface so that

$$\mathbf{v}^{(s)} \cdot \hat{\mathbf{n}} = -\frac{\partial s / \partial t}{|\nabla s|}. \quad (15.64)$$

Furthermore, the projection of the fluid velocity onto the normal direction can be written

$$\frac{Ds}{Dt} = \frac{\partial s}{\partial t} + \mathbf{v} \cdot \nabla s \implies \mathbf{v} \cdot \hat{\mathbf{n}} = \frac{1}{|\nabla s|} \left[\frac{Ds}{Dt} - \frac{\partial s}{\partial t} \right]. \quad (15.65)$$

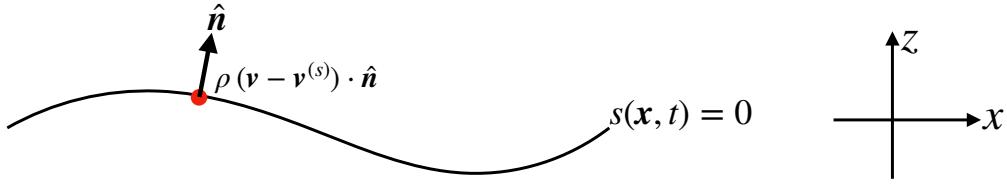


FIGURE 15.6: Illustrating the boundary condition for a moving permeable surface, such as the interface between two miscible fluid layers. On this surface, the boundary condition states that $\rho(\mathbf{v} - \mathbf{v}^{(s)}) \cdot \hat{\mathbf{n}} d\mathcal{S} = -\mathcal{Q}_m d\mathcal{S}$ (equation (15.63)). In the special case of an ocean free surface with no overturns, this boundary condition reduces to the surface kinematic boundary condition (15.79).

Bringing these results together leads to

$$\rho(\mathbf{v} - \mathbf{v}^{(s)}) \cdot \hat{\mathbf{n}} d\mathcal{S} = \frac{\rho d\mathcal{S}}{|\nabla s|} \frac{Ds}{Dt} \implies \mathcal{Q}_m = -\frac{\rho}{|\nabla s|} \frac{Ds}{Dt}. \quad (15.66)$$

This equation says that the mass transport crossing the surface is proportional to the material time derivative of the surface coordinate. The material time derivative vanishes when there is no mass transport across the surface (see discussion in Section 15.4.2).

In terms of the horizontal projection of the surface area

Assume that the surface is not vertical, so that its normal direction has a nonzero component in the vertical (e.g., waves that do not overturn). This assumption means that

$$\frac{\partial s}{\partial z} \neq 0, \quad (15.67)$$

so that we can further massage the boundary condition (15.66) by writing the area factor in the form

$$\frac{d\mathcal{S}}{|\nabla s|} = \frac{d\mathcal{S}}{\sqrt{(\partial s/\partial x)^2 + (\partial s/\partial y)^2 + (\partial s/\partial z)^2}} \quad (15.68a)$$

$$= \frac{d\mathcal{S}}{|\partial s/\partial z| \sqrt{[(\partial s/\partial x)/(\partial s/\partial z)]^2 + [(\partial s/\partial y)/(\partial s/\partial z)]^2 + 1}} \quad (15.68b)$$

$$= \frac{d\mathcal{S}}{|\partial s/\partial z| \sqrt{1 + \tan^2 \vartheta}} \quad (15.68c)$$

$$= \left| \frac{\partial z}{\partial s} \right| |\cos \vartheta| d\mathcal{S} \quad (15.68d)$$

$$= \left| \frac{\partial z}{\partial s} \right| dA. \quad (15.68e)$$

The equality (15.68c) introduced the angle, ϑ , between the boundary surface and the horizontal plane. The squared slope of this surface given by

$$\tan^2 \vartheta = \frac{\nabla_z s \cdot \nabla_z s}{(\partial s/\partial z)^2} = \nabla_s z \cdot \nabla_s z \quad (15.69)$$

with

$$\nabla_z = \hat{x} \left[\frac{\partial}{\partial x} \right]_{y,z} + \hat{y} \left[\frac{\partial}{\partial y} \right]_{x,z} \quad (15.70)$$

the horizontal gradient operator on constant z surfaces, and

$$\nabla_s = \hat{\mathbf{x}} \left[\frac{\partial}{\partial x} \right]_{y,s} + \hat{\mathbf{y}} \left[\frac{\partial}{\partial y} \right]_{x,s} \quad (15.71)$$

the horizontal gradient operator on constant s surfaces, and $z(x, y, s, t)$ the vertical position of the constant s surface.³ The equality (15.68d) made use of a trigonometric identity, and the equality (15.68e) introduced the horizontal projection of the area,

$$dA = |\cos \vartheta| d\mathcal{S}. \quad (15.72)$$

These results bring the kinematic boundary condition (15.66) into the form

$$\rho(\mathbf{v} - \mathbf{v}^{(s)}) \cdot \hat{\mathbf{n}} d\mathcal{S} = -Q_m d\mathcal{S} \quad (15.73a)$$

$$= \rho \frac{Ds}{Dt} \left| \frac{\partial z}{\partial s} \right| dA \quad (15.73b)$$

$$\equiv -Q_m dA. \quad (15.73c)$$

As defined, the flux Q_m is the net mass per time per horizontal area crossing the boundary surface

$$Q_m = -\rho(\mathbf{v} - \mathbf{v}^{(s)}) \cdot \hat{\mathbf{n}} \frac{d\mathcal{S}}{dA} = -\rho \frac{Ds}{Dt} \left| \frac{\partial z}{\partial s} \right|. \quad (15.74)$$

We motivate the minus sign through the ocean free surface case in the following.

Kinematic boundary condition at the ocean free surface

Consider the ocean free surface located at

$$s(x, y, z, t) = z - \eta(x, y, t) = 0 \quad \text{ocean free surface.} \quad (15.75)$$

For this boundary, $\partial s / \partial z = 1$ so that the area elements are related by

$$d\mathcal{S} = |\nabla(z - \eta)| dA = \sqrt{1 + |\nabla\eta|^2} dA. \quad (15.76)$$

The normal projection for the velocity of a point fixed on the free surface is given by

$$\mathbf{v}^{(\eta)} \cdot \hat{\mathbf{n}} = -\frac{\partial s / \partial t}{|\nabla s|} = \frac{\partial \eta / \partial t}{|\nabla(z - \eta)|} = \frac{\partial \eta / \partial t}{\sqrt{1 + |\nabla\eta|^2}} \implies \mathbf{v}^{(\eta)} \cdot \hat{\mathbf{n}} d\mathcal{S} = \partial_t \eta dA, \quad (15.77)$$

so that the mass flux crossing the free surface is

$$-\mathcal{Q}_m = \rho(\mathbf{v} - \mathbf{v}^{(\eta)}) \cdot \hat{\mathbf{n}}. \quad (15.78)$$

The boundary condition (15.74) thus takes the form

$$\rho(\mathbf{v} - \mathbf{v}^{(\eta)}) \cdot \hat{\mathbf{n}} \frac{d\mathcal{S}}{dA} = \rho \left[\frac{D(z - \eta)}{Dt} \right] = -Q_m \implies w + \rho^{-1} Q_m = \frac{\partial \eta}{\partial t} + \mathbf{u} \cdot \nabla \eta. \quad (15.79)$$

We now motivate the sign convention chosen for equation (15.73c) by considering the special case of a flat free surface and a resting fluid with $\mathbf{v} = 0$. Adding mass to the ocean raises the free surface, so that $\partial \eta / \partial t > 0$. Hence, the chosen sign convention means that $Q_m > 0$ corresponds to mass added to the ocean.

³We study in detail such operators in Chapter 43 when developing the mathematics of generalized vertical coordinates.

Kinematic boundary condition on a buoyancy surface

Now consider the boundary surface to be a surface of constant potential density in the ocean (or analogously a surface of constant specific entropy in the atmosphere). These buoyancy isosurfaces are also known as isopycnals, and we use the symbol⁴

$$s = \sigma(x, y, z, t) \quad (15.80)$$

for a particular isopycnal, σ . The mass transport crossing the isopycnal is written

$$Q_m = \rho \frac{D\sigma}{Dt} \left| \frac{\partial z}{\partial \sigma} \right| \equiv \rho w^{(\dot{\sigma})}, \quad (15.81)$$

where we introduced the *diapycnal transport velocity*

$$w^{(\dot{\sigma})} \equiv \frac{D\sigma}{Dt} \left| \frac{\partial z}{\partial \sigma} \right|. \quad (15.82)$$

A key aspect of physical oceanography concerns the development of theories for processes that cause a non-zero diapycnal transport. Examples include breaking waves, which act to mix matter across density surfaces; i.e., to *entrain* water from one density class to another.

15.5 Volume and mass budgets for an ocean column

We close this chapter by deriving the budget for the volume per horizontal area in a column of ocean fluid such as that shown in Figure 15.7, as well as the budget for the mass per horizontal area in this column. The derivations require much of the formalism discussed earlier, thus serving as a useful close to the chapter. Note that since the upper boundary of the domain is the free surface, and since the free surface is a function of time, the region is not strictly Eulerian even though the sides are fixed in space. Furthermore, the free surface is permeable, as are the sides, so that the region is not material.

15.5.1 Kinematic free surface equation

We here derive an equation for the free surface evolution, with this equation providing a budget for the volume per horizontal area in the column. In outline form, the derivation proceeds by vertically integrating the mass continuity equation (15.16) over the depth of an ocean column, from $z = \eta_b(x, y)$ at the bottom to $z = \eta(x, y, t)$ at the free surface. Use of the bottom and surface kinematic boundary conditions renders a kinematic expression for the free surface time tendency.

Vertically integrating the continuity equation (15.16) for a compressible fluid renders

$$-\int_{\eta_b}^{\eta} \frac{1}{\rho} \frac{D\rho}{Dt} dz = \int_{\eta_b}^{\eta} \nabla \cdot \mathbf{v} dz \quad (15.83a)$$

$$= w(\eta) - w(\eta_b) + \int_{\eta_b}^{\eta} \nabla_z \cdot \mathbf{u} dz \quad (15.83b)$$

$$= w(\eta) - w(\eta_b) + \nabla_z \cdot \left[\int_{\eta_b}^{\eta} \mathbf{u} dz \right] - \mathbf{u}(\eta) \cdot \nabla_z \eta + \mathbf{u}(\eta_b) \cdot \nabla_z \eta_b \quad (15.83c)$$

$$= [w(\eta) - \mathbf{u}(\eta) \cdot \nabla_z \eta] - [w(\eta_b) - \mathbf{u}(\eta_b) \cdot \nabla_z \eta_b] + \nabla_z \cdot \left[\int_{\eta_b}^{\eta} \mathbf{u} dz \right], \quad (15.83d)$$

⁴In this book, we use σ as an arbitrary generalized vertical coordinate, here chosen to be an isopycnal.

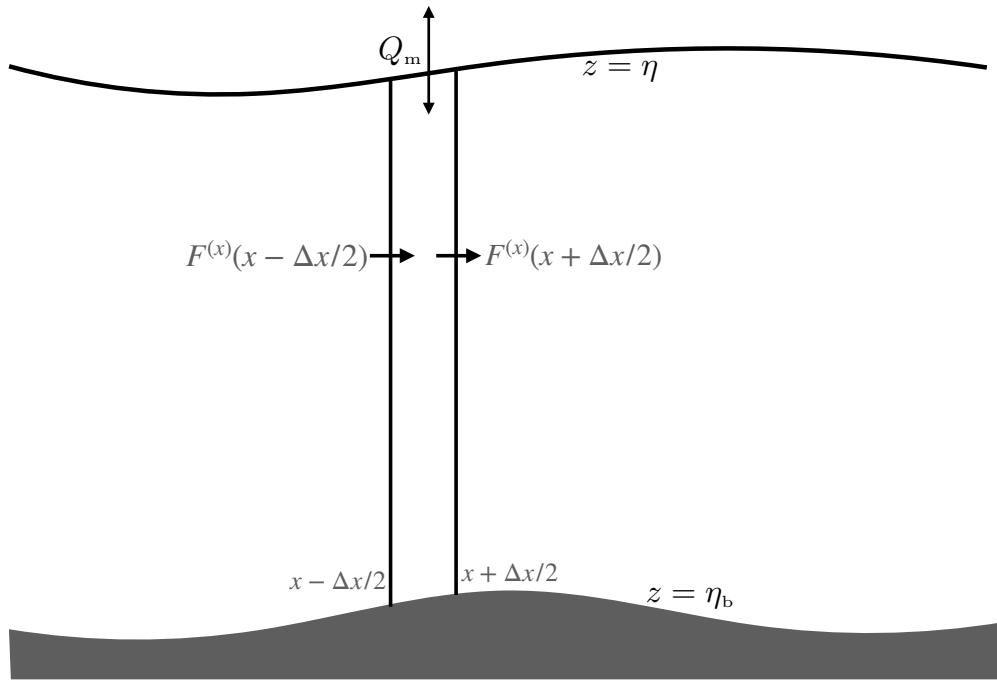


FIGURE 15.7: A longitudinal-vertical slice of ocean fluid from the surface at $z = \eta(x, y, t)$ to bottom at $z = \eta_b(x, y)$. The horizontal boundaries of the column at $x - \Delta x/2$ and $x + \Delta x/2$ are static and are penetrated by zonal mass transport, $F^{(x)}$. The zonal mass transport is computed by integrating the zonal mass flux, ρu over the area of the column sides. A similar transport acts in the meridional direction as well. The free surface is time dependent thus making the horizontal cross-sectional area for the fluid column time independent. The ocean bottom at the solid-earth boundary, $z = \eta_b(x, y)$, is static with no mass crossing this interface. The ocean surface at $z = \eta(x, y, t)$ is time dependent with mass flux Q_m crossing this interface.

where we made use of Leibniz's rule to move the horizontal divergence outside of the integral. Also note that $\nabla \cdot \mathbf{u} = \nabla_z \cdot \mathbf{u}$ since \mathbf{u} is the horizontal velocity, and likewise for $\nabla \eta_b$ and $\nabla \eta$ since η_b and η are both functions of horizontal space and time.

Use of the surface kinematic boundary condition (15.79) and no-normal flow bottom boundary condition yield

$$\frac{\partial \eta}{\partial t} = \frac{Q_m}{\rho(\eta)} - \nabla \cdot \mathbf{U} - \int_{\eta_b}^{\eta} \frac{1}{\rho} \frac{D\rho}{Dt} dz \quad (15.84)$$

where

$$\mathbf{U} = \int_{\eta_b}^{\eta} \mathbf{u} dz \quad (15.85)$$

is the depth integrated horizontal transport. Hence, as deduced from the mass continuity equation, the ocean free surface time tendency is affected by the passage of mass across the surface boundary (as normalized by the surface density), the convergence of depth integrated flow, and the depth integral of the material changes in density. [Griffies and Greatbatch \(2012\)](#) provide a more complete analysis of the sea surface height budget (15.84) by unpacking the physical processes leading to the material evolution of density.

15.5.2 Budget for mass per horizontal area

The mass per horizontal area in the fluid column is given by $\int_{\eta_b}^{\eta} \rho dz$. Use of Leibniz's rule, the bottom kinematic boundary condition, (15.41), surface kinematic boundary condition (15.79), and

the mass continuity equation (15.6), leads to

$$\frac{d}{dt} \left[\int_{\eta_b}^{\eta} \rho dz \right] = \rho(\eta) \frac{\partial \eta}{\partial t} + \int_{\eta_b}^{\eta} \frac{\partial \rho}{\partial t} dz \quad (15.86a)$$

$$= \rho(\eta) \frac{\partial \eta}{\partial t} - \int_{\eta_b}^{\eta} \nabla \cdot (\rho \mathbf{v}) dz \quad (15.86b)$$

$$= \rho(\eta) \left[\frac{\partial \eta}{\partial t} - w(\eta) \right] + \rho(\eta_b) w(\eta_b) - \int_{\eta_b}^{\eta} \nabla_z \cdot (\rho \mathbf{u}) dz \quad (15.86c)$$

$$= \rho(\eta) \left[\frac{\partial \eta}{\partial t} + \mathbf{u} \cdot \nabla \eta - w(\eta) \right] + \rho(\eta_b) [w(\eta_b) - \mathbf{u}(\eta_b) \cdot \nabla \eta_b] - \nabla_z \cdot \mathbf{U}^\rho \quad (15.86d)$$

$$= Q_m - \nabla_z \cdot \mathbf{U}^\rho, \quad (15.86e)$$

where

$$\mathbf{U}^\rho = \int_{\eta_b}^{\eta} \rho \mathbf{u} dz. \quad (15.87)$$

Hence, the mass per horizontal area within a column evolves according to

$$\frac{d}{dt} \left[\int_{\eta_b}^{\eta} \rho dz \right] = Q_m - \nabla \cdot \mathbf{U}^\rho, \quad (15.88)$$

with terms on the right hand side representing the convergence of mass onto the column either through the sides or surface. We consider an alternative derivation of this budget in Exercise 15.2.

15.6 Exercises

EXERCISE 15.1: CENTER OF MASS MOTION

Consider a material fluid region, $\mathcal{R}(\mathbf{v})$, with constant mass written as

$$M = \int_{\mathcal{R}(\mathbf{v})} \rho dV. \quad (15.89)$$

- (a) Show mathematically that the centre of mass for the region moves with the region's total linear momentum

$$\frac{d}{dt} \left[\frac{1}{M} \int_{\mathcal{R}(\mathbf{v})} \mathbf{x} \rho dV \right] = \frac{1}{M} \int_{\mathcal{R}(\mathbf{v})} \frac{D\mathbf{x}}{Dt} \rho dV = \frac{1}{M} \int_{\mathcal{R}(\mathbf{v})} \mathbf{v} \rho dV. \quad (15.90)$$

Precisely describe the reasoning behind each step. Note: a brief solution is sufficient, so long as the reasoning is sound.

- (b) Show mathematically (or precisely describe why) that the time change in the linear momentum for the region is given by

$$\frac{d}{dt} \left[\int_{\mathcal{R}(\mathbf{v})} \rho \mathbf{v} dV \right] = \int_{\mathcal{R}(\mathbf{v})} \frac{D\mathbf{v}}{Dt} \rho dV. \quad (15.91)$$

Precisely describe the reasoning behind each step. Note: a brief solution is sufficient, so long as the reasoning is sound.

EXERCISE 15.2: MASS BUDGET FOR A FLUID COLUMN

We here provide an alternative derivation of equation (15.88), the budget for the mass per horizontal area over a column of fluid.

The mass within an arbitrary fluid region is given by

$$M = \int \rho dV. \quad (15.92)$$

Consider the fluid mass within the column shown in Figure 15.7. In this column, the vertical side-walls are fixed in time, the bottom surface, $z = \eta_b(x, y)$, is at the solid-earth boundary, and the top, $z = \eta(x, y, t)$, is the fluctuating ocean free surface. Convince yourself that the mass for this column can be written

$$M = \iint \left[\int_{\eta_b(x, y)}^{\eta(x, y, t)} \rho dz \right] dx dy, \quad (15.93)$$

where the horizontal (x, y) integrals extend over the horizontal extent of the column. Mass conservation for this column means that the change in mass arises just through boundary fluxes, so that

$$\frac{dM}{dt} = - \int \rho \Delta \mathbf{v} \cdot \hat{\mathbf{n}} d\mathcal{S}, \quad (15.94)$$

where $\hat{\mathbf{n}}$ is the outward normal to the surface of the fluid region, $d\mathcal{S}$ is the area of an infinitesimal element on the surface, and the minus sign means that fluid leaving the region contributes to a reduction in mass within the region. The term

$$\Delta \mathbf{v} = \mathbf{v} - \mathbf{v}^{(s)} \quad (15.95)$$

is the velocity of the fluid relative to the velocity of the boundary; e.g., see the kinematic boundary condition discussion in Section 15.4.3. We also derive a general form of this relation in equation (16.38), though this exercise can be solved without knowing the details of that derivation.

- (a) Mass transported in the zonal direction ($\hat{\mathbf{x}}$) that crosses the column's vertical boundary at x is given by

$$F^{(x)}(x, y, t) = \int_{y-\Delta y/2}^{y+\Delta y/2} \left[\int_{\eta_b(x, y')}^{\eta(x, y', t)} u(x, y', z', t) \rho(x, y', z', t) dz' \right] dy' \quad (15.96a)$$

$$\equiv \int_{y-\Delta y/2}^{y+\Delta y/2} U^\rho(x, y', t) dy', \quad (15.96b)$$

and similarly for mass transport in the meridional direction

$$F^{(y)}(x, y, t) = \int_{x-\Delta x/2}^{x+\Delta x/2} \left[\int_{\eta_b(x', y)}^{\eta(x', y, t)} v(x', y, z', t) \rho(x', y, z', t) dz' \right] dx' \quad (15.97a)$$

$$\equiv \int_{x-\Delta x/2}^{x+\Delta x/2} V^\rho(x', y, t) dx', \quad (15.97b)$$

where

$$\mathbf{U}^\rho(x, y, t) = \int_{\eta_b(x, y)}^{\eta(x, y, t)} \mathbf{u}(x, y, z', t) \rho(x, y, z', t) dz' = \hat{\mathbf{x}} U^\rho + \hat{\mathbf{y}} V^\rho. \quad (15.98)$$

What are the physical dimensions [in terms of length (L), mass (M), and time (T)] for the mass transports, $F^{(x)}$ and $F^{(y)}$?

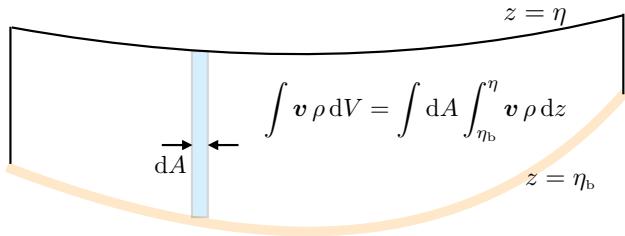


FIGURE 15.8: Cross-section of the integration region for Exercise 15.3, with the region extending from the ocean bottom at $z = \eta_b(x, y)$ and the free surface at $z = \eta(x, y, t)$. The sides are assumed to be vertical and rigid. An infinitesimal column is shown with cross-sectional area dA , extending from the bottom to the surface. The cross-sectional area for the column is time independent, so that a time derivative passes across the area integral to act only on the upper limit $z = \eta$ and the integrand in equation (15.103).

- (b) Using these expressions for the mass crossing the vertical side boundaries of a fluid column, take the limit as the horizontal cross-sectional area of the column becomes infinitesimally small to show that the evolution equation for the mass per unit area of the column is given by

$$\frac{d}{dt} \left[\int_{\eta_b}^{\eta} \rho dz \right] = -\nabla \cdot \mathbf{U}^\rho + Q_m, \quad (15.99)$$

where Q_m is the mass transport entering the ocean through the surface, per horizontal area, as defined by equation (15.73c), so that

$$\int Q_m dA = - \int \rho \Delta \mathbf{v} \cdot \hat{\mathbf{n}} d\delta \quad \text{at } z = \eta. \quad (15.100)$$

The derivation of equation (15.99) is part of this exercise, using methods distinct from those used in Section 15.5.2.

- (c) In words, the mass budget in equation (15.99) says that mass changes in a column of fluid if there is a convergence of mass into the column across its vertical boundaries (first term on right hand side), and a mass flux entering the column across the ocean surface (second term on right hand side). What are the physical dimensions of all terms in equation (15.99)?

EXERCISE 15.3: CHANGE IN LINEAR MOMENTUM OF A FLUID REGION

Consider a closed ocean basin with zero boundary fluxes of matter; i.e., zero precipitation/evaporation and zero mass fluxes through the solid-earth bottom. Consequently, this region is bounded by material surfaces and so it maintains constant matter content with fixed mass

$$M = \int_{\mathcal{R}} \rho dV. \quad (15.101)$$

Show that the time change in the linear momentum for this ocean basin is given by

$$\frac{d}{dt} \left[\int_{\mathcal{R}} \rho \mathbf{v} dV \right] = \int_{\mathcal{R}} \frac{D\mathbf{v}}{Dt} \rho dV. \quad (15.102)$$

This result is identical to that derived in Exercise 15.1. Rather than just repeating the solution method used there, make use of Leibniz's rule, the kinematic boundary condition detailed in Section 15.4.2, and mass conservation.

Hint: Refer to Figure 15.8 for a schematic of the integration where we have expanded the volume integral into the form

$$\frac{d}{dt} \left[\int_{\mathcal{R}} \rho \mathbf{v} dV \right] = \frac{d}{dt} \left[\int \left(\int_{\eta_b}^{\eta} \rho \mathbf{v} dz \right) dA \right], \quad (15.103)$$

where the horizontal integral extends over the rigid and fixed horizontal area of the basin, $dA = dx dy$ is the time independent horizontal area element, $z = \eta_b(x, y)$ is the solid-earth bottom and $z = \eta(x, y, t)$ is the ocean free surface. Time dependence appears in the upper boundary at $z = \eta$ and within the integrand. Perform the time derivative operation and make use of mass continuity and the kinematic boundary condition. Also make use of the trigonometry presented in Section 15.4.3 (in particular equation (15.72)). Unlike the formulation in Exercise 15.1, there is no use of a material time derivative in this approach. Rather, it is a straightforward use of integration over a domain with fixed horizontal/bottom boundaries and a time dependent free surface boundary.



Tracer conservation

Throughout this book, we assume that fluid mass is neither created nor destroyed anywhere within the fluid domain. We qualify this statement by noting that constituent matter may be created or destroyed via chemical reactions that convert matter from one form to the other. Such processes are typically represented as sources or sinks for matter components. Yet even in the presence of chemical reactions, we assume that mass is conserved. Hence, the net mass within a given fluid domain changes only via transport crossing the domain boundary. We here extend the discussion from Chapter 15, which focused on homogeneous fluids, to the case of a non-homogeneous fluid comprised of multiple matter constituents. In so doing we develop budgets for material tracers.

READER'S GUIDE TO THIS CHAPTER

The formulation pursued in this chapter is inspired by similar treatments in the chemical physics literature (e.g., Chapter 11 of *Aris* (1962) and Chapter II of *DeGroot and Mazur* (1984)), where they develop the theory for transport processes in multi-component fluids (along with chemical reactions, which we do not cover in this book). For this chapter, we assume an understanding of the Eulerian and Lagrangian kinematic descriptions detailed in Chapter 13 and the mass conservation analysis in Chapter 15. Much of the material from this chapter is used for the study of scalar fields such as potential enthalpy and material tracers. Furthermore, the Leibniz-Reynolds transport theorem of Section 16.3.4 is a kinematic expression that is central to all finite volume budgets in fluid mechanics.

16.1	The tracer equation	312
16.1.1	Mass conservation for each constituent	312
16.1.2	Total mass conservation	313
16.1.3	The tracer equation	314
16.1.4	Compatibility between total mass and tracer mass	315
16.1.5	Summary of some conceptual points	316
16.1.6	Further study	317
16.2	Passive tracers	317
16.3	Budgets for arbitrary fluid regions	318
16.3.1	Extensive and intensive fluid properties	318
16.3.2	General form of the finite domain integral	319
16.3.3	Eulerian (static) domain	320
16.3.4	Leibniz-Reynolds transport theorem	320
16.3.5	Revisiting Reynolds transport theorem	322
16.3.6	Summary of the time derivatives acting on integrals	324
16.4	Brute force illustration of Leibniz-Reynolds	324

16.4.1	Leibniz's rule plus kinematic boundary conditions	325
16.4.2	Summarizing the result	326
16.5	Boundary conditions	326
16.5.1	Interior layer boundary conditions	327
16.5.2	Solid-earth boundary conditions	327
16.5.3	Upper ocean surface boundary conditions	328
16.6	Evolution of region mean tracer concentration	329
16.6.1	Formulation	329
16.6.2	Application to a numerical ocean model grid cell	329
16.7	Exercises	330

16.1 The tracer equation

As defined in Section 13.2, a fluid element is an infinitesimal region of constant mass within the fluid continuum. Although possessing a constant mass, it generally has a non-constant material composition. That is, a fluid element is a non-material fluid parcel. Fluid element boundaries are open to the exchange of matter (i.e., tracers) with adjacent fluid elements. They are also open to the exchange of thermodynamic properties such as temperature and specific entropy. The kinematics of fluid elements share certain features with material fluid parcels. For example, we can uniquely specify the position of a fluid element by providing a material coordinate and time. Correspondingly, we can generalize Reynolds transport theorem for integration over a constant mass fluid region (Section 16.3.4). We make use of fluid elements to develop the mass budgets for non-homogeneous fluids such as the ocean and atmosphere. The constituent mass budgets are commonly referred to as *tracer equations*.

16.1.1 Mass conservation for each constituent

In this subsection we formulate the mass conservation equation for each constituent within the fluid. The mass equation is formulated by taking an integral or weak formulation approach with a fixed (Eulerian) region.

Density and velocity for each matter constituent

Consider a fluid with $n = 1, N$ matter constituents. For example, seawater has $N = 2$ when concerned just with its freshwater and salt content, whereas $N > 2$ when also concerned with other material constituents such as CO_2 and other biogeochemical species. Now focus on a fixed (Eulerian) region of the fluid with volume V and total mass M . Inside of the region, count the number of molecules of each constituent and determine their corresponding velocities. This information can be used to construct the molecular center of mass velocity for each constituent, $\mathbf{v}^{(n)}$, as well as the mass density,

$$\rho^{(n)} = V^{-1} M^{(n)}. \quad (16.1)$$

In the continuum limit, the volume and mass in the region get tiny ($V \rightarrow dV$ and $M \rightarrow dM$), yet the mass density remains smooth and finite (see Figure 12.2). Correspondingly, the constituent velocity and mass density are continuous fields whose values are available at each point within the continuum fluid.

Integral formulation of the constituent mass budget

Consider an arbitrary region, \mathcal{R} , assumed to be fixed in space (an Eulerian region). The mass of component n within \mathcal{R} is given by the integral

$$M^{(n)} = \int_{\mathcal{R}} \rho^{(n)} dV, \quad (16.2)$$

and it changes in time according to the finite volume budget equation (there is no implied summation on the right hand side)

$$\frac{d}{dt} \int_{\mathcal{R}} \rho^{(n)} dV = - \int_{\partial\mathcal{R}} \rho^{(n)} \mathbf{v}^{(n)} \cdot \hat{\mathbf{n}} d\mathcal{S}. \quad (16.3)$$

This equation is a constituent form of the finite volume mass budget given for a homogeneous fluid by equation (15.9). Since the region, \mathcal{R} , is assumed to be fixed in space, we can move the time derivative across the integral to reveal

$$\int_{\mathcal{R}} \left[\frac{\partial \rho^{(n)}}{\partial t} + \nabla \cdot (\rho^{(n)} \mathbf{v}^{(n)}) \right] dV = 0, \quad (16.4)$$

where we also used the divergence theorem to convert the surface integral to a volume integral. Arbitrariness of the region means that this integral expression must be satisfied at each point of the continuum, thus leading to the Eulerian form of the constituent mass continuity equation

$$\frac{\partial \rho^{(n)}}{\partial t} + \nabla \cdot (\rho^{(n)} \mathbf{v}^{(n)}) = 0. \quad (16.5)$$

This equation can also be written using a material time derivative

$$\frac{D^{(n)} \rho^{(n)}}{Dt} = -\rho^{(n)} \nabla \cdot \mathbf{v}^{(n)} \quad \text{for each of the } n = 1, N \text{ constituents}, \quad (16.6)$$

where the constituent material time derivative is given by

$$\frac{D^{(n)}}{Dt} = \frac{\partial}{\partial t} + \mathbf{v}^{(n)} \cdot \nabla. \quad (16.7)$$

We thus have N statements of mass conservation corresponding to each constituent material fluid parcel moving according to the velocity $\mathbf{v}^{(n)}$.

16.1.2 Total mass conservation

Summing the Eulerian mass continuity equation (16.5) over all constituents leads to the continuity equation for the total mass

$$\frac{\partial \rho}{\partial t} + \nabla \cdot (\rho \mathbf{v}) = 0, \quad (16.8)$$

where the total mass density and *barycentric velocity* are given by

$$\rho = \sum_{n=1}^N \rho^{(n)} \quad \text{and} \quad \mathbf{v} = \rho^{-1} \sum_{n=1}^N \rho^{(n)} \mathbf{v}^{(n)}. \quad (16.9)$$

Introducing the material time derivative following the barycentric velocity, $D/Dt = \partial/\partial t + \mathbf{v} \cdot \nabla$, leads to the equivalent material form for the mass conservation equation

$$\frac{D\rho}{Dt} = -\rho \nabla \cdot \mathbf{v}. \quad (16.10)$$

The *barycenter* of a distribution of matter is the center of inertia for that matter. We choose the term *barycentric velocity* to distinguish \mathbf{v} from the molecular *center of mass velocity*, $\mathbf{v}^{(n)}$, of each constituent. The barycentric velocity plays a key role in the conservation laws for multi-component fluids.

16.1.3 The tracer equation

Rather than keep track of each constituent velocity, $\mathbf{v}^{(n)}$, and the corresponding material parcels, it is generally more convenient to focus on the fluid element that moves with the barycentric velocity. For this purpose, we consider again the constituent mass continuity equation (16.5)

$$\left[\frac{\partial}{\partial t} + \mathbf{v}^{(n)} \cdot \nabla \right] \rho^{(n)} = -\rho^{(n)} \nabla \cdot \mathbf{v}^{(n)} \quad (16.11)$$

and insert the barycentric velocity to both sides by adding $0 = \mathbf{v} - \mathbf{v}$

$$\left[\frac{\partial}{\partial t} + (\mathbf{v} - \mathbf{v} + \mathbf{v}^{(n)}) \cdot \nabla \right] \rho^{(n)} = -\rho^{(n)} \nabla \cdot [\mathbf{v} - \mathbf{v} + \mathbf{v}^{(n)}]. \quad (16.12)$$

Rearrangement leads to

$$\left[\frac{\partial}{\partial t} + \mathbf{v} \cdot \nabla \right] \rho^{(n)} = -\rho^{(n)} \nabla \cdot \mathbf{v} - \nabla \cdot [\rho^{(n)} (\mathbf{v}^{(n)} - \mathbf{v})], \quad (16.13)$$

which can be written

$$\frac{D\rho^{(n)}}{Dt} = -\rho^{(n)} \nabla \cdot \mathbf{v} - \nabla \cdot \mathbf{J}^{(n)}, \quad (16.14)$$

where we defined the constituent mass flux

$$\mathbf{J}^{(n)} = \rho^{(n)} (\mathbf{v}^{(n)} - \mathbf{v}), \quad (16.15)$$

which arises from the difference between the constituent velocity and the barycentric velocity. The dimensions of $\mathbf{J}^{(n)}$ are mass of constituent n per time per area.

The material mass conservation equation (16.14) takes on the Eulerian form

$$\frac{\partial \rho^{(n)}}{\partial t} + \nabla \cdot (\mathbf{v} \rho^{(n)}) = -\nabla \cdot \mathbf{J}^{(n)}. \quad (16.16)$$

Introducing the tracer concentration $C^{(n)}$ according to

$$C^{(n)} = \frac{\rho^{(n)}}{\rho} = \frac{\delta M^{(n)}}{\delta M} = \frac{\text{mass of constituent } n \text{ in fluid element}}{\text{mass of fluid element}}, \quad (16.17)$$

leads to the tracer flux

$$\mathbf{J}^{(n)} = \rho C^{(n)} (\mathbf{v}^{(n)} - \mathbf{v}), \quad (16.18)$$

the Eulerian flux-form tracer budget

$$\frac{\partial(\rho C^{(n)})}{\partial t} + \nabla \cdot [\mathbf{v} \rho C^{(n)} + \mathbf{J}^{(n)}] = 0. \quad (16.19)$$

In Figure 16.1 we illustrate the contributions to the tracer evolution according to this Eulerian flux-form equation (16.19).

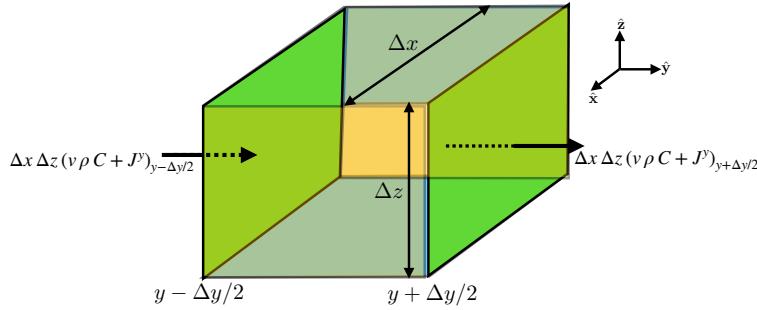


FIGURE 16.1: A finite sized cube as in Figure 15.1, here used to illustrate the budget of tracer mass over an Eulerian region. In addition to the advective flux of tracer moving with the barycentric velocity, \mathbf{v} , there is a diffusive flux, \mathbf{J} , that contributes to the transfer of tracer across the cell face.

Eulerian and Lagrangian forms of the tracer equation

The Eulerian flux-form equation (16.19) has a corresponding material time derivative form derived by expanding the derivatives to find

$$\begin{aligned} C^{(n)} \frac{\partial \rho}{\partial t} + \rho \frac{\partial C^{(n)}}{\partial t} + \rho \mathbf{v} \cdot \nabla C^{(n)} + C^{(n)} \nabla \cdot (\rho \mathbf{v}) \\ = C^{(n)} \left[\frac{\partial \rho}{\partial t} + \nabla \cdot (\rho \mathbf{v}) \right] + \rho \left[\frac{\partial C^{(n)}}{\partial t} + \mathbf{v} \cdot \nabla \right] C^{(n)}. \end{aligned} \quad (16.20)$$

The first term on the right hand side vanishes through mass continuity in the form of equation (16.8). The second term on the right hand side is the material time derivative of the tracer concentration. We are thus led to the equivalent forms for the tracer equation¹

$$\frac{\partial(\rho C^{(n)})}{\partial t} + \nabla \cdot [\mathbf{v} \rho C^{(n)}] = \rho \frac{DC^{(n)}}{Dt} = -\nabla \cdot \mathbf{J}^{(n)}. \quad (16.21)$$

Advective plus diffusive tracer fluxes

The above definitions allow us to decompose an advective tracer flux defined according to the tracer velocity into an advective flux based on the barycentric velocity plus a *diffusive flux*

$$\rho C^{(n)} \mathbf{v}^{(n)} = \rho C^{(n)} (\mathbf{v}^{(n)} - \mathbf{v} + \mathbf{v}) = \mathbf{J}^{(n)} + \rho C^{(n)} \mathbf{v}. \quad (16.22)$$

The diffusive flux vanishes when the tracer velocity equals to the barycentric velocity. Correspondingly, the diffusive flux also vanishes for a single-component fluid, since in that case there is only one matter component and so the constituent velocity equals to the barycentric velocity. We refer to this flux as “diffusive” since it is common in practice to parameterize this term as a downgradient diffusive flux. We study such parameterizations in Chapter 50.

16.1.4 Compatibility between total mass and tracer mass

By construction, the Eulerian flux-form of the tracer equation (16.19) is compatible with the flux-form continuity equation

$$\frac{\partial(\rho C^{(n)})}{\partial t} + \nabla \cdot [\rho \mathbf{v} C^{(n)} + \mathbf{J}^{(n)}] = 0 \iff \frac{\partial \rho}{\partial t} + \nabla \cdot (\rho \mathbf{v}) = 0. \quad (16.23)$$

¹The same result was also derived in equation (15.35c) when discussing Reynolds transport theorem for a homogeneous fluid.

Compatibility is manifest by summing the tracer equation over all constituents and using the identities

$$\sum_{n=1}^N C^{(n)} = 1 \quad \text{and} \quad \sum_{n=1}^N \mathbf{J}^{(n)} = 0. \quad (16.24)$$

Furthermore, through use of the barycentric velocity (16.9), we are ensured that the continuity equation for the total density of a fluid element is only transported by the barycentric velocity. There is no contribution from $\mathbf{J}^{(n)}$ since $\sum_{n=1}^N \mathbf{J}^{(n)} = 0$.

16.1.5 Summary of some conceptual points

What is a fluid element? How does it maintain constant mass but not constant matter? These questions can lead to some puzzles when encountering the notion of a fluid element. Here we aim to explain some of the conceptual points to resolve the puzzles, building from our initial specification of fluid elements and parcels in Section 13.2.

Revisiting a fluid element

The mass continuity equation (16.10) motivates us to define a fluid element as an infinitesimal fluid parcel that moves with barycentric velocity, \mathbf{v} , and maintains a constant total mass

$$\delta M = \sum_{n=1}^N \delta M^{(n)}. \quad (16.25)$$

The fluid element does not maintain a constant mass for each constituent, since the fluid element moves at the barycentric velocity, \mathbf{v} , which generally differs from the constituent velocities $\mathbf{v}^{(n)}$. Consequently, a fluid element boundary is permeable to matter transport that leaves its mass constant but allows for exchanges of matter constituents with adjacent fluid elements. Hence, if some matter leaves the fluid element, then an equal amount must enter the element in order to maintain a constant mass.

The exchange of matter across a fluid element's boundary can arise from the direct motion of matter crossing the boundary, or from the motion of the fluid element boundary relative to the matter. This point is central to resolving some conundrums associated with the notion of matter exchange constrained to retain constant mass. We have more to say on this point in the following.

Conceptual summary of the formulation

The formulation pursued in this section is based on considering the non-homogeneous fluid to be a continuum with distinct matter constituents (e.g., salt and freshwater for the ocean or water vapor and dry air for the atmosphere). Furthermore, the mass concentration for each constituent is represented by a scalar field whose value at any point in space-time gives the mass of tracer per mass of fluid. We then formulate mass conservation equations (i.e., continuity equations) for each matter constituent following methods used for the homogeneous fluid in Chapter 15. By choosing to use the *barycentric* (center of mass) velocity for describing fluid flow, the mass continuity equation for the total mass in a fluid element takes on the same form as for a homogeneous fluid. The resulting constituent mass budgets (i.e., tracer equations) have “diffusive” fluxes since the velocity of each matter constituent is generally distinct from the barycentric velocity.

To expose a bit of the details, we saw in this section that the tracer equation expresses the balance of mass for each trace constituent in the fluid. Furthermore, a nonzero tracer flux, $\mathbf{J}^{(n)} =$

$\rho C^{(n)} (\mathbf{v}^{(n)} - \mathbf{v})$, arises when the barycentric velocity, \mathbf{v} , differs from the constituent velocity, $\mathbf{v}^{(n)}$. In that case, matter and thermodynamic properties are exchanged between fluid elements, with the exchange made without altering the mass of a fluid element. In the presence of random motion within a turbulent fluid, or in the presence of random interactions with molecular degrees of freedom, tracer exchange is akin to a random walk. Such exchange is commonly parameterized by a diffusion process (see Section 48.3). Correspondingly, the mass of tracers in a fluid element is altered in the presence of differences in tracer concentration between fluid elements (i.e., tracer concentration gradients). However, the total mass of the element remains fixed.

How to maintain constant mass

As defined, a fluid element provides a generalization to multi-component fluids of the notion of a constant mass material fluid parcel that we used in describing a homogeneous fluid (see Section 13.2). Later in this chapter we will encounter a finite volume extension of the fluid element, which we refer to as a *Lagrangian region*. A Lagrangian region has boundaries that follow the barycentric velocity, \mathbf{v} , so that the region maintains constant mass as per our discussion of Reynolds transport theorem in Section 16.3.5.

To maintain constant mass, any matter that leaves the fluid element by crossing its boundary is compensated by an equal mass that enters the boundary. Kinematically, there are two means for matter to cross a boundary. First, the matter itself can move across the boundary, with the limiting case being a stationary boundary with matter moving across. Second, the boundary can move relative to the matter, with the limiting case being stationary matter with the boundary moving. In either case, by choosing to follow the barycentric velocity, a fluid element's boundary (or a corresponding Lagrangian region's boundary) adjusts so that mass remains constant.

The strategic choice to formulate the kinematics of multi-component fluids using the barycentric velocity is directly analogous to the choice in classical mechanics to describe motion relative to the center of mass for a system of many moving objects such as planets or point particles. In particular, by describing the motion of a multi-component fluid using the barycentric velocity, we simplify the kinematics by linking to the kinematics of homogeneous fluids while also supporting a generalization in the form of constituent tracer equations. As seen in Part IV of this book, a dynamical description of fluid motion is also facilitated by working with constant mass fluid elements/regions.

16.1.6 Further study

An extension of the ideas presented in this section allows for chemical reactions to transfer mass from one constituent to others, while retaining fixed net mass. This extension is relevant for discussion of atmospheric chemistry and ocean biogeochemistry. Fundamental discussions of these extensions, using nomenclature similar to that used here, are provided in Chapter 11 of [Aris \(1962\)](#) and Chapter II of [DeGroot and Mazur \(1984\)](#).

16.2 Passive tracers

As defined in equation (16.17), the concentration of a material tracer is the mass of the trace constituent per mass of a fluid element. Such material tracers modify the barycentric velocity (16.9) since they carry mass and thus affect the mass density. We here introduce the related construct known as a *passive tracer*. A passive tracer satisfies the advection-diffusion equation, but it has zero impact on the velocity and is thus dynamically passive. Mathematically, we conceive of a passive tracer as a material tracer in the limit where the tracer mass and mass of the fluid element

together go to zero. The passive tracer is thus analogous to the massless fluid particle of Section 13.2 whose trajectories define the Lagrangian reference frame. We make use of passive tracers to probe the advective and diffusive features of the flow without modifying the flow. For example, a passive tracer can be used to define tracer pathways and time scales for transport between fluid regions.

In Chapter 14 and Section 15.4.2, we discussed the notion of a material fluid object, which is an object comprised of fluid particles that follow the velocity, \mathbf{v} . In a homogeneous fluid, such material objects are impenetrable to matter, by construction. For a multi-component fluid, trace matter will generally cross the material object through diffusion since $\mathbf{v}^{(n)} \neq \mathbf{v}$. Hence, there is no perfectly impenetrable fluid object in a fluid with any form of diffusion, including molecular diffusion. However, we can conceive of a passive tracer that follows the barycentric velocity and either diffuses or not. Again, these enhanced features are afforded the passive tracer given that it is a massless idealization used to probe the fluid flow properties. Hence, it is not subject to the same physical constraints as a material tracer.

16.3 Budgets for arbitrary fluid regions

Thus far in this chapter we have considered the evolution of mass within a variety of fluid regions, including infinitesimal and finite domains either moving with the fluid or fixed in space. We have also considered similar domains in Chapter 15 where the fluid domains were typically material regions. In this section we synthesize these presentations by considering mass budgets over an arbitrary finite sized domain within multi-component fluids. The resulting mass equations form the basis for matter budget analyses used in geophysical fluid mechanics.

16.3.1 Extensive and intensive fluid properties

Physical properties can be characterized as *extensive* or *intensive*. An extensive property changes when the size of the sample changes. Examples are particle number, mass, length, volume, kinetic energy, entropy, enthalpy, and linear momentum. An intensive property is not affected when removing some of the sample. Examples are number density (number of particles per volume), mass density (mass of substance per volume), tracer concentration (mass of tracer per mass of fluid), temperature, velocity (linear momentum per mass), kinetic energy per mass, entropy per mass, and enthalpy per mass.

To further this discussion, consider a bucket of seawater that has homogeneous temperature and salinity. Removing a cup of water from this bucket does not alter the temperature and salinity (both are intensive properties), but it does alter the enthalpy, salt mass, and freshwater mass (each are extensive properties). We thus make the distinction between intensive and extensive properties for the purpose of developing budget equations, with budget equations formulated for extensive properties.²

We are concerned in this section with how scalar extensive properties change as a function of time. Determining the evolution of such properties constitutes a budget analysis for the scalar property. What are the processes responsible for these changes? Where are the changes coming from? Those are basic questions asked when performing a budget analysis. In addition to physical and biogeochemical processes active within the fluid, details of the region over which one performs a budget have an important impact on the budget. Is the region open to matter and energy transport,

²We have more to say about intensive and extensive properties when considering thermodynamics in Chapter 21.

or is it closed? Is the region static (Eulerian) or do boundaries move? If the boundaries move, do they move with fluid elements (Lagrangian) or are they moving in some other manner (arbitrary)?

In the following, let Π represent an intensive scalar fluid property of a fluid element so that $\Pi \rho \delta V$ is the corresponding extensive property

$$\Pi = \text{intensive fluid property such as tracer concentration} \quad (16.26a)$$

$$\Pi \rho \delta V = \text{extensive fluid property such as tracer mass.} \quad (16.26b)$$

For example, if Π is the tracer concentration in a fluid element (i.e., mass of tracer per mass of fluid), then the corresponding extensive property, $\Pi \rho \delta V$, is the mass of tracer in the fluid element. Anticipating our discussion in Section 22.8, let Π be the Conservative Temperature, Θ , of a fluid element, then the corresponding extensive property, $\Theta c_p \rho \delta V$, is the potential enthalpy with c_p the specific heat capacity.

We furthermore assume that Π satisfies the scalar conservation equation, written here in both its material form and Eulerian flux-form

$$\rho \frac{D\Pi}{Dt} = -\nabla \cdot \mathbf{J} \quad \iff \quad \frac{\partial(\rho \Pi)}{\partial t} + \nabla \cdot (\rho \Pi \mathbf{v} + \mathbf{J}) = 0, \quad (16.27)$$

where \mathbf{J} is a flux such as that associated with the tracer equation derived in Section 16.1.3. Depending on the context, the budget equation (16.27) is sometimes referred to as a *conservation law* for Π . Notably, satisfaction of a conservation law does not mean that Π is constant either at a point in space nor following a fluid particle. Instead, there are two cases of “constancy” that naturally arise. First, with $-\nabla \cdot \mathbf{J} = 0$, the scalar field is constant following a material fluid particle

$$-\nabla \cdot \mathbf{J} = 0 \implies \frac{D\Pi}{Dt} = 0. \quad (16.28)$$

In this case we say that Π is a *material invariant* or a *material constant*. Second, if the Eulerian time derivative vanishes, $\partial_t \Pi = 0$, then Π remains constant at a fixed spatial point in the fluid and we say that Π is in a *steady state*. Furthermore, recall that the Eulerian reference frame is stationary with respect to a laboratory frame, with the laboratory frame inertial when connected by a Galilean transformation to the universal Newtonian reference frame (Section 13.3). Hence, if the flow in one laboratory frame is steady, then flow in all laboratory frames is steady so long as the laboratory frames are connected by a Galilean transformation (see Section 13.6).

16.3.2 General form of the finite domain integral

We are concerned here with the evolution of extensive fluid properties integrated over an arbitrary region. Let us make use of the following notation for such integrals

$$\mathcal{I}[\mathcal{R}(t), t] = \int_{\mathcal{R}(t)} \Pi \rho dV \equiv \int_{\mathcal{R}(t)} \varphi dV, \quad (16.29)$$

where we introduced the shorthand

$$\varphi = \rho \Pi. \quad (16.30)$$

The integrand in equation (16.29) is a function of space and time, $\varphi = \varphi(\mathbf{x}, t)$, and the integration region is generally a function of time, $\mathcal{R}(t)$. In previous sections, \mathcal{R} was a material region of fixed matter content (Section 15.3) or a constant mass fluid region open to the exchange of matter with the surroundings (Section 16.1). In both of these cases the region was denoted by $\mathcal{R}(\mathbf{v})$ since it moved with the fluid flow. Here we make no *a priori* assumption about how the region moves.

The total time derivative of \mathcal{I} can be written as

$$\frac{d\mathcal{I}}{dt} = \left[\frac{\partial \mathcal{I}}{\partial t} \right]_{\mathcal{R}} + \frac{d\mathcal{R}}{dt} \left[\frac{\partial \mathcal{I}}{\partial \mathcal{R}} \right]_t. \quad (16.31)$$

The first term on the right hand side is the time derivative of the integral when holding the region fixed in space as per an Eulerian time derivative. The second term accounts for changes due to evolution of the region as weighted by dependence of the integral on the region itself. How the integral changes in time depends on both the evolution of the fluid property relative to the chosen region and evolution of the fluid region itself. Equation (16.31) is directly analogous to the total time derivative of a field in a moving fluid as given by equation (13.27).

16.3.3 Eulerian (static) domain

We first consider an Eulerian domain, which is fixed in space and thus static so that

$$\frac{d\mathcal{I}}{dt} = \left[\frac{\partial \mathcal{I}}{\partial t} \right]_{\mathcal{R}} = \frac{\partial}{\partial t} \left[\int_{\mathcal{R}} \Pi \rho dV \right] = \int_{\mathcal{R}} \left[\frac{\partial(\rho \Pi)}{\partial t} \right] dV. \quad (16.32)$$

Movement of the time derivative across the integral sign is available since the domain boundaries are static; i.e., the second term on the right hand side of equation (16.31) vanishes. Furthermore, since the domain is static, the volume element, dV , provides a static partition of the total domain volume so that dV does not appear inside the time derivative. This case corresponds to the Eulerian budgets depicted in Figures 15.1, 15.2, and 16.1.

16.3.4 Leibniz-Reynolds transport theorem

We now allow the domain boundaries to be time dependent so that both terms in the total time derivative (16.31) contribute. The resulting *Leibniz-Reynolds transport theorem* is a general expression of conservation over an arbitrary region.

A naive derivation of the transport theorem

Consider a one-dimensional domain with time dependent endpoints. Integrals of this type commonly arise when integrating over the depth of the atmosphere or ocean, in which case the boundary terms are replaced by kinematic boundary conditions (see Section 15.4). The chain rule for differentiating integrals is known as *Leibniz's rule*. It results in the time derivative acting on the upper integral limit, the lower limit, and the integrand

$$\frac{d}{dt} \left[\int_{x_1(t)}^{x_2(t)} \varphi(x, t) dx \right] = \int_{x_1(t)}^{x_2(t)} \frac{\partial \varphi}{\partial t} dx + \frac{d}{dt} \left[\int_{x_1(t)}^{x_2(t)} \right] \varphi(x, t) dx \quad (16.33a)$$

$$= \int_{x_1(t)}^{x_2(t)} \frac{\partial \varphi}{\partial t} dx + \frac{dx_2(t)}{dt} \varphi(x_2, t) - \frac{dx_1(t)}{dt} \varphi(x_1, t) \quad (16.33b)$$

$$= \int_{x_1(t)}^{x_2(t)} \left[\frac{\partial \varphi}{\partial t} + \frac{\partial}{\partial x} \left(\varphi \frac{dx}{dt} \right) \right] dx. \quad (16.33c)$$

As a matter of convenience we brought the boundary terms back inside the integral for the final equality. Generalization of this result to a three-dimensional integral yields

$$\frac{d}{dt} \left[\int_{\mathcal{R}} \varphi dV \right] = \int_{\mathcal{R}} \left[\frac{\partial \varphi}{\partial t} + \nabla \cdot (\varphi \mathbf{v}^{(b)}) \right] dV = \int_{\mathcal{R}} \frac{\partial \varphi}{\partial t} dV + \oint_{\partial \mathcal{R}} \varphi \mathbf{v}^{(b)} \cdot \hat{\mathbf{n}} d\mathcal{S}, \quad (16.34)$$

where the second equality made use of the divergence theorem to transfer the volume integral into a boundary integral, and where we introduced the shorthand for the velocity of a point on the boundary

$$\mathbf{v}^{(b)} = \frac{d\mathbf{x}}{dt}. \quad (16.35)$$

Comments on the Leibniz-Reynolds Transport Theorem

The identity (16.34) is the *Leibniz-Reynolds transport theorem*. It is an incredibly important kinematic expression in fluid mechanics since it forms the starting point for all finite volume budget analyses.

We purposefully presented a rather naive derivation of the Leibniz-Reynolds transport theorem (16.34). In particular, we made use of an assumption in moving from the one dimensional to three dimensional result; namely, that the three dimensional domain is fully expressible by Cartesian coordinates whose extents are mutually independent. More rigorous derivations can be found perhaps as early as [Truesdell \(1954\)](#). We nonetheless presented the naive derivation to support intuition.

Even though we presented a naive derivation, the resulting expression (16.34) offers a coordinate invariant measure of how an extensive property evolves within a region. Hence, by the rules of tensor analysis from Chapter 7, the result holds for arbitrary coordinates with arbitrary dependencies. Furthermore, it holds for multiply connected domains for which one merely sums over the distinct sub-domains to develop the complete budget. These results make good sense based on our intuition regarding extensive properties, such as fluid mass, tracer mass, and enthalpy. The general nature of the Leibniz-Reynolds transport theorem is of particular use when studying water mass and tracer transformation analysis in Chapter 52, where budgets are developed over arbitrary regions.

The appearance of the boundary velocity, $\mathbf{v}^{(b)}$, warrants some comment. As defined by equation (16.35), it measures the velocity of a point on the domain boundary. However, the resulting budget only requires information about the normal component to that velocity, $\mathbf{v}^{(b)} \cdot \hat{\mathbf{n}}$. For example, the domain boundary could be exhibiting arbitrary motion in the direction tangent to the bounding surface. However, such tangential motion is of no concern for a budget developed over the domain. The reason is that we are only concerned with transport across the boundary when developing the budget. Indeed, information concerning the tangential component is not available without making dynamical assumptions that go beyond the kinematics considered here.

Transport theorem for region volume

Setting $\varphi = 1$ in the transport theorem (16.34) yields the time change for the region volume

$$\frac{d}{dt} \left[\int_{\mathcal{R}} dV \right] = \oint_{\partial\mathcal{R}} \mathbf{v}^{(b)} \cdot \hat{\mathbf{n}} d\mathcal{S}. \quad (16.36)$$

This result says that the volume for an arbitrary region changes in time so long as there is motion of the region boundary normal to itself. We can compare this expression to that for a material region given by equation (15.25). The expressions are the same since $\mathbf{v}^{(b)} \cdot \hat{\mathbf{n}} = \mathbf{v} \cdot \hat{\mathbf{n}}$ for a material region. The volume budget (16.36) holds for both compressible and incompressible fluids, with specialization to the incompressible case in Section 17.6.2.

Transport theorem for a scalar field

We here derive a corollary to the transport theorem (16.34) that proves useful for budget analyses over moving regions, with generalization to the vector linear momentum provided in Section 19.6.

For this purpose, make use of the Eulerian flux-form of the scalar conservation equation (16.27) so that the transport theorem is written

$$\frac{d}{dt} \left[\int_{\mathcal{R}} \rho \Pi dV \right] = - \oint_{\partial \mathcal{R}} [\rho \Pi (\mathbf{v} - \mathbf{v}^{(b)}) + \mathbf{J}] \cdot \hat{\mathbf{n}} d\mathcal{S}. \quad (16.37)$$

Setting $\Pi = 1$ gives an expression for the change in mass for the region

$$\frac{d}{dt} \left[\int_{\mathcal{R}} \rho dV \right] = - \oint_{\partial \mathcal{R}} \rho (\mathbf{v} - \mathbf{v}^{(b)}) \cdot \hat{\mathbf{n}} d\mathcal{S}. \quad (16.38)$$

The transport theorem (16.37) has a straightforward interpretation. Namely, the left hand side is the time tendency for the total Π -stuff within the moving region. The right hand side is the surface area integral of the flux of Π -stuff through the boundary of the region. The first right hand side term arises from the difference between the barycentric fluid velocity and the velocity of the boundary, and the second term arises from the diffusive flux. Both fluxes are projected onto the outward normal at the boundary and then integrated over the surface area. Hence, the budget is not affected by fluxes tangential to the boundary. Finally, for the mass budget (16.38), the diffusive flux vanishes since the mass of a fluid element moves according to the barycentric velocity of Section (16.1.2).

In Figure 16.2 we illustrate the transport theorem (16.37) for the special case of a discrete numerical model grid cell. This cell has fixed positions for the vertical sides whereas the top and bottom interfaces are time dependent. This application of the transport theorem provides the framework for finite volume methods in numerical models (e.g., [Griffies et al. \(2020\)](#)). We offer further discussion of the kinematics of such *general vertical coordinate* models in Chapter 44 and their dynamics in Chapter 45.

16.3.5 Revisiting Reynolds transport theorem

We here consider a region that is moving with the fluid flow, in which case we provide a more general derivation of the *Reynolds transport theorem* than originally given for material regions in Section 15.3.4. The following results are special cases of the general expression (16.37).

Reynolds Transport Theorem

Let us apply the result (16.34) to a region that follows the fluid flow as defined by the barycentric velocity, \mathbf{v} . For this moving region, the time derivative of the region boundaries in equation (16.34) is given by the fluid velocity thus leading to

$$\frac{d}{dt} \left[\int_{\mathcal{R}(\mathbf{v})} \varphi dV \right] = \int_{\mathcal{R}(\mathbf{v})} \left[\frac{\partial \varphi}{\partial t} + \nabla \cdot (\mathbf{v} \varphi) \right] dV = \int_{\mathcal{R}(\mathbf{v})} \left[\frac{D\varphi}{Dt} + \varphi \nabla \cdot \mathbf{v} \right] dV. \quad (16.39)$$

This result is the Reynolds transport theorem. The derivation given here is more general than that in Section 15.3.4 where we assumed the region to be material (i.e., no matter crosses the region boundary). For the present derivation we only assumed that the region boundaries move with the barycentric velocity, \mathbf{v} . We did not assume the region boundaries are material. We can thus make use of Reynolds transport theorem (16.39) for constant mass regions of a multi-component fluid so long as the region moves with the the barycentric velocity. Furthermore, the region boundary is generally permeable via the diffusion of tracers.

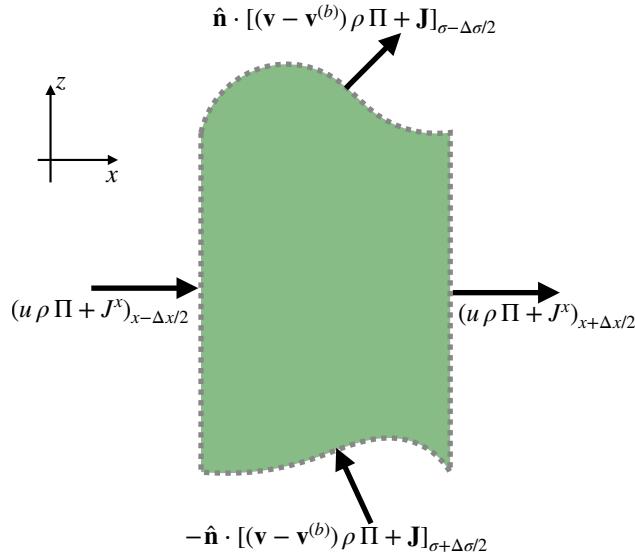


FIGURE 16.2: This figure depicts the contributions to the Leibniz-Reynolds transport theorem (16.37). The theorem is applied to a domain corresponding to a numerical model grid cell with the top and bottom interfaces defined by generalized vertical coordinates of Chapters 43, 44, and 45. In particular, the vertical cell faces are assumed to have fixed positions, so that $(\mathbf{v} - \mathbf{v}^{(b)}) \cdot \hat{\mathbf{n}} = \mathbf{v} \cdot \hat{\mathbf{n}}$ for these cell faces. Hence, the fluxes crossing these faces are due to advection by the barycentric velocity plus the diffusive flux. However, the top and bottom faces of the cell are allowed to move according to the generalized vertical coordinate surfaces. Hence, transport through these faces must take into account the nonzero velocity of the boundaries. Note that numerical models generally assume the top and bottom interfaces have a nonzero projection in the vertical direction so that they never overturn (e.g., Griffies et al. (2020)).

Alternative form of Reynolds Transport Theorem

We can put the Reynolds Transport Theorem (16.39) into another useful form by reintroducing $\varphi = \rho \Pi$ and making use of mass continuity

$$\frac{1}{\rho} \frac{D\rho}{Dt} = -\nabla \cdot \mathbf{v}. \quad (16.40)$$

Doing so yields the rather tidy result

$$\frac{d}{dt} \left[\int_{\mathcal{R}(\mathbf{v})} \Pi \rho dV \right] = \int_{\mathcal{R}(\mathbf{v})} \left[\frac{D\varphi}{Dt} + \varphi \nabla \cdot \mathbf{v} \right] dV \quad (16.41a)$$

$$= \int_{\mathcal{R}(\mathbf{v})} \left[\frac{D(\rho \Pi)}{Dt} + \rho \Pi \nabla \cdot \mathbf{v} \right] dV \quad (16.41b)$$

$$= \int_{\mathcal{R}(\mathbf{v})} \left[\Pi \left(\frac{D\rho}{Dt} + \rho \nabla \cdot \mathbf{v} \right) + \rho \frac{D\Pi}{Dt} \right] dV \quad (16.41c)$$

$$= \int_{\mathcal{R}(\mathbf{v})} \frac{D\Pi}{Dt} \rho dV. \quad (16.41d)$$

Heuristically, this result follows since ρdV is a constant when following the flow, so that passage of the time derivative across the integral only picks up the material derivative of Π .

We can take the result (16.41d) one more step by inserting the material form of the scalar

conservation equation (16.27) so that

$$\frac{d}{dt} \left[\int_{\mathcal{R}(\mathbf{v})} \Pi \rho dV \right] = - \oint_{\partial \mathcal{R}(\mathbf{v})} \mathbf{J} \cdot \hat{\mathbf{n}} d\mathcal{S}, \quad (16.42)$$

which is a special case of the general transport theorem (16.37) found by setting $(\mathbf{v} - \mathbf{v}^{(b)}) \cdot \hat{\mathbf{n}} = 0$ along the region boundary. This result says that the change in Π -stuff within a region moving with the barycentric velocity arises only from the area integrated diffusive flux crossing normal to the boundary. It is a finite volume generalization of the mass conservation statement for a fluid element as discussed in Section 16.1.3. We can set $\Pi = 1$ to render a statement of mass conservation for a Lagrangian region

$$\frac{d}{dt} \left[\int_{\mathcal{R}(\mathbf{v})} \rho dV \right] = 0, \quad (16.43)$$

where the diffusive flux, \mathbf{J} , vanishes for the mass.

16.3.6 Summary of the time derivatives acting on integrals

We here summarize the variety of time derivatives acting on integrals of scalar fields

$$\frac{d}{dt} \int_{\mathcal{R}} \rho \Pi dV = \begin{cases} \int_{\mathcal{R}} \frac{\partial(\rho \Pi)}{\partial t} dV = - \oint_{\partial \mathcal{R}} (\rho \mathbf{v} \Pi + \mathbf{J}) \cdot \hat{\mathbf{n}} d\mathcal{S} & \text{Eulerian } \mathcal{R} \\ \int_{\mathcal{R}(\mathbf{v})} \rho \frac{D\Pi}{Dt} dV = - \oint_{\partial \mathcal{R}(\mathbf{v})} \mathbf{J} \cdot \hat{\mathbf{n}} d\mathcal{S} & \text{Lagrangian } \mathcal{R}(\mathbf{v}) \\ - \oint_{\partial \mathcal{R}} [\rho \Pi (\mathbf{v} - \mathbf{v}^{(b)}) + \mathbf{J}] \cdot \hat{\mathbf{n}} d\mathcal{S} & \text{arbitrary } \mathcal{R}, \end{cases} \quad (16.44)$$

with the scalar fields assumed to satisfy the source-free differential equation

$$\rho \frac{D\Pi}{Dt} = -\nabla \cdot \mathbf{J} \iff \frac{\partial(\rho \Pi)}{\partial t} + \nabla \cdot (\rho \mathbf{v} \Pi + \mathbf{J}) = 0. \quad (16.45)$$

As discussed in Section 13.1, the partial differential equation (16.45) is referred to as the *strong formulation* of the scalar budget, whereas the integral expressions in equation (16.44) provide a variety of *weak formulations*.

16.4 Brute force illustration of Leibniz-Reynolds

The Leibniz-Reynolds transport theorem

$$\frac{d}{dt} \left[\int_{\mathcal{R}} \rho C dV \right] = - \int_{\partial \mathcal{R}} [\rho C (\mathbf{v} - \mathbf{v}^{(b)}) + \mathbf{J}] \cdot \hat{\mathbf{n}} d\mathcal{S}, \quad (16.46)$$

is an incredibly useful and elegant expression of the tracer budget over an arbitrary domain. Correspondingly, we make great use of it throughout this book. To further our understanding, we here consider the tracer budget for an ocean domain such as in Figure 16.3. Rather than make direct use of Leibniz-Reynolds, we use a brute force approach by expanding the volume integral according to

$$\frac{d}{dt} \left[\int_{\mathcal{R}} \rho C dV \right] = \frac{d}{dt} \left[\int_{A(t)} dA \int_{\eta_b}^{\eta} \rho C dz \right]. \quad (16.47)$$

In this equation, $\int_{A(t)} dA$ is an integral over the horizontal area of the domain, with the lateral boundaries of the domain generally a function of time. This exercise requires the use of various kinematic boundary conditions and provides further practice with the Leibniz rule.

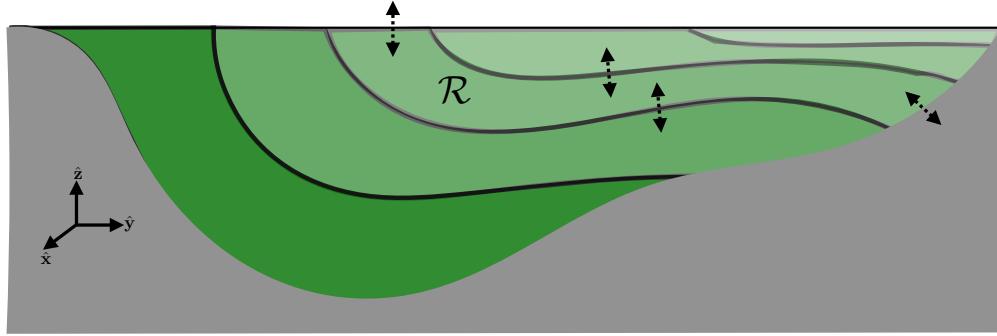


FIGURE 16.3: A depiction of fluid layers in which we formulate the budget for the total mass of tracer (or total potential enthalpy). The tracer mass within the layer, such as that one denoted by \mathcal{R} , is modified by dia-surface transport across interior layer interfaces, as well as transport across the surface and bottom boundaries. Note that an arbitrary layer might never intersect the bottom or surface boundaries. However, the layers depicted here each intersect boundaries, with such layers requiring extra care in formulating their tracer budgets.

16.4.1 Leibniz's rule plus kinematic boundary conditions

Performing the time derivative in equation (16.47) and using Leibniz's rule yields

$$\frac{d}{dt} \left[\int_{\mathcal{R}} \rho C dV \right] = \frac{dA}{dt} \left[\int_{\eta_b}^{\eta} \rho C dz \right]_{\text{bounds}} + \int_{A(t)} [\partial_t \eta (\rho C)_{z=\eta}] dA + \int_{A(t)} dA \int_{\eta_b}^{\eta} \frac{\partial(\rho C)}{\partial t} dz. \quad (16.48)$$

The first term on the right hand side is evaluated along the lateral boundaries of the domain. If the boundaries are fixed in time, as in a rectangular box of seawater or a periodic channel, then $dA/dt = 0$. The more general case has a lateral boundary that is time dependent such as along a beach where fluid moves up and down the sloping shoreline. However, in that case the thickness of fluid vanishes at the lateral boundary, $\eta - \eta_b = 0$, thus again revealing that the first term on the right hand side drops from the budget to render

$$\frac{d}{dt} \left[\int_{\mathcal{R}} \rho C dV \right] = \int_{A(t)} [\partial_t \eta (\rho C)_{z=\eta}] dA + \int_{A(t)} dA \int_{\eta_b}^{\eta} \frac{\partial(\rho C)}{\partial t} dz. \quad (16.49)$$

For the second term on the right hand side of equation (16.49) we make use of the tracer equation (16.59) and Leibniz's rule to write

$$\int_{\eta_b}^{\eta} \frac{\partial(\rho C)}{\partial t} dz = - \int_{\eta_b}^{\eta} \nabla_z \cdot (\rho C \mathbf{u} + \mathbf{J}_h) dz - \int_{\eta_b}^{\eta} \frac{\partial(\rho C w + J^z)}{\partial z} dz \quad (16.50a)$$

$$\begin{aligned} &= - \nabla_z \cdot \int_{\eta_b}^{\eta} (\rho C \mathbf{u} + \mathbf{J}_h) dz + \nabla(\eta - z) \cdot (\rho C \mathbf{v} + \mathbf{J})_{z=\eta} \\ &\quad + \nabla(z - \eta_b) \cdot (\rho C \mathbf{v} + \mathbf{J})_{z=\eta_b} \end{aligned} \quad (16.50b)$$

where we wrote $\mathbf{J} = \mathbf{J}_h + \hat{z} J^z$. The surface terms ($z = \eta$) combine with the $\partial_t \eta$ term appearing in equation (16.49) to yield

$$\rho C \left[\frac{\partial \eta}{\partial t} + \mathbf{u} \cdot \nabla \eta - w \right] = C Q_m, \quad (16.51)$$

where we used the surface kinematic boundary condition (15.79) to introduce the surface boundary mass flux Q_m . The bottom kinematic boundary condition eliminates the advective contribution at the bottom, $z = \eta_b$, via the no normal flow condition (15.41)

$$\nabla(z - \eta_b) \cdot \mathbf{v} = w - \mathbf{u} \cdot \nabla \eta_b = 0 \quad \text{at } z = \eta_b(x, y). \quad (16.52)$$

Finally, when integrated over the horizontal extent of the domain, the horizontal convergence term from equation (16.50b) vanishes. The reason it vanishes is because either the thickness of fluid vanishes at the horizontal boundaries (as along a beach); there is a no flux boundary condition if the boundary is a vertical wall; or the domain is periodic.

16.4.2 Summarizing the result

Bringing the results together yields the budget equation

$$\frac{d}{dt} \left[\int_{\mathcal{R}} \rho C dV \right] = \int_{z=\eta} (C Q_m + \nabla(\eta - z) \cdot \mathbf{J}) dA + \int_{z=\eta_b} \nabla(z - \eta_b) \cdot \mathbf{J} dA. \quad (16.53)$$

We now use the identity (15.68e) between horizontal area element, $dA = dx dy$, and area element on the surface

$$\hat{\mathbf{n}} d\mathcal{S} = -\nabla(\eta - z) dA \quad \text{at } z = \eta \quad (16.54a)$$

$$\hat{\mathbf{n}} d\mathcal{S} = -\nabla(z - \eta_b) dA \quad \text{at } z = \eta_b, \quad (16.54b)$$

to write

$$\frac{d}{dt} \left[\int_{\mathcal{R}} \rho C dV \right] = - \int_{z=\eta} (-C \mathcal{Q}_m + \hat{\mathbf{n}} \cdot \mathbf{J}) d\mathcal{S} - \int_{z=\eta_b} \hat{\mathbf{n}} \cdot \mathbf{J} d\mathcal{S}, \quad (16.55)$$

where

$$Q_m dA = \mathcal{Q}_m d\mathcal{S} = -\rho (\mathbf{v} - \mathbf{v}^{(\eta)}) \cdot \hat{\mathbf{n}} d\mathcal{S} \quad (16.56)$$

according to equation (15.73c). The budget for fluid mass is realized by setting C to a constant and thus dropping the non-advectional flux

$$\frac{d}{dt} \left[\int_{\mathcal{R}} \rho dV \right] = \int_{z=\eta} \mathcal{Q}_m d\mathcal{S} = - \int_{z=\eta} \rho (\mathbf{v} - \mathbf{v}^{(\eta)}) \cdot \hat{\mathbf{n}} d\mathcal{S}. \quad (16.57)$$

The manipulations in this section have succeeded in bringing the tracer and mass budgets into the form of the Leibniz-Reynolds transport theorem (16.46). The process of doing so required far more tedium as compared to the elegance of merely starting from equation (16.46). Even so, our efforts provide a useful means to ground the formalism by unpacking the many steps summarized by Leibniz-Reynolds.

16.5 Boundary conditions

We here extend our understanding of boundary conditions relevant at the variety of boundaries encountered by a fluid. We continue to focus on a fluid layer such as shown in Figure 16.3, paying particular interest to fluid layers that intersect surface (as for the ocean) and/or bottom boundaries (as for the ocean or atmosphere). We commonly think of this layer as defined by isosurfaces of generalized vertical coordinates whose layers are monotonically stacked in the vertical according to the discussion from Sections 43.8.1 and 44.3. However, the treatment given here allows for the layers to be non-monotonic in the vertical (e.g., overturns are allowed), so that these results can be used for the water mass transformation analysis discussed in Chapter 52. For example, the layers can be defined by surfaces of constant Conservative Temperature or salinity within the ocean, with these fields generally exhibiting regions of non-monotonic vertical stratification.

The Leibniz-Reynolds transport theorem (16.37) provides the starting point

$$\frac{d}{dt} \left[\int_{\mathcal{R}} \rho C dV \right] = - \int_{\partial \mathcal{R}} [\rho C (\mathbf{v} - \mathbf{v}^{(b)}) + \mathbf{J}] \cdot \hat{\mathbf{n}} dS, \quad (16.58)$$

where C is the concentration of a tracer that satisfies the advection-diffusion equation

$$\frac{\partial(\rho C)}{\partial t} + \nabla \cdot (\rho C \mathbf{v} + \mathbf{J}) = 0. \quad (16.59)$$

The left hand side of equation (16.58) is the time tendency for the mass of tracer within the region, such as the region \mathcal{R} shown in Figure 16.3. This tendency is affected by transport across the layer boundaries, with three boundaries considered here. We ignore interior sources, though note that the formalism can be readily extended in their presence.

16.5.1 Interior layer boundary conditions

The boundary transport across interior layer interfaces,

$$\text{interior boundary transport} = [\rho C (\mathbf{v} - \mathbf{v}^{(b)}) + \mathbf{J}] \cdot \hat{\mathbf{n}} dS, \quad (16.60)$$

measures the tracer mass transport due to advective fluxes across the moving layer boundaries (first term) and subgrid scale fluxes (second term). The advective flux is sometimes known as the *dia-surface transport*, with the kinematics of this transport discussed in Section 44.4.

16.5.2 Solid-earth boundary conditions

At the bottom boundary, the no-normal flow condition means that

$$\mathbf{v} \cdot \hat{\mathbf{n}} = 0. \quad (16.61)$$

Consider the velocity of a point attached to the layer interface, $\mathbf{v}^{(b)}$, and focus on where the interface intersects the bottom boundary. At this point, $\mathbf{v}^{(b)}$ tracks the position of the interface as it intersects the bottom boundary. By construction, the movement of this intersection point is tangential to the bottom boundary so that it too is orthogonal to the boundary outward normal direction

$$\mathbf{v}^{(b)} \cdot \hat{\mathbf{n}} = 0. \quad (16.62)$$

Hence, the only contribution to the tracer budget at the bottom boundary comes through the non-advective flux \mathbf{J}

$$\text{bottom boundary transport} = -\mathbf{J} \cdot \hat{\mathbf{n}} dS. \quad (16.63)$$

This equation says that if there is any transport through the bottom boundary (left hand side), then it induces a non-advective transport within the ocean whose normal component at the boundary equals to the bottom transport (right hand side).

Geothermal heating is the canonical solid-earth transport in the ocean. Assuming a known geothermal heat flux, $\mathcal{Q}_{\text{geo-heat}}$, it leads to a non-advective ocean boundary flux

$$\mathcal{Q}_{\text{geo-heat}} = -c_p \mathbf{J}(\Theta) \cdot \hat{\mathbf{n}}, \quad (16.64)$$

where c_p is the ocean heat capacity and Θ is the Conservative Temperature (see Chapter 51). Furthermore, if we assume the non-advective flux is parameterized as the downgradient diffusive flux (51.20), then the geothermal boundary condition (16.64) takes the form

$$\mathcal{Q}_{\text{geo-heat}} = c_p \rho (\mathbb{K} \cdot \nabla \Theta) \cdot \hat{\mathbf{n}}, \quad (16.65)$$

where \mathbb{K} is the diffusion tensor (Chapter 50).

For those cases where the geothermal heating vanishes, or more generally for tracers that have zero bottom boundary flux, then the tracer must satisfy the following no-normal flux (Neumann) boundary condition

$$\text{no flux bottom boundary} = (\mathbb{K} \cdot \nabla \Theta) \cdot \hat{\mathbf{n}} = 0. \quad (16.66)$$

In the case where diffusion next to the boundary is isotropic, as per molecular diffusion, then we reach the simpler result

$$\text{no flux bottom boundary} = \nabla C \cdot \hat{\mathbf{n}} = 0. \quad (16.67)$$

Namely, in this case, tracer isosurfaces are oriented normal to the boundary as depicted in Figure 16.4. Notably, this kinematic boundary condition holds at each point in time. For the dynamical tracers like temperature and salinity, this boundary condition affects flow near the boundary by modifying the density field and thus the pressure.

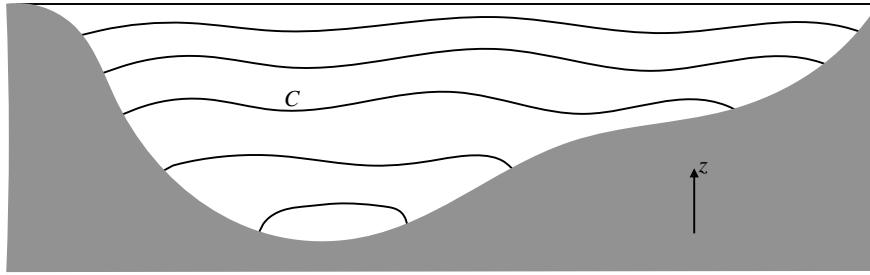


FIGURE 16.4: In the absence of a boundary tracer flux, and in the presence of isotropic downgradient diffusion, the isosurfaces of tracer, C , intersect solid boundaries normal to the boundary as per equation (16.67): $\nabla C \cdot \hat{\mathbf{n}} = 0$, where $\hat{\mathbf{n}}$ is the outward normal direction. This constraint holds at each time instance.

16.5.3 Upper ocean surface boundary conditions

Let us write the upper ocean surface boundary tracer transport as

$$\mathcal{Q}_C d\mathcal{S} = \text{net tracer mass per time crossing ocean surface}. \quad (16.68)$$

The surface boundary transport is generally comprised of two terms: a non-advection term just like at the solid-earth in Section 16.5.2, plus an advective term afforded since the ocean surface is permeable. If we assume that the tracer transported via the advected matter is either a dissolved tracer, such as salinity, or a thermodynamic tracer, such as Conservative Temperature, then we can write the net tracer flux as

$$\mathcal{Q}_C = C_m \mathcal{Q}_m + \mathcal{Q}_C^{\text{non-adv}}, \quad (16.69)$$

where $\mathcal{Q}_C^{\text{non-adv}}$ is the non-advection tracer flux, C_m is the tracer concentration within the mass transported across the surface, and \mathcal{Q}_m the mass per time per surface area of matter that crosses the boundary, as defined according to the kinematic boundary condition (15.63)

$$\mathcal{Q}_m = -\rho \hat{\mathbf{n}} \cdot (\mathbf{v} - \mathbf{v}^{(n)}) \quad \text{surface ocean boundary}. \quad (16.70)$$

As for the solid-earth boundary condition, specification of \mathcal{Q}_C requires information concerning the flux of tracer mass into or out of the ocean, and this flux equals to the net flux on the ocean side of the surface

$$\mathcal{Q}_C = C_m \mathcal{Q}_m + \mathcal{Q}_C^{\text{non-adv}} \equiv -[\rho C (\mathbf{v} - \mathbf{v}^{(b)}) + \mathbf{J}] \cdot \hat{\mathbf{n}} = C \mathcal{Q}_m - \mathbf{J} \cdot \hat{\mathbf{n}}, \quad (16.71)$$

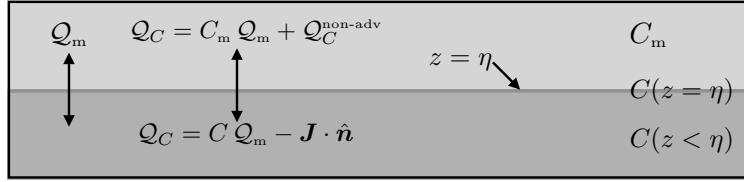


FIGURE 16.5: A schematic of an infinitesimal region of the ocean surface boundary at $z = \eta(x, y, t)$, with $z < \eta$ the ocean. $\mathcal{Q}_m d\mathcal{S}$ is the mass transport (mass per time) that crosses the interface and carries a tracer concentration C_m . $\mathcal{Q}_C d\mathcal{S}$ is the tracer mass transport that crosses the ocean surface. Continuity across the $z = \eta$ boundary means that the tracer mass transport at $z = \eta - \epsilon$ (ocean side) equals to that at $z = \eta + \epsilon$ (atmospheric side), with $\epsilon > 0$ a tiny number. The tracer concentration at the interface, $C(z = \eta)$, is not determined by kinematics. Many analyses and numerical model applications approximate $C(z = \eta)$ as the bulk tracer concentration within the upper few meters of the ocean. However, as model vertical resolution is refined, this assumption must be reconsidered.

We thus see that the surface transport of tracer mass induces the following non-advectional flux within the ocean at $z = \eta$

$$-\mathbf{J} \cdot \hat{\mathbf{n}} = \mathcal{Q}_C - C \mathcal{Q}_m = \mathcal{Q}_C^{\text{non-adv}} + (C_m - C) \mathcal{Q}_m. \quad (16.72)$$

Figure 16.5 offers a schematic to summarize these results. We make use of these results when discussing the surface ocean boundary conditions in Sections 51.3 and 52.9.3.

16.6 Evolution of region mean tracer concentration

We close this chapter by deriving the evolution equations for the domain averaged tracer concentration defined by

$$\langle C \rangle = M^{-1} \int_{\mathcal{R}} \rho C dV \quad \text{with} \quad M = \int_{\mathcal{R}} \rho dV. \quad (16.73)$$

We will then apply the resulting budgets to a numerical model grid cell.

16.6.1 Formulation

Use of the product rule leads to

$$\frac{d[\langle C \rangle M]}{dt} = M \frac{d\langle C \rangle}{dt} + \langle C \rangle \frac{dM}{dt} = M \frac{d\langle C \rangle}{dt} - \langle C \rangle \oint_{\partial\mathcal{R}} \rho (\mathbf{v} - \mathbf{v}^{(b)}) \cdot \hat{\mathbf{n}} d\mathcal{S}, \quad (16.74)$$

where the second equality made use of the mass budget (16.38). Inserting the transport theorem (16.37) for the left hand side yields

$$M \frac{d\langle C \rangle}{dt} = - \oint_{\partial\mathcal{R}} \left[\rho [C - \langle C \rangle] (\mathbf{v} - \mathbf{v}^{(b)}) + \mathbf{J} \right] \cdot \hat{\mathbf{n}} d\mathcal{S}. \quad (16.75)$$

The first term on the right hand side vanishes if the averaged concentration equals to the boundary concentration. That is, the region averaged tracer concentration is unchanged if the boundary fluxes of mass have a tracer concentration that matches the region average.

16.6.2 Application to a numerical ocean model grid cell

If the region is an ocean model grid cell that is adjacent to the ocean surface (see Figure 16.2 or 16.5), then use of the surface boundary condition (16.71) leads to

$$M \frac{d\langle C \rangle}{dt} = \int_{z=\eta} [C - \langle C \rangle] Q_m dA - \int_{\partial\mathcal{R}_{\text{int}}} \left[\rho [C - \langle C \rangle] (\mathbf{v} - \mathbf{v}^{(b)}) + \mathbf{J} \right] \cdot \hat{\mathbf{n}} d\mathcal{S} - \oint_{\partial\mathcal{R}} \mathbf{J} \cdot \hat{\mathbf{n}} d\mathcal{S}, \quad (16.76)$$

where $\partial\mathcal{R}_{\text{int}}$ is the interior boundary to the grid cell.

As noted above, the first term on the right hand side of equation (16.76) vanishes if $C(z = \eta) = \langle C \rangle$. This situation is commonly assumed for temperature in the surface grid cell of an ocean model. That is, the temperature of evaporation, precipitation, and river runoff is commonly taken as the temperature in the surface model grid cell. In contrast, $C(z = \eta) = 0$ is commonly the case for material tracers such as salt, whose concentration is zero within boundary water fluxes.

16.7 Exercises

EXERCISE 16.1: EQUATION FOR TRACER MASS PER FLUID VOLUME

In some treatments it can be suitable to define a volumetric tracer concentration as the mass of tracer per volume of fluid

$$\phi = \frac{\text{mass of constituent } n \text{ in fluid element}}{\text{volume of fluid element}} = C \rho, \quad (16.77)$$

where C is the mass concentration defined by equation (16.17) and satisfying the tracer equation (16.21). Derive the corresponding equation satisfied by ϕ .



Non-divergent flows

In this chapter, we study the kinematics of a non-divergent velocity field, $\nabla \cdot \mathbf{v} = 0$. In many areas of fluid mechanics, fluids with a non-divergent velocity are referred to as *incompressible* since the density is assumed to be constant everywhere; i.e., “incompressible”. However, as explored in Chapter 24, a non-divergent flow can still admit compressibility effects and the associated density variations, with the oceanic Boussinesq fluid the primary example studied in this book. Hence, we are motivated to distinguish between a fluid property and flow property. Namely, the Boussinesq fluid allows for an expression of compressibility (a property of the fluid), whereas the flow admitted by a Boussinesq fluid is non-divergent (a property of the flow). In this chapter we focus on the incompressible nature of the flow, with the non-divergent velocity written as the curl of a streamfunction. As we will see, the streamfunction plays a central role in the kinematics of this chapter.

READER’S GUIDE TO THIS CHAPTER

We presume an understanding of the kinematics of mass conservation from Chapter 15 as well as many of the results from Cartesian tensor analysis in Chapter 2. This chapter introduces many concepts and tools of use in the remainder of the book.

17.1	Introduction to non-divergent flow	332
17.2	Kinematic boundary conditions	332
17.3	Kinematic free surface equation	333
17.4	Streamfunction for two-dimensional flow	334
17.4.1	Streamfunction isolines are streamlines	334
17.4.2	Streamfunction is constant on material boundaries	335
17.4.3	The streamfunction and fluid transport	335
17.4.4	Gauge symmetry	336
17.4.5	Exact differential formulation	336
17.4.6	Concerning the Helmholtz decomposition	336
17.4.7	A caveat: transport with curl-free + divergent flow	338
17.5	Vector streamfunction for three-dimensional flow	338
17.5.1	Gauge symmetry	338
17.5.2	Vector streamfunction in the vertical gauge	338
17.5.3	The streamfunction and transport through a surface	339
17.5.4	Scalar streamfunctions and transport	340
17.5.5	Concerning a harmonic velocity potential	341
17.6	Evolution of volume and area	341
17.6.1	Material volumes and areas	341

17.6.2 Arbitrary volume and area	342
17.7 Meridional-depth overturning circulation	342
17.8 Exercises	344

17.1 Introduction to non-divergent flow

For many purposes in fluid mechanics, we can make the simplifying assumption regarding the fluid kinematics whereby the volume of a fluid element is approximated as a constant. In particular, this situation holds for fluid flow satisfying the oceanic Boussinesq approximation (see Chapter 24). Recalling the expression

$$\frac{1}{\delta V} \frac{D(\delta V)}{Dt} = \nabla \cdot \mathbf{v} \quad (17.1)$$

from Section 14.4.1, we see that a constant volume for a fluid element constrains the velocity field to be non-divergent

$$\frac{1}{\delta V} \frac{D(\delta V)}{Dt} = 0 \implies \nabla \cdot \mathbf{v} = 0. \quad (17.2)$$

Flow satisfying $\nabla \cdot \mathbf{v} = 0$ is said to be *incompressible* since the volume of a fluid element is materially invariant. We illustrate this situation in Figure 17.1.

A slightly less onerous constraint arises from the anelastic approximation, whereby

$$\nabla \cdot (\rho \mathbf{v}) = 0. \quad (17.3)$$

The anelastic approximation is sometimes motivated for the atmosphere. However, it is less commonly used for atmospheric dynamics than the Boussinesq approximation is used for the ocean. We thus focus on the Boussinesq case here, whereby $\nabla \cdot \mathbf{v} = 0$.

The non-divergence constraint (17.2) reduces by one the number of functional degrees of freedom possessed by the velocity field. What that means in practice is that we need one fewer velocity component to determine the flow. That is, one velocity component is specified by the other components. This property manifests by our ability to introduce a streamfunction to specify the velocity. We further develop this point in the following.

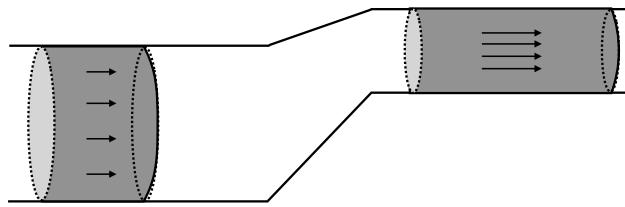


FIGURE 17.1: Illustrating volume continuity for an incompressible liquid flowing in a pipe. On the left the pipe has a relatively large diameter whereas on the right the pipe is narrower. A plug of water on the left moves through the pipe and becomes longer when it moves into the narrower region so that the volume of the plug remains the same. Correspondingly, the speed of the flow increases when moving into the narrower portion of the pipe.

17.2 Kinematic boundary conditions

For incompressible flow, there are slight modifications to the compressible boundary conditions detailed in Section 15.4. Whereas the material conditions remain identical, the non-material conditions are applied with a constant reference density, ρ_0 , rather than the local *in situ* density, ρ .

The reason is that we switch from specifying a mass transport condition as per equation (15.63) to a volume transport condition

$$\rho_0 (\mathbf{v} - \mathbf{v}^{(s)}) \cdot \hat{\mathbf{n}} d\mathcal{S} = -Q_m d\mathcal{S} = -Q_m dA \quad \text{moving non-material boundary condition,} \quad (17.4)$$

where the second equality introduced the mass flux per unit horizontal area, Q_m , according to equation (15.73c). Correspondingly, the kinematic boundary condition (15.79) applied at the ocean free surface takes on the form

$$\rho_0 \frac{D(z - \eta)}{Dt} = -Q_m \implies w + \frac{Q_m}{\rho_0} = \frac{\partial \eta}{\partial t} + \mathbf{u} \cdot \nabla \eta. \quad (17.5)$$

17.3 Kinematic free surface equation

We here derive the volume budget over a column of fluid. This budget provides a kinematic expression for the free surface evolution in an incompressible flow. For this purpose, we vertically integrate the non-divergence constraint, $\nabla \cdot \mathbf{v} = 0$, over the depth of an ocean column, from $z = \eta_b(x, y)$ at the bottom to $z = \eta(x, y, t)$ at the free surface and use the bottom and surface kinematic boundary conditions. This calculation yields

$$0 = \int_{\eta_b}^{\eta} \nabla \cdot \mathbf{v} dz \quad (17.6a)$$

$$= w(\eta) - w(\eta_b) + \int_{\eta_b}^{\eta} \nabla \cdot \mathbf{u} dz \quad (17.6b)$$

$$= w(\eta) - w(\eta_b) + \nabla \cdot \left[\int_{\eta_b}^{\eta} \mathbf{u} dz \right] - \mathbf{u}(\eta) \cdot \nabla \eta + \mathbf{u}(\eta_b) \cdot \nabla \eta_b \quad (17.6c)$$

$$= [w(\eta) - \mathbf{u}(\eta) \cdot \nabla \eta] - [w(\eta_b) - \mathbf{u}(\eta_b) \cdot \nabla \eta_b] + \nabla \cdot \left[\int_{\eta_b}^{\eta} \mathbf{u} dz \right], \quad (17.6d)$$

where we made use of Leibniz's Rule to move the horizontal divergence outside of the integral. We now make use of the surface kinematic boundary condition (17.5) and the bottom no-flow condition

$$w(\eta) - \mathbf{u} \cdot \nabla \eta = -\frac{Q_m}{\rho_0} + \frac{\partial \eta}{\partial t} \quad z = \eta \quad (17.7a)$$

$$w = \mathbf{u} \cdot \nabla \eta_b \quad z = \eta_b \quad (17.7b)$$

to render the free surface equation for a fluid with a non-divergent flow

$$\frac{\partial \eta}{\partial t} = \frac{Q_m}{\rho_0} - \nabla \cdot \mathbf{U}, \quad (17.8)$$

where

$$\mathbf{U} = \int_{\eta_b}^{\eta} \mathbf{u} dz \quad (17.9)$$

is the depth integrated horizontal transport. For the special case of a steady state with zero boundary mass flux, the depth integrated flow is non-divergent

$$\nabla \cdot \mathbf{U} = 0 \quad \text{if } Q_m = 0 \text{ and } \partial \eta / \partial t = 0. \quad (17.10)$$

Comparing the incompressible free surface quation (17.8) to the compressible free surface equation (15.84) indicates that the incompressible case is missing a contribution from the material changes in density. These changes arise from mixing and boundary fluxes of buoyancy. The absence of an impact from surface buoyancy fluxes means that the free surface in a Boussinesq fluid is not impacted by global thermal expansion, such as that arising from ocean warming. [Greatbatch \(1994\)](#) and [Griffies and Greatbatch \(2012\)](#) provide a recipe for diagnostically addressing this formulational limitation, thus enabling the study of global mean sea level with Boussinesq ocean models.

17.4 Streamfunction for two-dimensional flow

Vertical stratification of buoyancy plus the effects from planetary rotation inhibit vertical motion in geophysical flows. Therefore, as an idealization it is sometimes useful to assume the geophysical fluid flow is horizontal (two-dimensional) as well as non-divergent. The incompressible constraint for two-dimensional flow can be satisfied by writing the horizontal velocity in the form

$$\mathbf{u} = \nabla \wedge (z \nabla \psi) = \hat{\mathbf{z}} \wedge \nabla \psi = -\hat{\mathbf{x}} \frac{\partial \psi}{\partial y} + \hat{\mathbf{y}} \frac{\partial \psi}{\partial x}, \quad (17.11)$$

where

$$\nabla_z = \hat{\mathbf{x}} \partial_x + \hat{\mathbf{y}} \partial_y \quad (17.12)$$

is the horizontal gradient operator. The constraint $\nabla_z \cdot \mathbf{u} = 0$ is satisfied since the partial derivative operators commute

$$\frac{\partial^2 \psi}{\partial x \partial y} = \frac{\partial^2 \psi}{\partial y \partial x}. \quad (17.13)$$

We refer to ψ as the *streamfunction*, with this name motivated by the following considerations.

Note that in this section we could relax notation by dispensing with the z subscript on the horizontal gradient operator, ∇_z , since we are here concerned only with two-dimensional horizontal flow. However, it is useful to be a bit pedantic to distinguish the two-dimensional formulations in this section from the analogous three-dimensional case considered in Section 17.5.

17.4.1 Streamfunction isolines are streamlines

At any fixed time, the total differential of the streamfunction is

$$d\psi = \frac{\partial \psi}{\partial x} dx + \frac{\partial \psi}{\partial y} dy = v dx - u dy, \quad (17.14)$$

where the second equality follows from equation (17.11). Instantaneous lines along which ψ is a constant satisfy

$$d\psi = 0 \implies \frac{dx}{u} = \frac{dy}{v}. \quad (17.15)$$

Furthermore, the normal direction to constant ψ lines

$$\hat{\mathbf{n}} = \frac{\nabla_z \psi}{|\nabla_z \psi|} = \frac{v \hat{\mathbf{x}} - u \hat{\mathbf{y}}}{|\mathbf{u}|} \quad (17.16)$$

is normal to the velocity

$$\mathbf{u} \cdot \nabla_z \psi = u v - v u = 0. \quad (17.17)$$

Consequently, at each time instance, lines of constant ψ are streamlines (see Section 13.8.2 for discussion of streamlines). This property motivates the name *streamfunction*. Furthermore, through each point of a two-dimensional non-divergent flow and at any particular instance in time, there is one and only one streamline passing through that point.

17.4.2 Streamfunction is constant on material boundaries

As a corollary to the results from Section 17.4.1, we know that the streamfunction is a spatial constant when evaluated along material boundaries where $\mathbf{u} \cdot \hat{\mathbf{n}} = 0$. This property follows from equation (17.17). We can also see it from

$$\mathbf{0} = \mathbf{u} \cdot \hat{\mathbf{n}} = (\hat{\mathbf{z}} \wedge \nabla_z \psi) \cdot \hat{\mathbf{n}} = (\hat{\mathbf{n}} \wedge \hat{\mathbf{z}}) \cdot \nabla_z \psi = \hat{\mathbf{t}} \cdot \nabla_z \psi, \quad (17.18)$$

where $\hat{\mathbf{t}}$ a unit vector pointing tangent to the boundary. The operator $\hat{\mathbf{t}} \cdot \nabla_z \psi$ is the derivative of ψ computed along the boundary tangent at any given point along the boundary. Hence, $\hat{\mathbf{t}} \cdot \nabla_z \psi = 0$ means that ψ is a spatial constant along the boundary. Even though spatially constant, ψ along the boundary is generally a function of time.

17.4.3 The streamfunction and fluid transport

Consider an arbitrary curve in the fluid with endpoints \mathbf{x}_1 and \mathbf{x}_2 as depicted in Figure 17.2. At any particular time instance, the difference in streamfunction between these two points is given by

$$\psi(\mathbf{x}_2) - \psi(\mathbf{x}_1) = \int_{\mathbf{x}_1}^{\mathbf{x}_2} d\psi = \int_{\mathbf{x}_1}^{\mathbf{x}_2} \left[dx \frac{\partial \psi}{\partial x} + dy \frac{\partial \psi}{\partial y} \right] = \int_{\mathbf{x}_1}^{\mathbf{x}_2} \nabla_z \psi \cdot d\mathbf{x} = \int_{\mathbf{x}_1}^{\mathbf{x}_2} \nabla_z \psi \cdot \hat{\mathbf{t}} ds. \quad (17.19)$$

For the final equality we wrote

$$d\mathbf{x} = \hat{\mathbf{t}} ds, \quad (17.20)$$

where

$$ds = |\mathbf{dx}| \quad (17.21)$$

is the element of arc-length along the curve, and $\hat{\mathbf{t}}$ is a unit tangent vector that points in the direction along the curve from \mathbf{x}_1 to \mathbf{x}_2 . Now introduce the normal vector along the curve according to

$$\hat{\mathbf{t}} = \hat{\mathbf{n}} \wedge \hat{\mathbf{z}}, \quad (17.22)$$

where $\hat{\mathbf{n}}$ points to the left when facing in the $\hat{\mathbf{t}}$ direction. We thus have

$$\psi(\mathbf{x}_2) - \psi(\mathbf{x}_1) = \int_{\mathbf{x}_1}^{\mathbf{x}_2} \nabla_z \psi \cdot (\hat{\mathbf{n}} \wedge \hat{\mathbf{z}}) ds = \int_{\mathbf{x}_1}^{\mathbf{x}_2} (\hat{\mathbf{z}} \wedge \nabla_z \psi) \cdot \hat{\mathbf{n}} ds = \int_{\mathbf{x}_1}^{\mathbf{x}_2} \mathbf{u} \cdot \hat{\mathbf{n}} ds, \quad (17.23)$$

with the final equality an expression for the net area transport of fluid normal to the curve (dimensions of area per time). As the chosen curve connecting the points is arbitrary, we conclude that the difference in streamfunction values between two points measures the transport across any curve connecting the points. Correspondingly, the stronger the gradient in the streamfunction, the larger the transport since

$$|\mathbf{u}| = |\nabla_z \psi|. \quad (17.24)$$

The connection between the transport between two points and the value of the streamfunction at those two points sometimes prompts the name *transport streamfunction*. More specifically, this term is used when the streamfunction is computed for the depth integrated flow (17.9), in the special case when that flow has no divergence, $\nabla \cdot \mathbf{U} = 0$. In this case we can write

$$\mathbf{U} = \nabla \wedge (z \nabla \Psi) = \hat{\mathbf{z}} \wedge \nabla \Psi, \quad (17.25)$$

where Ψ is known as the *transport streamfunction* and it has dimensions $L^3 T^{-1}$.

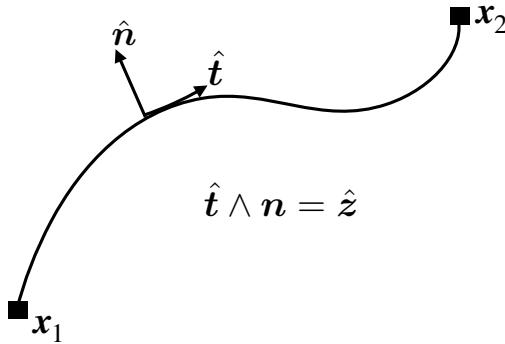


FIGURE 17.2: Depicting the transport between two points in a two-dimensional fluid. The transport is the line integral, $\int_{x_1}^{x_2} \mathbf{u} \cdot \hat{\mathbf{n}} ds$, from point x_1 to x_2 , with the normal direction $\hat{\mathbf{n}}$ pointing to the left when facing in the direction of the local tangent vector, $\hat{\mathbf{t}}$. By construction, $\hat{\mathbf{t}} \wedge \hat{\mathbf{n}} = \hat{\mathbf{z}}$, where $\hat{\mathbf{z}}$ points vertically out of the page. For a two-dimensionally incompressible flow, $\nabla_z \cdot \mathbf{u} = 0$, the transport between any two points is the difference in the streamfunction computed at these points, $\int_{x_1}^{x_2} \mathbf{u} \cdot \hat{\mathbf{n}} ds = \psi(x_1) - \psi(x_2)$. This result holds regardless the path taken between these two points, so long as the path remains simple; i.e., it does not intersect itself.

17.4.4 Gauge symmetry

For a two-dimensional non-divergent flow, the constraint $\nabla_z \cdot \mathbf{u} = 0$ reduces the functional degrees of freedom from two (the two velocity components (u, v)) to one (the streamfunction). However, the streamfunction is arbitrary up to a constant, k , since

$$\psi' = \psi + k \Rightarrow \mathbf{u}' = \mathbf{u}. \quad (17.26)$$

So the value of the streamfunction at a particular point has no unambiguous physical meaning. Rather, only the difference in streamfunction between two points is physically relevant. The ability to add a constant to the streamfunction is termed a *gauge symmetry*. We return to gauge symmetry for the three dimensional flow in Section 17.5.1.

17.4.5 Exact differential formulation

We here connect our discussion of velocity streamfunction to the discussion of exact differentials in Section 2.8. For that purpose introduce the differential

$$\mathbf{A} \cdot d\mathbf{x} \equiv (\mathbf{u} \wedge \hat{\mathbf{z}}) \cdot d\mathbf{x} = v dx - u dy. \quad (17.27)$$

By construction

$$\hat{\mathbf{z}} \cdot (\nabla_z \wedge \mathbf{A}) = 0 \quad \text{since} \quad \nabla_z \cdot \mathbf{u} = 0, \quad (17.28)$$

which means that $\mathbf{A} \cdot d\mathbf{x}$ is an exact differential (Section 2.8.1). Consequently, we can write

$$\mathbf{A} \cdot d\mathbf{x} = (\mathbf{u} \wedge \hat{\mathbf{z}}) \cdot d\mathbf{x} = \nabla_z \psi \cdot d\mathbf{x} = d\psi, \quad (17.29)$$

which then leads to the results derived earlier in this section where ψ is the transport streamfunction.

17.4.6 Concerning the Helmholtz decomposition

We close our discussion in this section by tidying up some mathematical niceties concerning the Helmholtz decomposition introduced in Section 2.9. For two-dimensional flows the decomposition takes the form

$$\mathbf{u} = \hat{\mathbf{z}} \wedge \nabla_z \Gamma + \nabla_z \Phi, \quad (17.30)$$

for some functions Γ and Φ . For non-divergent flows, Φ is constrained to be harmonic¹

$$\nabla \cdot \mathbf{u} = 0 \implies \nabla_z^2 \Phi = \nabla_z \cdot \nabla_z \Phi = 0. \quad (17.31)$$

As summarized in Table 17.1, it is sufficient to make use of just a streamfunction, ψ , for vortical flow and just a *velocity potential*, ϕ , for irrotational flow.² In the following we verify why it is sufficient to make use of this truncated version of the Helmholtz decomposition for non-divergent two-dimensional flows.

NON-DIVERGENT VORTICAL FLOW	NON-DIVERGENT IRROTATIONAL FLOW
$\nabla_z \cdot \mathbf{u} = 0$	$\nabla_z \cdot \mathbf{u} = 0$
$\nabla_z \wedge \mathbf{u} \neq 0$	$\nabla_z \wedge \mathbf{u} = 0$
$\mathbf{u} = \hat{\mathbf{z}} \wedge \nabla_z \psi$	$\mathbf{u} = \nabla_z \phi$
$\hat{\mathbf{z}} \cdot (\nabla_z \wedge \mathbf{u}) = \nabla_z^2 \psi$	$\nabla_z^2 \phi = 0$.

TABLE 17.1: Summarizing some mathematical properties of non-divergent two-dimensional velocity fields, $\nabla \cdot \mathbf{u} = 0$. The streamfunction is ψ whereas the harmonic velocity potential is ϕ .

Non-divergent vortical flow

Return to the exact differential formulation from Section 17.4.5. In that formulation we noted that $\nabla_z \cdot \mathbf{u} = 0$ means that the differential $\mathbf{A} \cdot d\mathbf{x} = (\mathbf{u} \wedge \hat{\mathbf{z}}) \cdot d\mathbf{x}$ is exact. Making use of the Helmholtz decomposition (17.30) renders

$$\mathbf{A} \cdot d\mathbf{x} = (\mathbf{u} \wedge \hat{\mathbf{z}}) \cdot d\mathbf{x} \quad (17.32a)$$

$$= [(\hat{\mathbf{z}} \wedge \nabla_z \Gamma) \wedge \hat{\mathbf{z}} + \nabla_z \Phi \wedge \hat{\mathbf{z}}] \cdot d\mathbf{x} \quad (17.32b)$$

$$= [\nabla_z \Gamma + \nabla_z \Phi \wedge \hat{\mathbf{z}}] \cdot d\mathbf{x}. \quad (17.32c)$$

To reveal the exactness of the right hand side requires the harmonic property of Φ so that we can write

$$\hat{\mathbf{z}} \cdot [\nabla_z \wedge (\nabla_z \Phi \wedge \hat{\mathbf{z}})] = -\nabla_z^2 \Phi = 0 \implies \nabla_z \Phi \wedge \hat{\mathbf{z}} \equiv \nabla_z \Upsilon, \quad (17.33)$$

in which case

$$\mathbf{A} \cdot d\mathbf{x} \equiv v dx - u dy = d(\Gamma + \Upsilon) \equiv d\psi. \quad (17.34)$$

We conclude that for non-divergent vortical flow, we lose no generality by working just with the streamfunction, ψ , of Section 17.4.3. There is no need to also include a harmonic function.

Non-divergent irrotational flow

The discussion is analogous for non-divergent and irrotational flow, with the irrotational condition holding so long as Γ is harmonic

$$\hat{\mathbf{z}} \cdot [\nabla_z \wedge (\hat{\mathbf{z}} \wedge \nabla_z \Gamma)] = \nabla_z^2 \Gamma = 0. \quad (17.35)$$

Consequently, we can write

$$\hat{\mathbf{z}} \wedge \nabla_z \Gamma = \nabla_z \gamma, \quad (17.36)$$

in which case

$$\mathbf{u} = \hat{\mathbf{z}} \wedge \nabla_z \Gamma + \nabla_z \Phi = \nabla_z (\gamma + \Phi) \equiv \nabla_z \phi. \quad (17.37)$$

Hence, for non-divergent irrotational flow, it is sufficient to work just with the harmonic velocity potential, ϕ .

¹Recall our discussion of harmonic functions in Sections 2.2.2 and 3.5.1.

²The vorticity, $\nabla_z \wedge \mathbf{u}$, is a measure of the spin in the fluid and is the focus of Part VI of this book.

17.4.7 A caveat: transport with curl-free + divergent flow

Consider a horizontal velocity that has a non-zero divergence, $\nabla \cdot \mathbf{u} \neq 0$, and yet it has a zero curl, $\nabla \wedge \mathbf{u} = 0$. The zero curl allows us to write $\mathbf{u} = \nabla_z \phi$, with ϕ the velocity potential. Hence, $d\Phi = \nabla_z \phi \cdot d\mathbf{x}$ is an exact differential and so its closed loop integral vanishes: $\oint d\Phi = 0$. However, there is no connection between velocity potential and transport. That is, we cannot conclude anything about the net transport across a closed curve based on properties of ϕ .

17.5 Vector streamfunction for three-dimensional flow

A three-dimensional non-divergent velocity, $\nabla \cdot \mathbf{v} = 0$, can be specified by a vector streamfunction

$$\mathbf{v} = \nabla \wedge \Psi. \quad (17.38)$$

The constraint $\nabla \cdot \mathbf{v} = 0$ is trivially satisfied since the divergence of the curl vanishes

$$\nabla \cdot (\nabla \wedge \Psi) = 0. \quad (17.39)$$

As we will see, the non-divergence condition reduces the three functional degrees of freedom for \mathbf{v} from three to two plus a gauge function.

17.5.1 Gauge symmetry

For three-dimensional non-divergent flow, the constraint $\nabla \cdot \mathbf{v} = 0$ reduces the three functional degrees of freedom down to the two available from the vector streamfunction, $\mathbf{v} = \nabla \wedge \Psi$. Gauge symmetry manifests through the ability to add the gradient of an arbitrary function to Ψ without altering \mathbf{v} :

$$\Psi' = \Psi + \nabla \lambda \Rightarrow \mathbf{v}' = \mathbf{v}, \quad (17.40)$$

which follows since $\nabla \wedge \nabla \lambda = 0$. Hence, the vector streamfunction has no absolute physical meaning since we can always modify the streamfunction by adding an arbitrary gauge function. Even so, we find it useful to consider certain vector streamfunctions that have direct connection to fluid transport, as discussed above in Section 17.5.3.

17.5.2 Vector streamfunction in the vertical gauge

We here exploit gauge symmetry to determine the transport streamfunction that is related to the horizontal volume transport of fluid beneath a chosen depth. For this purpose, write the component expression for the velocity

$$u = \frac{\partial \Psi_3}{\partial y} - \frac{\partial \Psi_2}{\partial z} \quad \text{and} \quad v = \frac{\partial \Psi_1}{\partial z} - \frac{\partial \Psi_3}{\partial x} \quad \text{and} \quad w = \frac{\partial \Psi_2}{\partial x} - \frac{\partial \Psi_1}{\partial y}, \quad (17.41)$$

where the Cartesian components to the vector streamfunction are

$$\Psi = \hat{\mathbf{x}} \Psi_1 + \hat{\mathbf{y}} \Psi_2 + \hat{\mathbf{z}} \Psi_3. \quad (17.42)$$

We follow studies of ocean mesoscale eddy parameterizations (e.g., see Sections 48.7.1 and 50.1) by choosing the *vertical gauge*. In this gauge we set $\nabla \lambda$ to precisely cancel the third component of the vector streamfunction so that

$$\Psi_3 = 0 \iff \text{vertical gauge}, \quad (17.43)$$

which then renders

$$u = -\frac{\partial \Psi_2}{\partial z} \quad \text{and} \quad v = \frac{\partial \Psi_1}{\partial z} \quad \text{and} \quad w = \frac{\partial \Psi_2}{\partial x} - \frac{\partial \Psi_1}{\partial y}. \quad (17.44)$$

Vertically integrating the u, v equations from the bottom at $z = \eta_b(x, y)$ up to an arbitrary geopotential leads to

$$\Psi = \int_{\eta_b}^z \mathbf{u} dz' \wedge \hat{\mathbf{z}} \equiv \mathbf{U}(z) \wedge \hat{\mathbf{z}}, \quad (17.45)$$

where $\mathbf{U}(z) = \int_{\eta_b}^z \mathbf{u} dz'$ is the horizontal transport of fluid beneath a chosen vertical position. By construction, $u = -\partial \Psi_2 / \partial z$ and $v = \partial \Psi_1 / \partial z$. We verify that this streamfunction also renders w through noting that

$$\frac{\partial \Psi_2}{\partial x} = u(\eta_b) \partial_x \eta_b - \int_{\eta_b}^z \partial_x u dz' \quad \text{and} \quad \frac{\partial \Psi_1}{\partial y} = \int_{\eta_b}^z \partial_y v dz' - v(\eta_b) \partial_y \eta_b, \quad (17.46)$$

so that

$$\frac{\partial \Psi_2}{\partial x} - \frac{\partial \Psi_1}{\partial y} = \mathbf{u}(\eta_b) \cdot \nabla \eta_b - \int_{\eta_b}^z \left[\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} \right] dz' = w(\eta_b) + \int_{\eta_b}^z \frac{\partial w}{\partial z'} dz' = w(z), \quad (17.47)$$

where we used the bottom kinematic boundary condition (15.41) to write $w - \mathbf{u} \cdot \nabla \eta_b = 0$ at $z = \eta_b(x, y)$. Hence, knowledge of the vector streamfunction (17.45) offers a means to compute the three velocity components. Of course, the velocity is needed to compute the vector streamfunction! Even though the logic is circular, we are satisfied that the circle closes to reveal self-consistency of the formalism. Furthermore, in some contexts it is more suitable to first compute the transport than the velocity, such as in ocean mesoscale eddy parameterizations (e.g., Chapter 9 of [Griffies \(2004\)](#)). We also note that the transport is a smoother field than the velocity given that it is computed as an integral of the velocity.

We close by noting that the transport through the solid-earth bottom at $z = \eta_b(x, y)$ vanishes according to equation (17.50). We can trivially verify this result for the vertical gauge since

$$\Psi(z = \eta_b) = 0, \quad (17.48)$$

so that $\oint_{\partial\mathcal{S}} \Psi \cdot \hat{\mathbf{t}} ds = 0$ on the bottom.

17.5.3 The streamfunction and transport through a surface

The volume transport (volume per time) of fluid crossing a surface is defined by the area integral

$$\mathcal{T}(\mathcal{S}) = \int_{\mathcal{S}} \mathbf{v} \cdot \hat{\mathbf{n}} d\mathcal{S}, \quad (17.49)$$

where $\hat{\mathbf{n}}$ is the outward unit normal vector on the surface. Introducing the vector streamfunction and making use of Stokes' Theorem (Section 2.6) then leads to

$$\mathcal{T}(\mathcal{S}) = \int_{\mathcal{S}} \mathbf{v} \cdot \hat{\mathbf{n}} d\mathcal{S} = \int_{\mathcal{S}} (\nabla \wedge \Psi) \cdot \hat{\mathbf{n}} d\mathcal{S} = \oint_{\partial\mathcal{S}} \Psi \cdot \hat{\mathbf{t}} ds, \quad (17.50)$$

where $\hat{\mathbf{t}} ds$ is the oriented arc distance increment along the boundary of \mathcal{S} , and $\oint_{\partial\mathcal{S}}$ is the oriented line integral around the boundary $\partial\mathcal{S}$. Hence, the volume transport of fluid through the surface depends only on the vector streamfunction on the perimeter of the surface. Furthermore, if the

transport through the surface vanishes (e.g., no-flux material surface such as a solid earth boundary), then on the surface the vector streamfunction can be written as the gradient of an arbitrary scalar field, $\Psi = \nabla\chi$, since

$$\oint_{\partial S} \Psi \cdot \hat{t} ds = \oint_{\partial S} \nabla\chi \cdot \hat{t} ds = \oint_{\partial S} \nabla\chi \cdot dx = \oint_{\partial S} d\chi = 0. \quad (17.51)$$

Because Ψ has a connection to fluid transport, we sometimes refer to it as the *transport streamfunction*, just as for the streamfunction ψ in two-dimensional non-divergent flows (Section 17.4.3).

17.5.4 Scalar streamfunctions and transport

We can expose the two degrees of freedom of the vector streamfunction by writing it as the product

$$\Psi = \gamma \nabla\psi \quad (17.52)$$

so that the velocity is given by

$$\mathbf{v} = \nabla \wedge \Psi = \nabla\gamma \wedge \nabla\psi. \quad (17.53)$$

By construction the velocity satisfies

$$\mathbf{v} \cdot \nabla\gamma = \mathbf{v} \cdot \nabla\psi = 0, \quad (17.54)$$

so that the velocity is parallel to surfaces of constant γ and ψ . Correspondingly, the velocity streamlines are intersections of the γ and ψ isosurfaces, as depicted in Figure 17.3. We thus refer to γ and ψ as the two *scalar streamfunctions* for the three dimensional non-divergent flow. However, note that γ and ψ have different dimensions. By convention, we choose γ to have dimensions of length, so that it is not a traditional streamfunction, whereas ψ has the typical streamfunction dimensions of squared length per time.

As a check that the formalism is sensible, consider the special case of two-dimensional flow so that all streamlines are in the horizontal x - y plane. Taking $\gamma = z$ then renders

$$\Psi = z \nabla\psi \quad \text{and} \quad \mathbf{v} = \nabla \wedge \Psi = \hat{z} \wedge \nabla\psi, \quad (17.55)$$

which agrees with the scalar streamfunction in equation (17.11) for two dimensional non-divergent flow.

The volume transport through a surface defined by the two streamfunction isosurfaces takes the form

$$\mathcal{T}(S) = \int_S \mathbf{v} \cdot \hat{n} dS = \oint_{\partial S} \Psi \cdot \hat{t} ds = \oint_{\partial S} \gamma \nabla\psi \cdot \hat{t} ds = \oint_{\partial S} \gamma d\psi = - \oint_{\partial S} \psi d\gamma. \quad (17.56)$$

To reach the penultimate step we set

$$d\psi = \nabla\psi \cdot dx = \nabla\psi \cdot \hat{t} ds, \quad (17.57)$$

and for the final step we used the identity

$$\oint_{\partial S} \gamma d\psi = \oint_{\partial S} d(\gamma\psi) - \oint_{\partial S} \psi d\gamma = - \oint_{\partial S} \psi d\gamma, \quad (17.58)$$

which follows since $\oint_{\partial S} d(\gamma\psi) = 0$. The volume transport for the particular surface shown in Figure 17.3 is given by

$$\mathcal{T}(S) = \oint_{\partial S} \gamma d\psi = \gamma_1 (\psi_2 - \psi_1) + \gamma_2 (\psi_1 - \psi_2) = (\gamma_1 - \gamma_2) (\psi_2 - \psi_1). \quad (17.59)$$

Hence, the volume transport through a streamtube defined by isosurfaces of γ and ψ is given by the product of the difference between the isosurfaces.

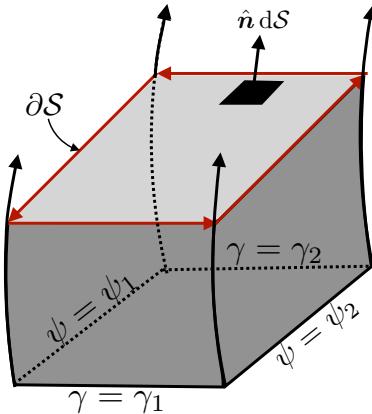


FIGURE 17.3: Isosurfaces of constant scalar streamfunctions, γ and ψ , for a three dimensional non-divergent flow. Streamlines are defined by the intersections of the γ and ψ isosurfaces, as shown by four streamlines along the corners of this particular volume. The volume transport of fluid through the surface, S , is determined by the line integral, $\oint_{\partial S} \gamma \, d\psi = -\oint_{\partial S} \psi \, d\gamma = (\gamma_1 - \gamma_2)(\psi_2 - \psi_1)$, around the boundary circuit.

17.5.5 Concerning a harmonic velocity potential

As for the two-dimensional case discussed in Section 17.4.6, we consider the relevance of an arbitrary harmonic velocity potential, χ , so that the velocity takes the form

$$\mathbf{v} = \nabla \wedge \boldsymbol{\Gamma} + \nabla \chi \quad \text{with} \quad \nabla^2 \chi = 0. \quad (17.60)$$

Since $\nabla \cdot \nabla \chi = 0$ we can write $\nabla \chi$ as the curl of another vector

$$\nabla \cdot \nabla \chi = 0 \implies \nabla \chi = \nabla \wedge \boldsymbol{\Lambda}, \quad (17.61)$$

in which case the velocity takes the form

$$\mathbf{v} = \nabla \wedge \boldsymbol{\Gamma} + \nabla \phi = \nabla \wedge (\boldsymbol{\Gamma} + \boldsymbol{\Lambda}) \equiv \nabla \wedge \boldsymbol{\Psi}. \quad (17.62)$$

Consequently, just as for the two-dimensional case, we are at liberty to work with the transport streamfunction $\boldsymbol{\Psi}$ if that suits our needs. Otherwise, we can work with the harmonic potential, χ , which is commonly used when the velocity is both non-divergent and irrotational.

17.6 Evolution of volume and area

In this section we develop kinematic equations for the evolution of volume and area within a non-divergent flow. We start by considering a material region and then study an arbitrary region.

17.6.1 Material volumes and areas

As shown by equation (17.2), the volume of a material parcel remains materially constant in a non-divergent flow. Correspondingly, a material fluid region maintains a constant volume

$$\frac{d}{dt} \int_{\mathcal{R}(\mathbf{v})} dV = \int_{\mathcal{R}(\mathbf{v})} \frac{D(\delta V)}{Dt} = \int_{\mathcal{R}(\mathbf{v})} (\nabla \cdot \mathbf{v}) dV = 0. \quad (17.63)$$

The appearance of a material time derivative on the inside of the integral arises since the integral is computed following material fluid particles whose trajectories define integral curves of the flow

(see Section 16.3.6). Likewise, following from the material area element equation (14.67), the area of a material region in a two-dimensional incompressible flow remains materially constant

$$\frac{d}{dt} \int_{\mathcal{S}(\mathbf{v})} d\mathcal{S} = \int_{\mathcal{S}(\mathbf{v})} \frac{D(\delta\mathcal{S})}{Dt} = \int_{\mathcal{S}(\mathbf{v})} (\nabla \cdot \mathbf{u}) d\mathcal{S} = 0. \quad (17.64)$$

This area preservation property is illustrated in Figure 17.4, in which a two-dimensional flow is seen to deform an initially square patch of fluid.



FIGURE 17.4: Illustrating the stretching and straining of an initially square patch of fluid in a two-dimensional non-divergent flow field. As discussed in Section 17.6, the area of a material region remains fixed in this flow field.

17.6.2 Arbitrary volume and area

We make use of the Leibniz-Reynolds transport theorem from Section 16.3.4 to develop the evolution equation for the volume of an arbitrary region. In particular equation (16.36) gives

$$\frac{d}{dt} \left[\int_{\mathcal{R}} dV \right] = \oint_{\partial\mathcal{R}} \mathbf{v}^{(b)} \cdot \hat{\mathbf{n}} d\mathcal{S}. \quad (17.65)$$

This result holds for both compressible and incompressible flows. But for incompressible flows we can go one step further by noting that

$$0 = \int_{\mathcal{R}} \nabla \cdot \mathbf{v} dV = \oint_{\partial\mathcal{R}} \mathbf{v} \cdot \hat{\mathbf{n}} d\mathcal{S}. \quad (17.66)$$

Importantly, this result holds only when integrating around the boundary of the closed volume, \mathcal{R} . It does not necessarily mean that $\mathbf{v} \cdot \hat{\mathbf{n}} = 0$ holds at every point along the boundary. Indeed, when the boundary is time dependent $\mathbf{v} \cdot \hat{\mathbf{n}} \neq 0$ generally holds everywhere along the boundary.

Making use of equation (17.66) allows us to write

$$\frac{d}{dt} \left[\int_{\mathcal{R}} dV \right] = \oint_{\partial\mathcal{R}} \mathbf{v}^{(b)} \cdot \hat{\mathbf{n}} d\mathcal{S} = - \oint_{\partial\mathcal{R}} (\mathbf{v} - \mathbf{v}^{(b)}) \cdot \hat{\mathbf{n}} d\mathcal{S} = - \int_{\mathcal{R}} \nabla \cdot (\mathbf{v} - \mathbf{v}^{(b)}) dV. \quad (17.67)$$

This result is identical to the mass budget equation (16.38) for the special case of a constant reference density appropriate for a Boussinesq fluid. The dia-surface transport, $(\mathbf{v} - \mathbf{v}^{(b)}) \cdot \hat{\mathbf{n}} d\mathcal{S}$, measures the volume per time crossing the boundary of the region, whether that region has a static or moving boundary. For example, if the boundary is the ocean free surface, then we can make use of the surface kinematic boundary condition (17.4).

17.7 Meridional-depth overturning circulation

Geophysical fluid flow is generally three-dimensional. However, it is sometimes useful to summarize aspects of that flow by integrating the mass transport over one of the directions. A common approach is to integrate over the zonal direction either between two solid-wall boundaries (as in an ocean basin) or over a periodic domain (as in the atmosphere or within the Southern Ocean).

Doing so leaves a two-dimensional transport in the (y, z) plane known as the meridional-depth overturning circulation

$$V^\rho = \int_{x_1}^{x_2} \rho v \, dx \quad \text{and} \quad W^\rho = \int_{x_1}^{x_2} \rho w \, dx. \quad (17.68)$$

We can go even further for incompressible flow, or for steady state compressible flow, in which case we introduce a streamfunction for the meridional-depth circulation.

Creating the overturning streamfunction

To see how to create an overturning streamfunction, consider the zonal integrated area transport for an incompressible flow³

$$V = \int_{x_1}^{x_2} v \, dx \quad \text{and} \quad W = \int_{x_1}^{x_2} w \, dx. \quad (17.69)$$

For a non-periodic basin domain (e.g., North Atlantic Ocean), let the zonal bounds, x_1 and x_2 , be well within rock (where the velocity vanishes) so that we encompass the full basin without having any meridional or depth dependence to the zonal bounds. Taking the meridional derivative of the meridional transport leads to

$$\frac{\partial V}{\partial y} = \frac{\partial}{\partial y} \left[\int_{x_1}^{x_2} v \, dx \right] \quad (17.70a)$$

$$= \int_{x_1}^{x_2} \frac{\partial v}{\partial y} \, dx \quad (17.70b)$$

$$= - \int_{x_1}^{x_2} \left[\frac{\partial u}{\partial x} + \frac{\partial w}{\partial z} \right] dx \quad (17.70c)$$

$$= - \int_{x_1}^{x_2} \frac{\partial w}{\partial z} \, dx \quad (17.70d)$$

$$= - \frac{\partial}{\partial z} \left[\int_{x_1}^{x_2} w \, dx \right] \quad (17.70e)$$

$$= - \frac{\partial W}{\partial z}. \quad (17.70f)$$

These results show that the two-dimensional zonally integrated transport is non-divergent

$$\frac{\partial V}{\partial y} + \frac{\partial W}{\partial z} = 0. \quad (17.71)$$

As a consequence of the non-divergence condition (17.71), we can introduce a meridional-depth streamfunction

$$\Psi(y, z, t) = - \int_{\eta_b^{\max}}^z V \, dz' = - \int_{\eta_b^{\max}}^z \left[\int_{x_1}^{x_2} v \, dx \right] dz', \quad (17.72)$$

to specify the zonally integrated flow. The lower limit of $z = \eta_b^{\max}$ is a constant that is deeper than the maximum depth in the full domain and with the convention that there is zero transport for any region below the fluid bottom. This convention follows that taken for the zonal bounds, x_1

³The case for a steady compressible flow follows analogously through use of $(\rho v, \rho w)$ rather than (v, w) .

and x_2 , which are constants that we well within the solid earth when the domain is not periodic. We verify that Ψ is a streamfunction by computing

$$\frac{\partial \Psi}{\partial z} = -V \quad (17.73a)$$

$$\frac{\partial \Psi}{\partial y} = -\frac{\partial}{\partial y} \left[\int_{\eta_b^{\max}}^z V dz' \right] = -\int_{\eta_b^{\max}}^z \frac{\partial V}{\partial y} dz' = \int_{\eta_b^{\max}}^z \frac{\partial W}{\partial z} dz' = W. \quad (17.73b)$$

To reach this result, we made use of the non-divergent condition (17.71), and set

$$W(z = \eta_b^{\max}) = 0. \quad (17.74)$$

Also, we are able to move the derivative across the lower limit of the integral since η_b^{\max} is a constant. An idealized version of the meridional-depth circulation is shown in Figure 17.5.

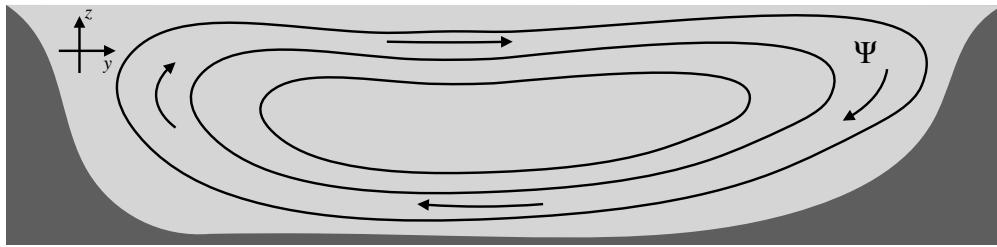


FIGURE 17.5: An idealized depiction of a steady meridional-depth overturning circulation. Shown here are streamlines (isolines of constant Ψ) for the zonally integrated flow between two solid boundaries or over a zonally periodic domain. The flow is assumed to be non-divergent, as per equation (17.71). In the upper reaches of the fluid, flow moves northward (positive y), with downward motion as it reaches the northern boundary, then southward motion at depth and eventual return towards the surface near the southern boundary.

The overturning streamfunction at the ocean surface

To evaluate the overturning streamfunction (17.72) at the ocean surface, $z = \eta(x, y, t)$, we follow a method similar to how we deal with the bottom. Namely, introduce a constant η^{\max} that is larger than any value of $\eta(x, y, t)$ and with the convention that the transport is zero in regions above the ocean surface. In this case the streamfunction computed across the full depth of the domain is given by minus the net meridional transport across the chosen latitude

$$\Psi(y, z = \eta^{\max}, t) = - \int_{\eta_b^{\max}}^{\eta^{\max}} V dz' = - \int_{\eta_b^{\max}}^{\eta^{\max}} \left[\int_{x_1}^{x_2} v dx \right] dz', \quad (17.75)$$

Volume conservation means that this transport vanishes in the steady state but it is generally nonzero due to transients.



17.8 Exercises

EXERCISE 17.1: STREAMLINES FOR CELLULAR FLOW

Sketch the velocity field for this streamfunction

$$\psi(x, y) = A \sin(k x) \sin(l y), \quad (17.76)$$

where (k, l) are the zonal and meridional wavenumbers, respectively. Hint: assume any convenient value for the wavenumbers k, l and the amplitude, A , but indicate what values were chosen. Furthermore, show vectors so that the sense of the flow is clear.

EXERCISE 17.2: ZERO NET AREA TRANSPORT THROUGH STATIC CLOSED CURVE

For a two-dimensional non-divergent flow, show that there is zero net transport of fluid crossing an arbitrary static and simply connected closed curve. Note that in two space dimensions, the transport of fluid across a line has dimensions $L^2 T^{-1}$, thus representing an area transport.

EXERCISE 17.3: ZERO NET VOLUME TRANSPORT THROUGH STATIC CLOSED SURFACE

For a three-dimensional non-divergent flow, show that there is zero net transport of fluid crossing an arbitrary static and simply connected closed surface within the fluid interior. Note that in three space dimensions, the transport of fluid across a surface has dimensions $L^3 T^{-1}$, thus representing a volume transport.

EXERCISE 17.4: NET FLUID TRANSPORT ACROSS AN ARBITRARY SURFACE

Consider flow in a container with static sides/bottom. Draw an arbitrary static surface, \mathcal{S} , within the fluid from one side of the container to the other as in Figure 17.6. Integrate the fluid transport over the surface, $\int_{\mathcal{S}} \mathbf{v} \cdot \hat{\mathbf{n}} d\mathcal{S}$.

- For an incompressible flow, show that the transport $\int_{\mathcal{S}} \mathbf{v} \cdot \hat{\mathbf{n}} d\mathcal{S}$ vanishes. That is, the net transport across the surface is zero.
- Specialize the above result to a horizontal surface so that we see there is zero integrated vertical transport across the surface, $\int_{\mathcal{S}} w dx dy = 0$. Discuss these results. Note: see Section 44.4.8 for the more general case of a non-static surface.
- Rework part (a) for the case of a compressible fluid. That is, derive an equation for the net fluid transport across the surface if the fluid is compressible.

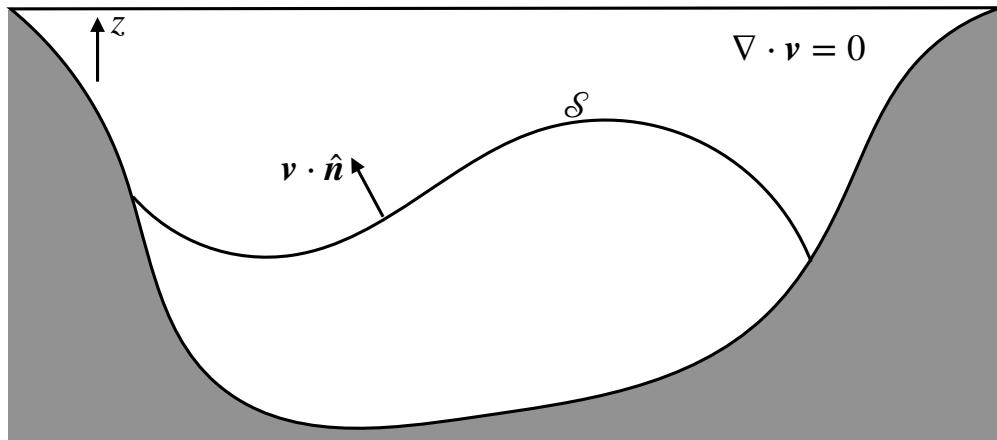


FIGURE 17.6: Schematic for exercise 17.4, whereby we show that the net flow vanishes across a static surface, \mathcal{S} , that extends from one boundary to the other within an incompressible flow.

EXERCISE 17.5: SOLID BODY ROTATION

Consider a velocity field corresponding to a time-independent solid-body rotation on a plane

$$\mathbf{u} = \Omega \hat{\mathbf{z}} \wedge \mathbf{x} = \Omega (-y \hat{\mathbf{x}} + x \hat{\mathbf{y}}), \quad (17.77)$$

where $\Omega > 0$ is a constant rotation rate.

- Compute the relative vorticity, $\boldsymbol{\omega} = \nabla \wedge \mathbf{u}$.
- Compute the streamfunction $\mathbf{u} = \hat{\mathbf{z}} \wedge \nabla \psi$. Draw streamfunction contours; i.e., lines of constant streamfunction. Put arrows to orient the flow along the streamlines.
- Describe the geometry of material lines. Hint: since the velocity field is time-independent, material parcel trajectories are coincident with streamlines.

EXERCISE 17.6: ALTERNATIVE FORM OF MERIDIONAL-DEPTH STREAMFUNCTION

In equation (17.72), we introduced the meridional-depth overturning streamfunction

$$\Psi(y, z, t) = - \int_{-H_{\max}}^z V \, dz'. \quad (17.78)$$

Show that an alternative streamfunction is given by

$$\Gamma(y, z, t) = \int_{y_s}^y W \, dy', \quad (17.79)$$

where y_s is a constant latitude just south of the southern-most latitude in the domain.

EXERCISE 17.7: VOLUME TRANSPORT THROUGH STREAMTUBE ENDS

Recall our discussion of streamtubes in Section 13.8.2 (see in particular Figure 13.5). For a steady non-divergent three-dimensional flow, show that the volume transport (volume per time) through the two streamtube ends balances

$$\int_{\mathcal{S}_1} \mathbf{v} \cdot \hat{\mathbf{n}}_1 \, d\mathcal{S} + \int_{\mathcal{S}_2} \mathbf{v} \cdot \hat{\mathbf{n}}_2 \, d\mathcal{S} = 0, \quad (17.80)$$

where $\hat{\mathbf{n}}_1$ and $\hat{\mathbf{n}}_2$ are the outward normals at the two end caps \mathcal{S}_1 and \mathcal{S}_2 . Since the end caps have oppositely directed outward normals, equation (17.80) says that the volume transport entering one streamtube end equals to that leaving the other end. Furthermore, the area of the streamtube is inversely proportional to the local normal velocity, so that flow speeds up when moving through a narrower region of the tube.

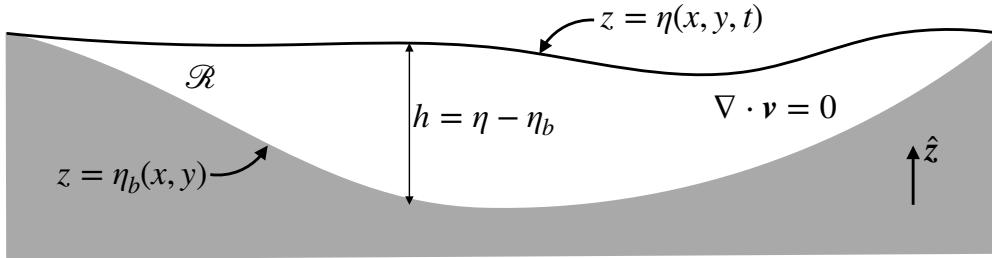


FIGURE 17.7: Schematic for exercise 17.8 with $z = \eta(x, y, t)$ the free surface at the top of the fluid. This exercise shows that the area integrated time tendency for the free surface vanishes in the absence of mass transport across the free surface.

EXERCISE 17.8: AREA AVERAGE OF FREE SURFACE TIME TENDENCY

Consider an incompressible ocean flow bounded by a free upper surface and a solid bottom. Let $z = \eta_b(x, y)$ be the vertical position of the static bottom, and $z = \eta(x, y, t)$ be the position of

the transient free surface, so that the thickness of the layer is $h = -\eta_b + \eta$ (see Figure 17.7). The horizontal extent of the layer is a function of time, and is defined by a vanishing thickness $h = -\eta_b + \eta = 0$ (e.g., ocean water reaching the shoreline). Assume no material crosses either the surface or bottom boundaries, so that both boundaries are material surfaces. Show that the free surface has a time derivative, $\partial\eta/\partial t$, whose area average vanishes. Discuss this result.

EXERCISE 17.9: VOLUME INTEGRAL OF THE NON-DIVERGENT VELOCITY FIELD

Consider an incompressible ocean flow bounded by a free upper surface and a solid bottom over a domain \mathcal{R} . Let $z = \eta_b(x, y)$ be the vertical position of the static bottom, and $z = \eta(x, y, t)$ be the position of the transient free surface as in Figure 17.7. Prove that the domain integral of the velocity is given by

$$\int_{\mathcal{R}} \mathbf{v} \, dV = \int_{z=\eta} \mathbf{x} (\mathbf{v} \cdot \hat{\mathbf{n}}) \, d\mathcal{S} = \int_{z=\eta} \mathbf{x} (\partial_t \eta - Q_m / \rho_0) \, dx \, dy. \quad (17.81)$$

Hint: make use of the results from Section 2.7.6 and then use the kinematic boundary conditions from Section 17.2.



Part IV

Dynamics of geophysical fluid flows

Dynamics is the area of mechanics that examines the causes of motion. For a Newtonian system, such as a geophysical fluid, the cause of motion is understood when we understand the forces. In this part of the book, we encounter a suite of theoretical concepts that form the foundations of geophysical fluid mechanics. Our presentation moves largely from the general to the specific, with each chapter written in a manner that allows it to be picked up without relying too much on other chapters. The “general to specific” route also allows us to establish general principles based on foundational principles and concepts and to then see how they manifest in specific contexts.

Forces of concern in geophysical fluid mechanics include the body force acting on a fluid element from the earth’s gravity field along with the contact forces from pressure and friction that act between adjacent fluid elements. Additionally, by choosing to work in a non-inertial rotating terrestrial reference frame, we encounter body forces from the Coriolis and planetary centrifugal accelerations, just as encountered for geophysical particle mechanics in Part II. Each of these forces play important roles in determining the diversity of geophysical fluid motion, and their analysis leads to dynamical insights into the nature and causes of fluid motion. We observe that forces in fluids are commonly inferred from kinematic properties of the motion, thus making use of the fluid kinematics from Part III.

SUMMARY OF THE DYNAMICS CHAPTERS

We start the development of dynamics by formulating the equations of motion (linear momentum and angular momentum) in Chapter 18. In subsequent chapters we study the forces appearing in these equations, including friction (Chapter 19), pressure (Chapter 20), and buoyancy (Chapter 25). Buoyancy is the vertical pressure force, arising from density inhomogeneities, that are not balanced by gravity. As such, our study of buoyancy focuses on vertical forces, which contrasts to our study of pressure form stresses in Chapter 20, which focus on horizontal forces.

When studying buoyancy in Chapter 25, we make use of an equation of state that provides the mass density of a fluid element as a function of thermodynamical properties such as temperature, pressure, and matter concentration. This role for thermodynamics prompts the development in Chapter 21 of selected topics in equilibrium thermodynamics. In Chapter 22 we extend these equilibrium thermodynamic relations to linear irreversible thermodynamics relevant for a moving geophysical fluid where we assume that each fluid element is in a local thermodynamic equilibrium. This extension to moving fluids allows us to study how mechanical energy is exchanged with internal energy in the presence of work done by pressure and heat generated by friction.

Chapters 23 and 24 introduce a variety of approximate equations that allow us to focus on selected dynamical regimes by filtering away uninteresting phenomena. It is here that we encounter the hydrostatic approximation and the Boussinesq approximation, both of which are commonly used for large-scale models of the ocean and atmosphere. Additionally, we simplify the spherical geometry by making the tangent plane approximation, with the f -plane and β -plane commonly used in idealized theoretical studies. Approximate balances are further examined in Chapter 26, where we study the mechanics of a rapidly rotating fluid. We here encounter the geostrophic balance, which is a diagnostic balance appropriate for describing large-scale geophysical flows in which the horizontal pressure acceleration is balanced by the Coriolis acceleration. Geostrophic balance is one of the variety of balances considered in Chapter 27, which introduces balanced flow regimes pertaining to horizontal motions. We close this part of the book in Chapter 28, which examines the physics of an Ekman boundary layer in which the Coriolis acceleration balances vertical friction.

Throughout the previous chapters we assumed a constant effective gravitational acceleration. However, much of the observed motion of the ocean arises from the spatial-temporal variations of the gravity field that give rise to tides, thus motivating our brief look at the subject in Chapter 29.

Finally, in Chapter 30 we briefly visit the topic of gravity waves on the surface of a homogenous and non-hydrostatic fluid. Such waves have a vertical extent that decays exponentially, thus leading to the transport of particles through Stokes drift. Stokes drift appears in many guises in fluid mechanics, with the presentation in this chapter rather standard but helping to support the later view in Chapter 49 from the basis of the generalized Lagrangian mean.

Momentum dynamics

We here formulate the fluid mechanical equations for linear and axial angular momentum relevant to geophysical fluid motions. These equations of geophysical fluid dynamics (GFD) are based on Newton's laws of motion applied to a stratified fluid continuum moving on a rotating spherical planet where the rotation rate is constant in time. Relative to the point particle, the new dynamical feature afforded to the continuum concerns contact forces between fluid elements, which lead to pressure and frictional forces from mechanical interactions.

READER'S GUIDE TO THIS CHAPTER

We make liberal use of results from point particle mechanics studied in Part II as well as fluid kinematics from Part III. Much of the material in this chapter is introductory, with further details and elaboration offered in later chapters.

18.1	Linear momentum equation	354
18.1.1	Body forces	354
18.1.2	Contact forces	354
18.1.3	Equation of motion	356
18.1.4	Euler equation for perfect fluid motion	357
18.1.5	Further study	357
18.2	Special forms for the momentum equation	357
18.2.1	Spherical coordinates	358
18.2.2	Vector-invariant velocity equation	359
18.2.3	Dynamic pressure and the Magnus acceleration	359
18.3	Exact hydrostatic balance	359
18.3.1	Properties of exact hydrostatic balance	360
18.3.2	Comparison to approximate hydrostatic balance	361
18.4	Axial angular momentum	361
18.4.1	Axial angular momentum conserving motion of a ring of air	362
18.4.2	Sketching the atmospheric angular momentum budget	363
18.4.3	Further study	364
18.5	Acoustic waves	364
18.5.1	Rudiments of acoustic waves	365
18.5.2	Further study	366
18.6	Exercises	366

18.1 Linear momentum equation

We here summarize elements of classical continuum mechanics and in turn apply Newton's second law to derive the linear momentum budget for a fluid continuum. We present the momentum budget over both a finite volume region of the fluid (weak formulation) and for an infinitesimal fluid element (strong formulation).

18.1.1 Body forces

Forces acting on an arbitrary volume, \mathcal{R} , of a continuous media are of two general kinds. The first kind involves *external* or *body* forces, such as gravitation (including astronomical tidal forces), Coriolis, planetary centrifugal, and electromagnetic forces. These forces act throughout the extent of the media. Consequently, the total body force acting on a volume of fluid is the integral of the body force per unit mass, \mathbf{f}_{body} , multiplied by the mass of the media

$$\mathbf{F}_{\text{body}} = \int_{\mathcal{R}} \mathbf{f}_{\text{body}} \rho dV. \quad (18.1)$$

For example, the effective gravitational force (combination of central gravity plus planetary centrifugal) acting on a volume of fluid is given by

$$\mathbf{F}_{\text{effective gravity}} = \int_{\mathcal{R}} \mathbf{g} \rho dV, \quad (18.2)$$

where $\mathbf{g} = -\nabla\Phi$ is the effective acceleration of gravity with Φ the geopotential (Section 10.12.4). Likewise, the Coriolis force is given by

$$\mathbf{F}_{\text{Coriolis}} = -2 \int_{\mathcal{R}} (\boldsymbol{\Omega} \wedge \mathbf{v}) \rho dV. \quad (18.3)$$

These body forces have the same appearance as for the point particle in Chapter 10, with the only difference being the material is now a continuous media rather than a point mass.

18.1.2 Contact forces

The second kind of forces are *internal* or *contact* forces, such as pressure forces and frictional forces. These forces act on a region of a continuous media through the area integrated stresses acting on the boundary enclosing the region. Mathematically, we compute the contact force exerted on the region through its boundaries through the area integral of the *stress tensor* projected onto the normal direction

$$\mathbf{F}_{\text{contact}} = \oint_{\partial\mathcal{R}} \mathbb{T} \cdot \hat{\mathbf{n}} d\mathcal{S}, \quad (18.4)$$

where $\hat{\mathbf{n}}$ is the outward normal direction orienting the domain boundary with $d\mathcal{S}$ the associated area element, and \mathbb{T} is the second order stress tensor. We have more to say about the stress tensor in the following as well as in Chapter 19.

Contact forces affect continuous media and they do so through the nonzero spatial extent of elements within the media, with this spatial extent allowing for interactions between adjacent fluid elements. Point particles (Part II) do not experience contact forces since point particles have no spatial extent. Hence, contact forces represent a fundamentally new feature, conceptually and operationally, to the fluid dynamical equations relative to the equations of point particles.

Stresses from friction and pressure

As studied in detail in Chapter 19, there are two types of stress that concern us: diagonal stresses associated with reversible momentum exchange through pressure, and stresses associated with the irreversible exchange of momentum through friction. Hence, it is convenient to decompose the stress tensor components according to

$$\mathbb{T}^{ab} = \tau^{ab} - p g^{ab}. \quad (18.5)$$

In this equation, p is the pressure, which is a force per unit area acting in a compressive manner on the area of a surface. The second order tensor, g^{ab} , is a chosen coordinate representation of the inverse metric tensor and it equals to the Kronecker or unit tensor when choosing Cartesian coordinates in Euclidean space (Section 7.1). The frictional stress tensor is written τ^{ab} . It is also known as the *deviatoric* stress tensor as it represents deviations from the static case when stress is due solely to pressure. The friction stress tensor generally has zero trace, with pressure comprising the trace portion of the full stress tensor.

Substitution of the stress tensor (18.5) into the contact force expression (18.4) leads to

$$\mathbf{F}_{\text{contact}} = \oint_{\partial\mathcal{R}} (\tau \cdot \hat{\mathbf{n}} - p \hat{\mathbf{n}}) dS, \quad (18.6)$$

where the integral is taken over the bounding surface of the domain whose outward normal is $\hat{\mathbf{n}}$. Given this expression for contact forces acting on the boundary of a fluid domain, it is seen that positive pressure ($p > 0$) acts in the direction opposite to the surface's outward normal so that pressure always acts in a compressive manner. Deviatoric stresses create more general forces on the bounding surface, which can have compressive, expansive, shearing, and/or rotational characteristics.

Exchange of momentum between fluid elements

We mathematically represent the exchange of momentum between fluid elements via a symmetric stress tensor, with symmetry implied by statements about angular momentum conservation (detailed in Section 19.3). The divergence of the stress tensor then leads to a force acting on the fluid element. The forces arising from molecular viscosity provide an irreversible exchange of momentum that reduce the kinetic energy of fluid elements (Section 22.3.2). This process is dissipative and thus referred to as friction. Furthermore, when averaging over turbulent realizations of a fluid, the impacts on the mean flow are generally far larger than those associated with molecular viscosity, with these exchanges commonly parameterized via a symmetric stress tensor.

A gauge symmetry of pressure force

The contribution from pressure in the contact force (18.6) remains invariant if pressure is shifted by an arbitrary function of time

$$p(\mathbf{x}, t) \rightarrow p(\mathbf{x}, t) + F(t). \quad (18.7)$$

We see this invariance by noting that

$$\oint_{\partial\mathcal{R}} F(t) \hat{\mathbf{n}} dS = F(t) \oint_{\partial\mathcal{R}} \hat{\mathbf{n}} dS = 0, \quad (18.8)$$

where the final equality follows from a corollary of the divergence theorem as discussed in Section 2.7.3. Briefly, through the divergence theorem in Section 2.7.2 we know that

$$\oint_{\partial\mathcal{R}} p \hat{\mathbf{n}} d\mathcal{S} = \int_{\mathcal{R}} \nabla p dV, \quad (18.9)$$

so that if pressure is shifted by a spatial constant then the pressure gradient body force remains unchanged, as will the pressure contact force.

We refer to this invariance of the pressure force as a *gauge symmetry*. It means that motion of the fluid remains unchanged if pressure is modified by a spatial constant that can generally be a function of time.

18.1.3 Equation of motion

The linear momentum of a fluid region is given by

$$\mathbf{P} = \int_{\mathcal{R}} \mathbf{v} \rho dV. \quad (18.10)$$

Applying Newton's law of motion to the continuum leads to the finite volume equation of motion

$$\frac{d}{dt} \int_{\mathcal{R}} \mathbf{v} \rho dV = \int_{\mathcal{R}} \rho \mathbf{f}_{\text{body}} dV + \oint_{\partial\mathcal{R}} \mathbb{T} \cdot \hat{\mathbf{n}} d\mathcal{S}. \quad (18.11)$$

The time derivative can be material, as for a constant mass fluid region moving with the barycentric velocity, Eulerian, as for a fixed region in space (see Section 16.3), or otherwise. Applying the divergence theorem (Section 2.7.2) to the area integral yields

$$\frac{d}{dt} \int_{\mathcal{R}} \rho \mathbf{v} dV = \int_{\mathcal{R}} (\rho \mathbf{f}_{\text{body}} + \nabla \cdot \mathbb{T}) dV, \quad (18.12)$$

where we brought the contact forces into the volume integral through exposing the divergence of the stress tensor.

General form of the equation of motion for a fluid element

Since the volume under consideration is arbitrary, the integral relation (18.12) is satisfied for an arbitrary region. We apply the result to an infinitesimal fluid element moving with the flow so that

$$\frac{D(\rho \mathbf{v} \delta V)}{Dt} = \delta V (\rho \mathbf{f}_{\text{body}} + \nabla \cdot \mathbb{T}). \quad (18.13)$$

Assuming the mass for the fluid element is constant then leads to the equation of motion

$$\rho \frac{D\mathbf{v}}{Dt} = \rho \mathbf{f}_{\text{body}} + \nabla \cdot \mathbb{T}. \quad (18.14)$$

This equation is a continuum expression of Newton's equation of motion, and it is sometimes referred to as *Cauchy's equation of motion* in honor of the work Cauchy did in its formulation.

Momentum equation for a rotating fluid in a gravitational field

We now specialize the momentum equation (18.14) to suit the needs of geophysical fluid mechanics. We first write the stress tensor in terms of the deviatoric component from friction and a diagonal component from pressure (equation (18.6))

$$\rho \frac{D\mathbf{v}}{Dt} = \rho \mathbf{f}_{\text{body}} - \nabla p + \nabla \cdot \boldsymbol{\tau}. \quad (18.15)$$

Next, move to a rotating terrestrial reference frame and thus expose the Coriolis acceleration and the effective gravitational force (Section 10.13)

$$\rho \frac{D\mathbf{v}}{Dt} + 2\rho \boldsymbol{\Omega} \wedge \mathbf{v} = -\rho \nabla \Phi - \nabla p + \nabla \cdot \boldsymbol{\tau}. \quad (18.16)$$

This form of the equation of motion arises from extracting the solid-body motion of the basis vectors to define the Coriolis acceleration (see Section 10.11). Any remaining changes to the basis vectors arise from motion of the fluid relative to the solid-body rotating reference frame, and thus appear when expanding the material time derivative. The form (18.16) for the equation of motion offers a suitable starting point for studies of geophysical fluid dynamics. It sometimes goes by the name of *Navier-Stokes* equation. However, that name is more commonly applied to the non-rotating case with a specific form for the friction operator (see Section 19.7.6). We thus refer to equation (18.16) as Newton's law of motion for a rotating fluid.¹

18.1.4 Euler equation for perfect fluid motion

This inviscid form of the momentum equation (18.15) is known as the *Euler equation* of perfect fluid mechanics

$$\rho \frac{D\mathbf{v}}{Dt} = \rho \mathbf{f}_{\text{body}} - \nabla p, \quad (18.17)$$

where the body force is conservative. That is, the Euler equation is concerned just with fluid motion in the absence of dissipative processes. The inviscid form of the momentum equation (18.16) leads to the Euler equation in the presence of rotation and gravitation

$$\rho \frac{D\mathbf{v}}{Dt} + 2\rho \boldsymbol{\Omega} \wedge \mathbf{v} = -\rho \nabla \Phi - \nabla p. \quad (18.18)$$

We have many occasions in this book to ignore dissipation, in which case we work with a particular form of the Euler equation of motion. We further comment on the Euler equation in Section 19.7.6.

18.1.5 Further study

Chapter 5 of [Aris \(1962\)](#) offers an insightful discussion of continuum mechanics as applied to a fluid. Section 2.2 [Vallis \(2017\)](#) provides a thorough derivation of the dynamical equations of motion for the atmosphere and ocean. We offer further discussion of the mathematics and physics of stress in fluids in Chapters 19 and 20.

18.2 Special forms for the momentum equation

We here display some special forms of the momentum equation that will be of use in our studies.

¹The Navier-Stokes equations were first derived by Claude-Louis Navier in 1822 and later independently derived by George Stokes in 1845.

18.2.1 Spherical coordinates

Geophysical fluids move on a rotating planet with the planet commonly assumed to have an oblated spherical geometry, though with the equations approximated by their spherical form using the geopotential vertical coordinate. To display the equations of motion, we make use of the acceleration as derived in Section 10.13.3 for the point particle, using the geopotential coordinate to measure radial distances from the center of the sphere (equation (22.144)), as well as the longitude and latitude angular coordinates defined by Figure 8.1. The point particle time derivative, which is computed following the particle, translates into a material time derivative for fluid elements. We are thus led to the spherical equations of motion in their full glory

$$\frac{Du}{Dt} + \frac{u(w - v \tan \phi)}{r} + 2\Omega(w \cos \phi - v \sin \phi) = -\frac{1}{\rho r \cos \phi} \frac{\partial p}{\partial \lambda} + F^\lambda \quad (18.19)$$

$$\frac{Dv}{Dt} + \frac{v w + u^2 \tan \phi}{r} + 2\Omega u \sin \phi = -\frac{1}{\rho r} \frac{\partial p}{\partial \phi} + F^\phi \quad (18.20)$$

$$\frac{Dw}{Dt} - \frac{u^2 + v^2}{r} - 2\Omega u \cos \phi = -g - \frac{1}{\rho} \frac{\partial p}{\partial r} + F^r, \quad (18.21)$$

where we introduced the spherical components to the friction acceleration

$$\mathbf{F} = F^\lambda \hat{\lambda} + F^\phi \hat{\phi} + F^r \hat{r}, \quad (18.22)$$

which is determined by the divergence of the frictional stress tensor. We also note the spherical coordinate form for the gradient operator (Section 8.2.8)

$$\nabla = \frac{\hat{\lambda}}{r \cos \phi} \frac{\partial}{\partial \lambda} + \frac{\hat{\phi}}{r} \frac{\partial}{\partial \phi} + \hat{r} \frac{\partial}{\partial r}, \quad (18.23)$$

as well as the material time derivative operator

$$\frac{D}{Dt} = \frac{\partial}{\partial t} + \mathbf{v} \cdot \nabla = \frac{\partial}{\partial t} + \frac{u}{r \cos \phi} \frac{\partial}{\partial \lambda} + \frac{v}{r} \frac{\partial}{\partial \phi} + w \frac{\partial}{\partial r}. \quad (18.24)$$

We can write the spherical momentum equations in a bit more compact form by introducing the spherical coordinate velocity field (see equation (10.67))

$$\mathbf{v} = \mathbf{u} + \hat{r} w = u \hat{\lambda} + v \hat{\phi} + w \hat{r} \quad (18.25)$$

and the corresponding spherical coordinate acceleration

$$\mathbf{A}_{\text{sphere}} = \frac{Du}{Dt} \hat{\lambda} + \frac{Dv}{Dt} \hat{\phi} + \frac{Dw}{Dt} \hat{r}. \quad (18.26)$$

We also introduce the expression (10.104c) for the metric acceleration to render

$$\rho \frac{D\mathbf{v}}{Dt} + 2\rho \boldsymbol{\Omega} \wedge \mathbf{v} = -\rho \nabla \Phi - \nabla p + \rho \mathbf{F}, \quad (18.27)$$

where we have the acceleration relative to the rotating frame

$$\frac{D\mathbf{v}}{Dt} = \mathbf{A}_{\text{sphere}} + \frac{1}{r} [u \tan \phi (\hat{r} \wedge \mathbf{v}) + w \mathbf{u} - \hat{r} \mathbf{u} \cdot \mathbf{u}]. \quad (18.28)$$

For some purposes it is convenient to combine one piece of the metric acceleration to the Coriolis acceleration to yield

$$\mathbf{A}_{\text{sphere}} + \frac{1}{r} [w \mathbf{u} - \hat{r} \mathbf{u} \cdot \mathbf{u}] + \left[2\boldsymbol{\Omega} + \frac{u \tan \phi \hat{r}}{r} \right] \wedge \mathbf{v} = - \left[\frac{\rho \nabla \Phi + \nabla p}{\rho} \right] + \mathbf{F}. \quad (18.29)$$

18.2.2 Vector-invariant velocity equation

The metric terms appearing in the momentum equation (18.29) are those terms proportional to r^{-1} that arise from spatial dependence of the spherical unit vectors. An alternative formulation removes these terms in favor of the vorticity and kinetic energy. For that purpose we make use of the identity (equation (2.39)) for the nonlinear self-advection term

$$(\mathbf{v} \cdot \nabla) \mathbf{v} = \boldsymbol{\omega} \wedge \mathbf{v} + \nabla(\mathbf{v} \cdot \mathbf{v})/2, \quad (18.30)$$

where $\boldsymbol{\omega} = \nabla \wedge \mathbf{v}$ is the vorticity from Part VI of this book. We derive the corresponding *vector-invariant* form of the velocity equation using Cartesian coordinates and then invoke general coordinate invariance (Section 6.1) to extend the result to arbitrary coordinates.² Making use of equation (18.30) thus leads to the material acceleration

$$\frac{D\mathbf{v}}{Dt} = \frac{\partial \mathbf{v}}{\partial t} + \boldsymbol{\omega} \wedge \mathbf{v} + \nabla(\mathbf{v} \cdot \mathbf{v})/2 \quad (18.31)$$

so that the momentum equation (18.16) becomes the vector-invariant velocity equation

$$\frac{\partial \mathbf{v}}{\partial t} + (2\boldsymbol{\Omega} + \boldsymbol{\omega}) \wedge \mathbf{v} = -\nabla(\Phi + \mathbf{v} \cdot \mathbf{v}/2) + (1/\rho)(-\nabla p + \nabla \cdot \mathbf{t}). \quad (18.32)$$

18.2.3 Dynamic pressure and the Magnus acceleration

The *velocity equation* (18.32) is mathematically equivalent to the momentum equation (18.16). Even so, it provides a more convenient means to derive Bernoulli's theorem in Section 22.6.2 and the vorticity equation in Chapter 36. Furthermore, it highlights certain physical processes affecting accelerations that are not obviously seen from the momentum equation. For example, the velocity equation reveals that kinetic energy gradients contribute a *dynamic pressure gradient* that accelerates the fluid down the kinetic energy gradient, from regions of high kinetic energy to low.

Additionally, the acceleration $-\boldsymbol{\omega} \wedge \mathbf{v}$ appearing in the velocity equation (18.32) couples vorticity and velocity. This acceleration is known as the *Magnus effect* or *Magnus acceleration*. Since it acts only when there is both motion and vorticity, it is sometimes referred to as a *vortex force*, with the force increasing in magnitude where the velocity and/or vorticity are larger in magnitude. As discussed in Chapter 33, vorticity is a measure of the spin of a fluid element, so that the Magnus acceleration deflects a spinning fluid element in a direction perpendicular to its trajectory just like the Coriolis acceleration. We illustrate a particular case in Figure 18.1 for a zonal flow with a negative meridional shear that creates a positive vertical relative vorticity. Besides causing a moving and spinning fluid element to deflect, the Magnus acceleration provides the mechanism whereby a solid spinning body immersed in a moving fluid is deflected, such as commonly experienced by spinning balls used for baseball, tennis, and cricket. Notably, as per Newton's third law, the spinning ball is deflected in a direction opposite to that of the spinning fluid element.

18.3 Exact hydrostatic balance

We are mostly interested in moving fluids within this book. Even so, it is useful to expose the signature of a static fluid supporting the trivial solution, $\mathbf{v} = 0$. The equation of motion (18.16) has an exact static solution so long as the pressure gradient force balances the effective gravitational force

$$\nabla p = -\rho \nabla \Phi, \quad (18.33)$$

²See Section 4.4.4 of [Griffies \(2004\)](#) for a detailed derivation using general coordinates.

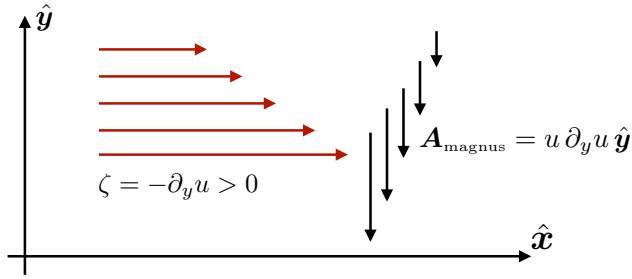


FIGURE 18.1: The Magnus acceleration, $\mathbf{A}_{\text{magnus}} = -\boldsymbol{\omega} \wedge \mathbf{v}$, acting on a fluid element with nonzero vorticity. In this case the fluid element has a zonal velocity with a linear meridional shear, $\mathbf{v} = u(y) \hat{\mathbf{x}}$, so that the relative vorticity is $\boldsymbol{\omega} = \zeta \hat{\mathbf{z}} = -(\partial_y u) \hat{\mathbf{z}}$ and the Magnus acceleration is $\mathbf{A}_{\text{magnus}} = u \partial_y u \hat{\mathbf{y}}$. When the fluid has a negative meridional shear, $\partial_y u < 0$, its relative vorticity is positive (cyclonic), $\zeta = -\partial_y u > 0$, so that the Magnus acceleration deflects the fluid element to the right of the motion, just as for the Coriolis acceleration in the northern hemisphere.

and where the frictional stress tensor has zero divergence. Equation (18.33) constitutes the *exact hydrostatic balance*. As justified in Section 23.3, the hydrostatic balance is a very good approximation for the vertical momentum equation in large-scale geophysical fluids even when those fluids are moving. We will thus commonly make the *hydrostatic approximation* for moving fluids. For the current considerations, we are interested in a static fluid, in which case the hydrostatic balance (18.33) is an exact solution to the equation of motion.

18.3.1 Properties of exact hydrostatic balance

We make the following observations of the exact hydrostatic balance.

- Since ∇p is directly proportional to $\nabla \Phi$, surfaces of constant pressure in a static fluid correspond to surfaces of constant geopotential.
- Since the curl of the pressure gradient vanishes, a static fluid maintains its density gradients parallel to geopotential gradients

$$\nabla \rho \wedge \nabla \Phi = 0, \quad (18.34)$$

which in turn means that density surfaces are parallel to geopotentials so that

$$\rho = \rho(\Phi) \quad \text{static fluid}. \quad (18.35)$$

For the geopotential $\Phi = g z$, a static fluid is realized if the density has no horizontal (on the sphere) gradients so that

$$\rho = \rho(z) \quad \text{static fluid with } \Phi = g z. \quad (18.36)$$

If the density gradient has any component perpendicular to $\nabla \Phi$, then pressure forces will affect fluid flow thus implying that the fluid is not in exact hydrostatic balance.

- Projecting both sides of equation (18.33) onto an infinitesimal space increment, $d\mathbf{x}$, renders

$$d\mathbf{x} \cdot \nabla p = -\rho d\mathbf{x} \cdot \nabla \Phi \implies \frac{dp}{d\Phi} = -\rho. \quad (18.37)$$

Hence, the difference in hydrostatic pressure between any two geopotentials is given by the integral

$$p(\Phi_2) - p(\Phi_1) = - \int_{\Phi_1}^{\Phi_2} \rho(\Phi) d\Phi. \quad (18.38)$$

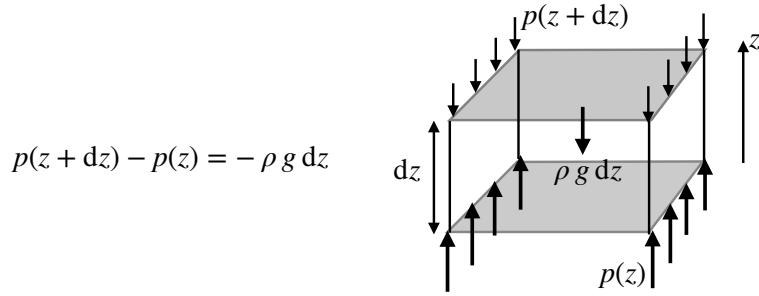


FIGURE 18.2: Illustrating the forces acting in a hydrostatically balanced fluid layer placed in a geopotential field $\Phi = g z$ with g constant. The layer has an infinitesimal thickness $dz > 0$, density ρ , and horizontal cross-sectional area dA . The pressure force acting on the top and bottom of the layer are compressive. Hence, the pressure force at the top of the layer acts downward, $\mathbf{F}^{\text{press}}(z+dz) = -\hat{z} p(z+dz) dA$, whereas the pressure force at the bottom of the layer acts upward, $\mathbf{F}^{\text{press}}(z) = +\hat{z} p(z) dA$. In a hydrostatically balanced fluid, the difference in pressure across the layer is exactly balanced by the weight per area of fluid within the layer. Consequently, $p(z+dz) - p(z) = -g \rho(z) dz$, so that pressure at the top of the layer is less than that at the bottom.

If $\Phi = g z$ then we recover

$$p(z_2) - p(z_1) = -g \int_{z_1}^{z_2} \rho(z) dz, \quad (18.39)$$

so that the difference in hydrostatic pressure between two geopotentials is given by the weight per horizontal area of fluid between the two geopotentials. This relation is illustrated for an infinitesimally thin layer in Figure 18.2.

18.3.2 Comparison to approximate hydrostatic balance

A static fluid in a gravitational field exhibits hydrostatic balance whereby pressure at a point is a function solely of the geopotential, in which case $p = p(z)$ when $\Phi = g z$. Correspondingly, $dp/dz = -\rho z$, which means that we determine hydrostatic pressure at a point by computing the weight per horizontal area of fluid above that point. Likewise, density is just a function of geopotential since $\nabla \rho \wedge \nabla \Phi = 0$.

For an approximate hydrostatic fluid, pressure is a function of space and time, $p = p(\mathbf{x}, t)$, as is density, $\rho = \rho(\mathbf{x}, t)$. However, an approximate hydrostatic fluid satisfies $\partial p / \partial z = -\rho g$. Hence, column by column, the pressure at a point in an approximate hydrostatic fluid is determined by the weight per horizontal area of fluid above that point. This key property is thus shared between fluids in exact and approximate hydrostatic balance. We have more to say regarding approximate hydrostatic balance in Section 23.3.

18.4 Axial angular momentum

Following our discussion of a point particle in Section 11.7, the axial angular momentum of a fluid element is given by

$$L^z = (\rho \delta V) r_\perp (u + r_\perp \Omega) \equiv (\rho \delta V) l^z \quad (18.40)$$

where

$$l^z = r_\perp (u + r_\perp \Omega) \quad (18.41)$$

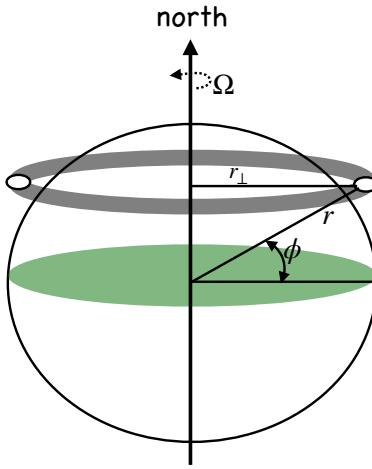


FIGURE 18.3: A ring of inviscid air circulating around a constant latitude circle over a smooth planet. This ring conserves is axial angular momentum. Consequently, axial angular momentum conserving motion of the ring induces a zonal acceleration if the ring alters its distance from the rotation axis, r_{\perp} , by moving meridionally or vertically.

is the axial angular momentum per unit mass, and the distance to the polar rotation axis,

$$r_{\perp} = r \cos \phi \quad (18.42)$$

is the moment-arm for determining the torques acting on a fluid element. Making use of the zonal momentum equation (18.19) we find the material time change

$$\frac{Dl^z}{Dt} = (u + 2\Omega r_{\perp}) \frac{Dr_{\perp}}{Dt} + r_{\perp} \frac{Du}{Dt} \quad (18.43a)$$

$$= (u + 2\Omega r_{\perp}) \frac{Dr_{\perp}}{Dt} + (u + 2\Omega r_{\perp})(v \sin \phi - w \cos \phi) - \frac{1}{\rho} \frac{\partial p}{\partial \lambda} \quad (18.43b)$$

$$= (u + 2\Omega r_{\perp}) \left[\frac{Dr_{\perp}}{Dt} + v \sin \phi - w \cos \phi \right] - \frac{1}{\rho} \frac{\partial p}{\partial \lambda} \quad (18.43c)$$

$$= -\frac{1}{\rho} \frac{\partial p}{\partial \lambda}, \quad (18.43d)$$

so that

$$\rho \frac{Dl^z}{Dt} = -\frac{\partial p}{\partial \lambda} \implies \frac{\partial(\rho l^z)}{\partial t} + \nabla \cdot (\rho l^z \mathbf{v}) = -\frac{\partial p}{\partial \lambda}. \quad (18.44)$$

It is only in the absence of a zonal pressure gradient that the axial angular momentum for a fluid element is materially invariant just like for the point particle discussed in Section 11.7. The physical constraints for motion of the point particle, as described in Section 11.8, also hold for the fluid element. In particular, we can equate the zonal Coriolis acceleration to the zonal acceleration induced by axial angular momentum conservation. For example, a fluid element initially at rest in a fluid with zero zonal pressure gradient will zonally accelerate when moved meridionally (e.g., as from a meridional pressure gradient) according to the needs of axial angular momentum conservation.

18.4.1 Axial angular momentum conserving motion of a ring of air

Atmospheric and oceanic flows rarely experience a zero zonal pressure gradient. However, on a smooth spherical planet without meridional boundaries there is a zero zonally integrated zonal

pressure gradient

$$\int_{\text{ring}} \frac{\partial p}{\partial \lambda} d\lambda = 0. \quad (18.45)$$

Hence, a constant mass ring of fluid circling the smooth planet (Figure 18.3) will preserve its axial angular momentum in the absence of friction

$$\frac{d}{dt} \oint_{\text{ring}} \rho l^z dV = \oint_{\text{ring}} \rho \frac{Dl^z}{Dt} dV = - \oint_{\text{ring}} \frac{\partial p}{\partial \lambda} dV = 0. \quad (18.46)$$

For example, consider a latitudinal ring of constant mass inviscid fluid circling around the equator at radius R_e , and assume it is at rest relative to the rotating earth. The angular momentum per mass for this ring is $l^z = \Omega R_e^2$. Moving the ring vertically to $r \neq R_e$ while maintaining constant axial angular momentum induces a zonal velocity given by

$$u = \frac{\Omega (R_e^2 - r^2)}{r}. \quad (18.47)$$

Movement downward to $r < R_e$ leads to an eastward flow, $u > 0$, (westerly winds), whereas upward motion leads to the opposite. Likewise, moving the ring latitudinally while keeping $r = R_e$ leads to the zonal velocity

$$u = \frac{\Omega R_e \sin^2 \phi}{\cos \phi}. \quad (18.48)$$

Since $\cos \phi \geq 0$ on the sphere, latitudinal motion away from the equator while preserving axial angular momentum leads to eastward flow ($u > 0$) whether the ring is moved northward or southward.

18.4.2 Sketching the atmospheric angular momentum budget

How realistic is it to have coherent rings of inviscid air circulating around the planet at all latitudes? To answer this question we insert some numbers for a ring of radius R_e that starts with zero relative velocity at the equator. Equation (18.48) says that the westerly winds induced by axial angular momentum conserving motion have the following speeds at a selection of latitudes

$$u(10^\circ) = 14 \text{ m s}^{-1} \quad u(20^\circ) = 58 \text{ m s}^{-1} \quad u(30^\circ) = 134 \text{ m s}^{-1}. \quad (18.49)$$

The values at higher latitudes grow unbounded since $\cos \phi \rightarrow 0$ as the poles are approached. So there is a problem with an idealized theory of atmospheric circulation based on axial angular momentum conserving rings of air. It turns out that inviscid axial angular momentum conserving ideas extend only so far as the Hadley circulation is concerned, with the latitude extents of the Hadley Cell extending only to the middle latitudes. There are two missing ingredients to a more realistic theory: (i) frictional dissipation between the atmosphere and land; (ii) baroclinic eddies that contribute to poleward transport of angular momentum. It is outside of our scope to detail these physical processes and the corresponding atmospheric circulation. Instead, we here simply sketch the scene following Section 10.3 of [Holton \(1992\)](#) and Section 8.2 of [Marshall and Plumb \(2008\)](#).

Given that the earth has a near constant rotation rate, we can examine a zonally integrated axial momentum budget for the atmosphere and assume that the axial angular momentum of the earth is fixed.³ Rather than an inviscid fluid, we consider the dissipation due to boundary layer

³In fact, the length of a day fluctuates by roughly 10^{-3} s over a seasonal cycle due to transfer of axial angular momentum between the atmosphere and land.

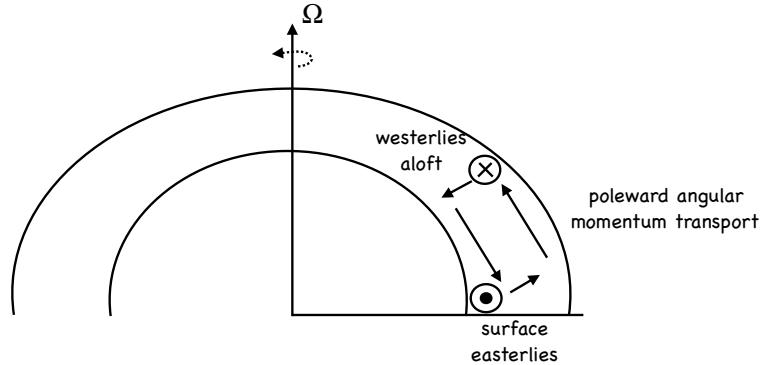


FIGURE 18.4: A sketch of the Hadley circulation. At the low latitudes, low level winds are predominantly easterly (blowing to the west) and so the atmosphere gains positive axial angular momentum from the earth via frictional boundary layer processes. As air rises and moves poleward it zonally accelerates according to conservation of axial angular momentum, thus becoming the mid-latitude westerlies (blowing to the east). The westerlies are generally unstable thus producing baroclinic eddies that vertically transfer axial angular momentum via form stress, thus depositing axial angular momentum back to the earth to close the axial angular momentum cycle.

friction and/or vertical eddy form stress from baroclinic eddies (we study form stress in Chapter 20). In the tropical atmosphere, the atmosphere generally has less angular momentum than the earth given that the winds are predominantly easterly (lower branch of the Hadley Cell). In this region, boundary layer interactions transfer angular momentum from the earth to the atmosphere. In contrast, the presence of middle latitude westerlies aloft (upper branch of the Hadley Cell) signal that the atmosphere has more angular momentum than the earth at these latitudes. There is an implied transfer of angular momentum from the atmosphere to the earth, largely mediated through vertical form stresses due to baroclinic eddies.

As sketched in Figure 18.4, a steady state axial angular momentum budget is realized through a meridional transport of axial angular momentum from the lower latitude atmosphere to the middle latitudes. This transport is mediated partly through an atmospheric overturning circulation (Hadley Cell) and partly through the vertical transfer of momentum via synoptic scale baroclinic eddies (see Section 50.2.4). Furthermore, dissipation by eddies and boundary layer interactions is crucial to realize the observed zonal wind profile. Namely, as seen above, a non-dissipative angular momentum conserving atmosphere exhibits winds that are far larger in magnitude than observed, particularly when moving to higher latitudes.

18.4.3 Further study

In Section 20.4 we consider the zonally integrated axial angular momentum budget for the ocean with sloping solid-earth bottom boundary as well as the upper surface (ocean) boundary. In particular, we see how boundary form stresses (Section 20.1) affect the angular momentum in addition to boundary frictional stresses. That analysis is analogous to that for the atmosphere given in Section 10.3 of [Holton \(1992\)](#).

18.5 Acoustic waves

In building up our understanding of geophysical fluid dynamics, it is useful to focus on selected pieces of the dynamical equations to see what sorts of phenomena they encompass. In that spirit, we here temporarily discard all portions of the momentum equation (18.16) having to do with geophysics. Namely, we consider a homogeneous perfect non-rotating fluid in free space, so that

the only acceleration arises from pressure gradients

$$\rho \frac{D\mathbf{v}}{Dt} = -\nabla p. \quad (18.50)$$

Mass remains materially conserved, so that this form of the Euler equation is coupled to the mass continuity equation

$$\partial_t \rho + \nabla \cdot (\mathbf{v} \rho) = 0, \quad (18.51)$$

with an assumed equation of state

$$\rho = \rho(p) \quad (18.52)$$

providing a relation between pressure and density.⁴ Although highly simplified relative to their geophysically more complete form, these equations are coupled nonlinear partial differential equations that are tough to solve in general. In this situation it is generally quite useful to examine the linearized form of the equations. As we now show, the linearized form of these equations supports acoustic waves.

18.5.1 Rudiments of acoustic waves

Consider a motionless base state and examine small amplitude fluctuations about that state. Since we are ignoring gravity, the pressure and density of the base state are uniform constants, p_0 and ρ_0 . Linearizing the Euler equation (18.50) and mass continuity equation (18.51) leads to the linear system

$$\rho_0 \frac{\partial \mathbf{v}}{\partial t} = -\nabla p \quad (18.53a)$$

$$\frac{\partial p}{\partial t} + c_s^2 \rho_0 \nabla \cdot \mathbf{v} = 0. \quad (18.53b)$$

In the continuity equation we introduced the inverse squared acoustic (sound) speed of the base state

$$c_s^{-2} = \left[\left(\frac{\partial \rho}{\partial p} \right)_S \right]_{p=p_0}, \quad (18.54)$$

where the pressure derivative is computed with a fixed entropy, S (matter concentration is also held fixed when considering sound speeds in a non-homogeneous fluid). We assume the entropy is fixed when computing the density derivative since the frequency of acoustic oscillations are so high that there is no time for heat to flow. Hence, sound waves are reversible adiabatic waves, which accords with our use of the perfect fluid equations. We offer further discussion of the sound speed in Sections 21.5.2 and 21.9.8, as part of the thermodynamics chapter.

The velocity, \mathbf{v} , in equation (18.53a) and pressure, p in equation (18.53b) are fluctuations relative to the base state. Taking the divergence of the velocity equation (18.53a) and making use of the continuity equation (18.53b) leads to

$$(\partial_{tt} - c_s^2 \nabla^2) p = 0, \quad (18.55)$$

which is the canonical wave equation discussed in Section 3.7. The solutions to this equation are linear and non-dispersive *acoustic waves* or *sound waves* with phase speed c_s . Since the waves are non-dispersive, their group speed equals to their phase speed. Additionally, the linear velocity equation (18.53a) reveals that pressure fluctuations and fluid particle fluctuations are in the same

⁴We study equations of state, including the ideal gas equation, in Chapters 21 and 25.

direction, with such waves known as *longitudinal waves*. Waves encountered later in this book appearing in incompressible flows are *transverse waves*, where $\mathbf{k} \cdot \mathbf{v} = 0$, so that fluid particle motion is perpendicular to the wave direction.

Acoustic waves propagate in all directions and they carry information about pressure fluctuations. Their squared speed is determined by isentropic compressibility of the base state. For example, a more compressible media, such as the atmosphere, has a smaller sound speed ($c_s \approx 350 \text{ m s}^{-1}$) than a less compressible media, such as the ocean ($c_s \approx 1500 \text{ m s}^{-1}$). Indeed, acoustic waves have infinite phase speed when the media is incompressible, with the infinite phase speed a signature that the hyperbolic wave system has converted to an elliptic system (see Chapter 3).

The *Mach number* is the ratio of the fluid particle speed to the sound speed. If a fluid is moving with Mach number greater than unity (super-sonic), then there can be discontinuities (shocks) that break the continuum approximation (Chapter 12). In this case, the continuous fluid equations must be supplemented by other physical conditions such as those afforded by molecular dynamics. We have no occasion to study super-sonic flows in this book.

18.5.2 Further study

Our treatment of acoustic waves is very terse, largely since we only wish to expose their role in propagating pressure fluctuations rather than embark on the rather interesting, yet sizable, job of studying the kinematics and dynamics of sound in the atmosphere and ocean. That topic is outside our scope. There are many pedagogical treatments that one can find, such as the physics treatment given in Section 16.5 of *Thorne and Blandford* (2017) and the oceanographic treatment in Chapter 7 of *Apel* (1987). *Vallis* (2017) discusses acoustic-gravity waves for an ideal gas in his Section 7.8, with these waves feeling both compressibility and gravity. See also his Chapter 6 for a lucid summary of wave kinematics.

18.6 Exercises

EXERCISE 18.1: THEOREM OF STRESS MEANS (*Aris* (1962) EXERCISE 5.12.2)

Make use of Cauchy's equation of motion (18.14) and the divergence theorem to prove the *theorem of stress means*

$$\oint_{\partial\mathcal{R}} \Psi \mathbb{T}_{pq} \hat{n}_q d\mathcal{S} = \int_{\mathcal{R}} \left[\mathbb{T}_{pq} \partial_q \Psi + \rho \Psi \left(\frac{Dv_p}{Dt} - f_p \right) \right] dV, \quad (18.56)$$

where Ψ is an arbitrary differentiable function, and \hat{n}_q is the q 'th component of the outward normal vector on $\partial\mathcal{R}$.

EXERCISE 18.2: TIME SYMMETRY OF THE EULER EQUATION

In Section 11.3 we studied the needs for time reversal symmetry in the equations of a point particle. Here we ask the same question for the Euler equation (18.18) in a rotating reference frame. Consider a solution, $\mathbf{v}(\mathbf{x}, t)$, to the Euler equation (18.18) in a rotating reference frame in either free space (i.e., no boundaries) or with fixed material boundaries. What transformation properties are needed for $p(\mathbf{x}, t)$, $\rho(\mathbf{x}, t)$, $\Phi(\mathbf{x}, t)$, and $\boldsymbol{\Omega}$ to ensure that $\mathbf{v}^*(\mathbf{x}, t^*) = -\mathbf{v}(\mathbf{x}, -t)$ is also a solution? That is, what time symmetry properties ensure that the velocity is reversible so that if $\mathbf{v}(\mathbf{x}, t)$ is a solution then so is $-\mathbf{v}(\mathbf{x}, -t)$? Discuss your answer.

EXERCISE 18.3: AREA OF A STEADY 1D LAMINAR JET EMANATING FROM A DOWNWARD NOZZLE

Consider a steady state laminar jet of constant density and inviscid water emanating from a downward facing nozzle with a constant prescribed volume flow rate, Q (dimensions volume per time).

Ignore surface tension and assume the air pressure acting on the surface of the jet is constant all along the jet. You can solve this exercise by making use of the steady vertical momentum equation and the steady mass continuity equation.

- (a) Explain why we can set $dp/dz = 0$ within the jet once it leaves the nozzle.
- (b) Determine an expression for the area of the jet, $A(z)$, as a function of distance, z , from the nozzle, with the nozzle placed at $z = 0$ and $z < 0$ a position beneath the nozzle. In addition to z , your expression will contain Q , g , and $A(0)$.
- (c) Is the area of the jet getting smaller or larger as the water moves downward away from the nozzle? Does this answer agree with your experience?
- (d) If the downward speed of water at the nozzle is $w(0) = 0.5 \text{ m s}^{-1}$, then at what vertical position, z , is the area of the jet four times different than at $z = 0$?



Stress in fluids

A fluid element experiences two kinds of forces. *External* or *body* forces act throughout the fluid element and arise from sources external to the matter acted upon. The accumulated effects from a body force within a fluid region results from integrating the body force over each point within a region. In geophysical fluid mechanics, we are concerned with body forces from the effective gravitational acceleration (central gravity plus planetary centrifugal) as well as the Coriolis acceleration. Body forces are also experienced by the point particles studied in Part II in this book.

The second kind of forces are called *internal* or *contact* forces. These forces act on a fluid element due to its local interactions with the surrounding fluid media through the exchange of dynamical properties (e.g., momentum, kinetic energy, vorticity), the exchange of matter, and the exchange of thermodynamical properties such as enthalpy and entropy. Contact forces are specific to the mechanics of continuous media. Their existence arises since a fluid element is not a point particle but instead it has nonzero volume with contact forces acting on its bounding surface. Hence, contact forces embody the fundamental distinction between the dynamics of a fluid element and the dynamics of a point particle.

The exchange of dynamical properties between fluid elements is mediated by mechanical interactions, with this exchange mathematically formulated as a stress (force per unit area) acting on the surface boundary area of the fluid element. As a force per unit area, a stress is associated with two directions: the direction of the force and the direction normal to the area acted upon by the force. Correspondingly, stresses acting on a fluid element are naturally organized into a second order *stress tensor*. Contact forces (given by the stress times an area element) satisfy Newton's third law, also known as the action/reaction law. Hence, the net contact force acting on a finite region arises just from the contact forces acting at the region boundary.

Surface tension is a rather unique kind of stress acting on fluid boundaries such as oil and water or air and water. As shown in Section 19.10, surface tension is generally negligible for length scales larger than a few centimeters. It is for this reason that surface tension is commonly absent from books on geophysical fluid mechanics, where most concern is with larger length scales. Nevertheless, the effects are important if studying physical processes associated with air-sea interactions, such as tracer, heat, and momentum exchange through bubbles, droplets, and capillary-gravity waves.

READER'S GUIDE TO THIS CHAPTER

We introduced contact forces in Chapter 18 when deriving the fluid equations of motion. In this chapter we dive deeper into contact forces and their corresponding stresses. We also discuss conditions placed on stress and velocity at boundaries. The material in this chapter builds from our study of forces, acceleration, and Newton's laws as applied to a fluid continuum in Chapter 18. Fluid stresses can be organized into a second order tensor and further decomposed into

pressure stresses and viscous/frictional stresses. Understanding the mathematical and physical aspects of stress is important for the suite of fluid models studied in this book. Because the material involves vectors and tensors it can require a bit more patience from the reader than analogous chapters that discuss scalar fields. To make the formalism less mathematically intense, we employ Cartesian tensors as discussed in Chapters 1 and 2. Results can be generalized to arbitrary coordinates through the general tensor analysis detailed in Chapter 7.

19.1	Cauchy's stress principle and Newton's laws	371
19.1.1	Cauchy's stress principle	371
19.1.2	Newton's third law and local equilibrium of stresses	371
19.1.3	Comments on the local equilibrium relation	372
19.2	The stress tensor	373
19.3	Angular momentum and stress tensor symmetry	376
19.3.1	Basic formulation	376
19.3.2	Physical interpretation	377
19.3.3	Symmetry of the stress tensor	377
19.4	Forces and torques in a static fluid	378
19.4.1	Force balance	379
19.4.2	Pressure force balance for a homogeneous fluid	379
19.4.3	Torque balance	381
19.5	Flux-form Eulerian momentum equation	382
19.6	Linear momentum for arbitrary regions	383
19.7	Constitutive relation between stress and rate of strain	384
19.7.1	Thermodynamic pressure and mechanical pressure	384
19.7.2	Couette flow and the viscous tensor	385
19.7.3	Guidance from Galilean invariance	386
19.7.4	A comment on Rayleigh drag	386
19.7.5	Constitutive relation for Newtonian fluids	386
19.7.6	Navier-Stokes and Euler equations	389
19.7.7	Laplacian friction in terms of vorticity and divergence	389
19.7.8	Frictional stresses in a sheared flow	390
19.7.9	The net stress tensor	390
19.7.10	Comments and further study	391
19.8	Reynolds number and flow regimes	391
19.8.1	Non-dimensional Navier-Stokes	392
19.8.2	Ratio of inertial to frictional accelerations	392
19.8.3	Reynolds numbers for some example flows	392
19.9	Stress on an interface	394
19.9.1	General formulation	394
19.9.2	Solid material boundary	394
19.9.3	No-slip boundary condition	395
19.9.4	Lagrangian interface	396
19.9.5	Permeable interface	396
19.9.6	Summary comments	397
19.9.7	Comments on boundary layers	397
19.10	Surface tension	398
19.10.1	Force balance on an air-water interface	399
19.10.2	Young-Laplace formula	400
19.10.3	Some oceanographic examples	402
19.10.4	Further study	403
19.11	Exercises	403

19.1 Cauchy's stress principle and Newton's laws

We here develop some general properties of contact forces and the associated stresses. For that purpose, consider an arbitrary smooth closed region, \mathcal{R} , of fluid with volume $V = \int_{\mathcal{R}} dV$ and mass $M = \int_{\mathcal{R}} \rho dV$ (Figure 19.1). Furthermore, let $\partial\mathcal{R}$ be the bounding surface for the region, and let $\hat{\mathbf{n}}$ be the outward normal at a point on the boundary.

19.1.1 Cauchy's stress principle

The bounding surface of the region experiences mechanical interactions with the surrounding fluid continuum and these interactions lead to contact forces acting on the boundary. Let $\boldsymbol{\tau}$ be the stress vector (force per unit area) acting at a point on $\partial\mathcal{R}$. Cauchy's stress principle asserts that the stress vector is a function of the position, time, and boundary normal

$$\boldsymbol{\tau} = \boldsymbol{\tau}(\mathbf{x}, t, \hat{\mathbf{n}}). \quad (19.1)$$

The dependence on boundary normal means that the stress acting on a surface is generally a function of the orientation of that surface. This form of the stress trivially holds for an exactly hydrostatic fluid where the stress vector is proportional to the pressure (Section 18.3). Furthermore, the stress from pressure is oriented along the inward normal, thus reflecting the purely compressive nature of pressure

$$\boldsymbol{\tau} = -p(\mathbf{x}, t) \hat{\mathbf{n}} \quad \text{static fluid in hydrostatic balance.} \quad (19.2)$$

Cauchy's stress principle is sensible for points within the fluid media, and its relevance has been supported by experimental studies over the time since Cauchy made this assertion in the year 1823. Furthermore, it holds for pressure and viscous stresses at the interface between fluid media or at solid-earth boundaries. However, Cauchy's stress principle does not hold for surface tension, which is proportional to the curvature of the surface separating two fluid media (e.g., atmosphere and ocean), where curvature involves spatial gradients of the normal vector. As discussed in Section 19.10, surface tension effects are important for relatively small scales (order centimeters), and as such play a minor role in this book. Hence, with the single exception of surface tension, we rely on Cauchy's stress principle to formulate the fluid dynamical equations.

19.1.2 Newton's third law and local equilibrium of stresses

Newton's second law says that in an inertial reference frame, unbalanced forces acting on a physical system affect a time change to the linear momentum. Consider a region, $\mathcal{R}(\mathbf{v})$, whose fluid elements follow the barycentric velocity. Newton's law then states that the material time evolution of the region's linear momentum is given by

$$\frac{d}{dt} \int_{\mathcal{R}(\mathbf{v})} \mathbf{v} \rho dV = \int_{\mathcal{R}(\mathbf{v})} \mathbf{f} \rho dV + \oint_{\partial\mathcal{R}(\mathbf{v})} \boldsymbol{\tau} d\mathcal{S}, \quad (19.3)$$

where $\int_{\mathcal{R}(\mathbf{v})} \mathbf{f} \rho dV$ is the domain integrated body force (from central gravity, planetary centrifugal, and Coriolis). To develop a general property for the contact forces, consider this balance for a region whose size gets infinitesimally small. Assuming the integrands for the two volume integrals are well behaved (i.e., smooth and bounded) as the region size goes to zero, we see that the volume integrals are proportional to L^3 , where L is a length scale measuring the size of the region (e.g., side for a cubical region or diameter for a spherical region). In the same manner, we assume the stresses

are well behaved in the case of an infinitesimal region. However, the integral of the contact forces goes to zero at the slower rate that is proportional to L^2 . Self-consistency for the balance (19.3) over a region of infinitesimal size thus requires the contact forces to satisfy the limiting behavior

$$\lim_{L \rightarrow 0} \frac{1}{L^2} \oint_{\partial \mathcal{R}(\mathbf{v})} \boldsymbol{\tau} d\mathcal{S} = 0. \quad (19.4)$$

This behavior means that contact forces at a point in the fluid must be in local equilibrium. Equation (19.4) is sometimes referred to as *Cauchy's fundamental lemma*.

A direct implication of the local equilibrium statement is that stress vectors that respect Cauchy's principle (19.1) satisfy

$$\boldsymbol{\tau}(\mathbf{x}, t, \hat{\mathbf{n}}) = -\boldsymbol{\tau}(\mathbf{x}, t, -\hat{\mathbf{n}}). \quad (19.5)$$

For example, the stress vector on one side of a surface is equal and oppositely directed to the stress vector acting on the other side. This equation is an expression of Newton's third law of mechanics (the action/reaction law), here written in terms of the stresses acting in a continuous media. It is of fundamental importance throughout our study of contact forces and their associated stresses acting within the fluid and at boundaries. We thus see how an application of Newton's second law, the linear momentum principle (19.3) for a continuous media, leads to a statement of Newton's third law, (19.5), for contact forces.

As an example of the above ideas, the simplest stress we consider in this chapter is that from pressure, with pressure acting solely in a compressive manner so that the stress vector takes the form

$$\boldsymbol{\tau}^{\text{press}}(\mathbf{x}, t, \hat{\mathbf{n}}) = -p(\mathbf{x}, t) \hat{\mathbf{n}}. \quad (19.6)$$

This stress trivially satisfies the Newton's third law relation (19.5) since

$$\boldsymbol{\tau}^{\text{press}}(\mathbf{x}, t, \hat{\mathbf{n}}) = -\boldsymbol{\tau}^{\text{press}}(\mathbf{x}, t, -\hat{\mathbf{n}}). \quad (19.7)$$

19.1.3 Comments on the local equilibrium relation

The local equilibrium relation (19.4), and the corresponding expression of Newton's third law, (19.5), might suggest that stresses cannot lead to motion. However, that suggestion is incorrect since stresses integrated over a finite region can lead to a net force that causes motion. Since contact forces within the domain interior cancel pointwise, the local equilibrium relation (19.4) says that the net contact force acting on the region arises only from the area integrated stresses acting on the region boundary. Local or pointwise mechanical equilibrium does not imply mechanical equilibrium for finite regions.

To further emphasize the above point, consider an ocean region bounded at its bottom by the solid earth and its upper surface by a massive atmosphere. Variations (divergences) in stresses over finite regions within the ocean fluid lead to accelerations; e.g., ocean circulation. However, when integrated over the full ocean domain, all stresses cancel pointwise. Consequently, the net contact force acting on the full ocean domain reduces to the contact force acting just on the ocean boundaries. The boundary contact forces arise from mechanical interactions with the solid-earth and the overlying atmosphere. The center of mass for the ocean basin remains static if the accumulation of forces sum to zero, which includes the contact forces acting over its boundaries plus the volume integrated body forces from effective gravity (central gravity plus planetary centrifugal) and Coriolis.

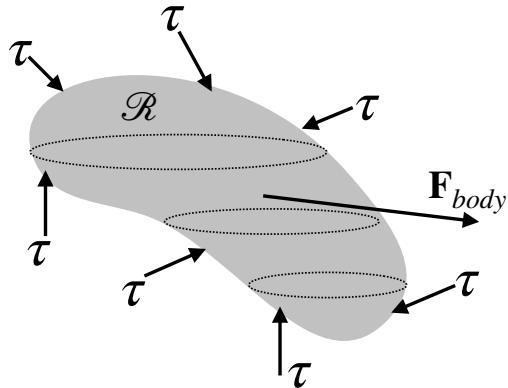


FIGURE 19.1: Schematic of the net body force, \mathbf{F}_{body} , acting on a finite region of fluid, plus the accumulation of stress vectors, τ , acting on the region boundaries. The net body force is determined by a volume integral of the body force (gravity, centrifugal, and Coriolis) at each point within the volume. In contrast, since the stresses are in local equilibrium, the volume integral of the stress divergence reduces to an area integral of the stress over the region boundary. Stress arises from pressure (compressive and normal) and strains (which then lead to viscous stresses when there is viscosity). The area integrated contribution from pressure to horizontal accelerations is referred to as *form stress*. The form stress coming from the bottom boundary is called the *topographic form stress*. The form stress appearing at the air-sea boundary is the *atmospheric form stress* if considering ocean dynamics and *oceanic form stress* if considering atmospheric dynamics. We study form stress in Chapter 20.

In Figure 19.2 we illustrate the net pressure force acting on an arbitrary fluid domain. Pressure acts solely in a compressive manner as directed along the inward normal to the domain. Area integration over a domain boundary renders the net pressure force acting on the domain

$$\mathbf{F}^{\text{pressure}} = - \oint_{\partial\mathcal{R}} p \hat{\mathbf{n}} d\mathcal{S} = - \int_{\mathcal{R}} \nabla p dV, \quad (19.8)$$

where the second equality follows from application of Gauss's divergence theorem for a scalar field (Section 2.7.2). When decomposed according to coordinate axes, the pressure force acting on the boundary has a component in both the vertical and horizontal directions, thus contributing to both vertical and horizontal accelerations. The vertical accelerations are closely balanced by the weight of fluid, with exact balance in the case of a hydrostatic fluid. The horizontal stresses from pressure are known as *form stress*. This name arises since the stress depends on the form, or shape, of the interface on which pressure acts.

19.2 The stress tensor

Cauchy's stress principle reduces the mathematical complexity of describing stress vectors. A further implication of this principle leads to *Cauchy's theorem*, which states that the stress vector, which is a function of space, time, and normal direction, can be expressed in terms of a stress tensor (a function of space and time) projected into the direction of the normal. The purpose of this section is to provide arguments supporting this theorem.

For this purpose, consider the tetrahedron fluid region shown in Figure 19.3, where three of the four sides are aligned according to the Cartesian coordinate axes and the fourth side has an outward normal, $\hat{\mathbf{n}} = (\hat{n}_1, \hat{n}_2, \hat{n}_3)$, projecting into all three directions. The results developed for this rather contrived region hold for an arbitrary region using arbitrary coordinates. The reason for this generality is that once we derive a tensorially correct result using one choice of coordinates,

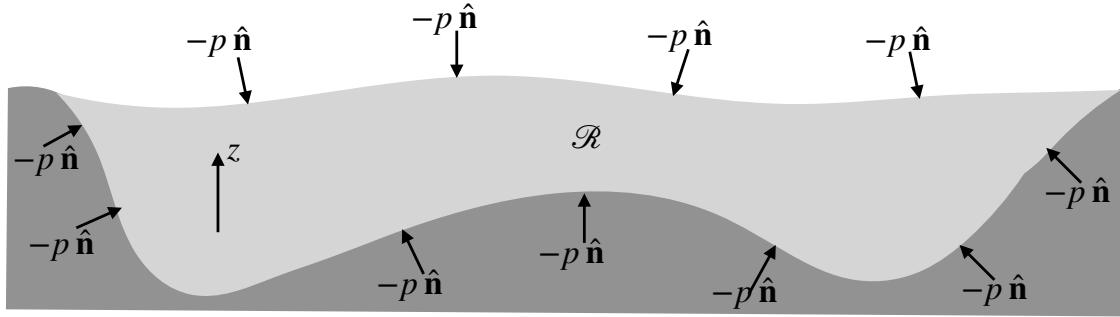


FIGURE 19.2: Schematic of contact forces from pressure acting on the boundaries to an ocean domain. Pressure forces are directed according to minus the local normal since pressure is a compressive force aligned with the inward normal direction. As with all contact forces, the pressure forces acting in the interior of the ocean are locally in mechanical equilibrium. Hence, when integrated over the global domain the net pressure forces only arise at the domain boundaries. That is, the net pressure force acting on the full ocean domain arises only at the interface between the solid-earth and the ocean, plus the interface between the atmosphere and the ocean. Note that the pressure force has a component in both the vertical and horizontal directions as per the orientation of the local normal vector. Further boundary stresses arise from viscous exchange, which generally have components perpendicular to the boundary normal; i.e., tangential to the boundary. Such stresses also satisfy Newton's third law.

such as Cartesian used here, we can make use of general tensor analysis (Chapters 6 and 7) to move from specific coordinates to arbitrary coordinates.

In the limit that the tetrahedron size goes to zero, local equilibrium of the contact forces means that

$$-\sum_{m=1}^3 \boldsymbol{\tau}_{(m)} dA_m + \boldsymbol{\tau}_{\hat{n}} dA = 0, \quad (19.9)$$

where we use the shorthand expression for the outward normal directed stress vector

$$\boldsymbol{\tau}(\mathbf{x}, t, \hat{n}) = \boldsymbol{\tau}_{\hat{n}}. \quad (19.10)$$

In equation (19.9), $\boldsymbol{\tau}_{(m)} dA_m$ (no implied summation) is the contact force vector acting on the face with outward normal parallel to the corresponding coordinate axis and $\boldsymbol{\tau}_{\hat{n}} dA$ is the contact force acting on the slanted face with outward normal \hat{n} . The minus sign arises for the summation term since the outward normals for these three faces point in the negative coordinate directions, and our convention is for $\boldsymbol{\tau}_{(m)}$ to align with the positive coordinate directions. The areas for each face are related to the slanted face area through

$$dA_m = \hat{n}_m dA, \quad (19.11)$$

so that the local equilibrium relation (19.9) becomes

$$\boldsymbol{\tau}_{\hat{n}} = \sum_{m=1}^3 \hat{n}_m \boldsymbol{\tau}_{(m)}. \quad (19.12)$$

Equation (19.12) can be organized into a matrix-vector equation

$$\begin{bmatrix} (\boldsymbol{\tau}_{\hat{n}})_1 \\ (\boldsymbol{\tau}_{\hat{n}})_2 \\ (\boldsymbol{\tau}_{\hat{n}})_3 \end{bmatrix} = \begin{bmatrix} \hat{n}_1 & \hat{n}_2 & \hat{n}_3 \end{bmatrix} \begin{bmatrix} \boldsymbol{\tau}_{(1)1} & \boldsymbol{\tau}_{(1)2} & \boldsymbol{\tau}_{(1)3} \\ \boldsymbol{\tau}_{(2)1} & \boldsymbol{\tau}_{(2)2} & \boldsymbol{\tau}_{(2)3} \\ \boldsymbol{\tau}_{(3)1} & \boldsymbol{\tau}_{(3)2} & \boldsymbol{\tau}_{(3)3} \end{bmatrix}, \quad (19.13)$$

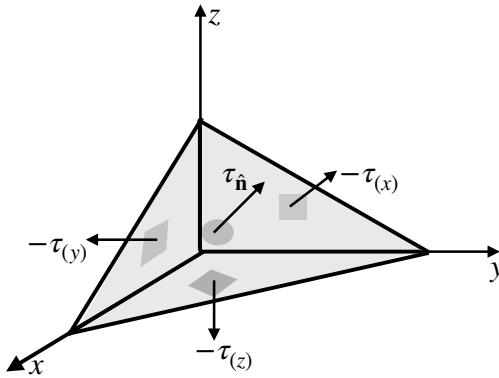


FIGURE 19.3: Tetrahedron region of fluid with stresses acting on the four faces. Note that the stresses are not necessarily directed normal to the faces. Local equilibrium of stresses means that the accumulation of these four stresses around the region goes to zero as the region volume goes to zero.

where each matrix element is the n -component of the m -stress $\tau_{(m)}$. We introduce a less cumbersome notation by writing

$$\begin{bmatrix} (\tau_{\hat{n}})_1 \\ (\tau_{\hat{n}})_2 \\ (\tau_{\hat{n}})_3 \end{bmatrix} = [\hat{n}_1 \quad \hat{n}_2 \quad \hat{n}_3] \begin{bmatrix} \mathbb{T}_{11} & \mathbb{T}_{12} & \mathbb{T}_{13} \\ \mathbb{T}_{21} & \mathbb{T}_{22} & \mathbb{T}_{23} \\ \mathbb{T}_{31} & \mathbb{T}_{32} & \mathbb{T}_{33} \end{bmatrix}, \quad (19.14)$$

so that \mathbb{T}_{mn} measures the force per area in the n -direction along a surface whose outward normal points in the m -direction, as depicted in Figure 19.4. Making use of \mathbb{T}_{mn} in the expression (19.12) leads to

$$(\tau_{\hat{n}})_n = \sum_{m=1}^3 \hat{n}_m \mathbb{T}_{mn}, \quad (19.15)$$

which can be written more succinctly as

$$\tau_{\hat{n}} = \hat{n} \cdot \mathbb{T}. \quad (19.16)$$

We thus see that the stress vector acting on a surface oriented according to a normal vector, \hat{n} , equals to the projection of the *stress tensor*, \mathbb{T} , onto the normal vector. Exposing functional dependence reveals

$$\tau_{\hat{n}}(\mathbf{x}, t, \hat{n}) = \hat{n} \cdot \mathbb{T}(\mathbf{x}, t), \quad (19.17)$$

which manifests Cauchy's theorem. Namely, the stress vector $\tau_{\hat{n}}$, which is a function of (\mathbf{x}, t, \hat{n}) , has been decomposed into a stress tensor, \mathbb{T} , which is a function of (\mathbf{x}, t) , as well as the projection of the stress tensor into a direction \hat{n} .

How do we know that \mathbb{T}_{mn} form the components to a tensor rather than just being elements of a 3×3 matrix? To answer this question we note that each component of \mathbb{T} is a force per area, with force a vector and area orientable by its outward normal. As it is built from vectors, which are first order tensors, we suspect that \mathbb{T} should be a proper second order tensor. This suspicion is supported by the quotient rule from tensor analysis (Section 6.1.1). Namely, the quotient rule means that equation (19.15) indeed yields \mathbb{T}_{mn} that are components to a second order tensor. As components to a second order tensor, the tensor components, \mathbb{T}_{mn} , transform under a coordinate transformation according to the rules developed in Chapter 1 for Cartesian tensors and in Chapter 7 for general tensors. Through the power of tensor analysis, we thus see that our considerations, based on the rather contrived tetrahedron region in Figure 19.3, hold for an arbitrary region described by an arbitrary coordinate system.

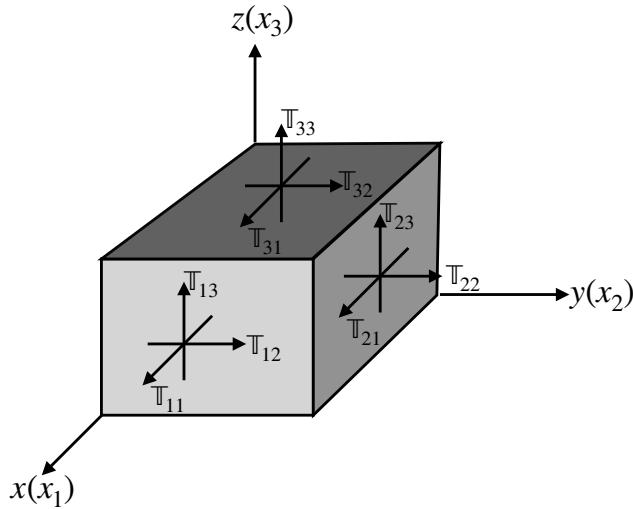


FIGURE 19.4: Illustrating the components to the stress tensor, \mathbb{T}_{mn} and how they are organized according to the coordinate axes. The component \mathbb{T}_{mn} is the stress that points in the n -direction along the face with outward normal in the m -direction.

19.3 Angular momentum and stress tensor symmetry

The linear momentum principle afforded by Newton's law of motion allowed us to deduce the local equilibrium property (19.4) of the stress. We here derive a constraint placed on the stress tensor that is imposed by studying angular momentum. Phenomenologically, we observe that geophysical fluids, as with most common fluids, experience torques only as the moments of body forces acting throughout the volume of a fluid region, or as moments of contact forces acting on the surface bounding the fluid region. We now make use of this observation to deduce symmetry of the stress tensor.¹

19.3.1 Basic formulation

Consider a constant mass fluid element that has a Cartesian position \mathbf{x} relative to an arbitrary origin. The angular momentum of the fluid element with respect to the origin is

$$\mathbf{L} = \rho \delta V (\mathbf{x} \wedge \mathbf{v}), \quad (19.18)$$

and its material time evolution is

$$\frac{D\mathbf{L}}{Dt} = \rho \delta V \mathbf{x} \wedge \frac{D\mathbf{v}}{Dt}, \quad (19.19)$$

which follows since $D(\rho \delta V)/Dt = 0$, $D\mathbf{x}/Dt = \mathbf{v}$, and $\mathbf{v} \wedge \mathbf{v} = 0$. Making use of Cauchy's form for the equation of motion (18.14)

$$\rho \frac{D\mathbf{v}}{Dt} = \rho \mathbf{f} + \nabla \cdot \mathbb{T} \quad (19.20)$$

allows us to write the angular momentum evolution as

$$\frac{D\mathbf{L}}{Dt} = \delta V \mathbf{x} \wedge (\rho \mathbf{f} + \nabla \cdot \mathbb{T}). \quad (19.21)$$

¹Page 11 of [Batchelor \(1967\)](#) and Section 5.13 of [Aris \(1962\)](#) offer brief discussions of fluids in which internal force couplets lead to torques distinct from those considered here, and in which the stress tensor has an anti-symmetric component. [Dahler and Scriven \(1961\)](#) provide a more thorough account.

The first term arises from body forces (e.g., central gravity, planetary centrifugal, and Coriolis) and the second term arises from the divergence of stresses. Expanding the stress divergence term renders

$$\left[\frac{DL_m}{Dt} \right]_{\text{stress}} = \delta V \epsilon_{mnp} x_n (\nabla \cdot \mathbb{T})_p \quad (19.22a)$$

$$= \delta V \epsilon_{mnp} x_n \partial_q \mathbb{T}_{pq} \quad (19.22b)$$

$$= \delta V \epsilon_{mnp} [\partial_q (x_n \mathbb{T}_{pq}) - (\partial_q x_n) \mathbb{T}_{pq}] \quad (19.22c)$$

$$= \delta V \epsilon_{mnp} [\partial_q (x_n \mathbb{T}_{pq}) - \mathbb{T}_{pn}], \quad (19.22d)$$

where the final equality follows since $\partial_q x_n = \delta_{qn}$. Bringing this result back into the full expression (19.21) leads to

$$\frac{DL_m}{Dt} = \delta V \epsilon_{mnp} [\rho x_n f_p + \partial_q (x_n \mathbb{T}_{pq}) - \mathbb{T}_{pn}]. \quad (19.23)$$

19.3.2 Physical interpretation

To facilitate a physical interpretation of the terms appearing in equation (19.23), integrate over an arbitrary Lagrangian region (region moving with the barycentric velocity, \mathbf{v}) so that

$$\frac{d}{dt} \int_{\mathcal{R}(\mathbf{v})} L_m = \int_{\mathcal{R}(\mathbf{v})} \epsilon_{mnp} [\rho x_n f_p + \partial_q (x_n \mathbb{T}_{pq}) - \mathbb{T}_{pn}] dV. \quad (19.24)$$

As noted earlier, the first term on the right hand side arises from torques due to body forces acting over the region

$$\int_{\mathcal{R}(\mathbf{v})} \epsilon_{mnp} (\rho x_n f_p) dV = \int_{\mathcal{R}(\mathbf{v})} (\mathbf{x} \wedge \mathbf{f})_m \rho dV. \quad (19.25)$$

The second term on the right hand side of equation (19.24) can be transferred into a surface integral using the divergence theorem

$$\int_{\mathcal{R}(\mathbf{v})} \epsilon_{mnp} \partial_q (x_n \mathbb{T}_{pq}) dV = \int_{\partial \mathcal{R}(\mathbf{v})} \epsilon_{mnp} x_n \mathbb{T}_{pq} \hat{n}_q d\mathcal{S} = \int_{\partial \mathcal{R}(\mathbf{v})} (\mathbf{x} \wedge \boldsymbol{\tau})_m d\mathcal{S}, \quad (19.26)$$

where \hat{n}_q is the q 'th component of the normal vector on the region boundary, $\partial \mathcal{R}(\mathbf{v})$, and

$$\tau_p = \mathbb{T}_{pq} \hat{n}_q \quad (19.27)$$

is the p 'th component to the stress vector that is normal to $\hat{\mathbf{n}}$ (see equation (19.16)). Hence, the second term is the contribution to angular momentum evolution due to torques arising from the moment of contact forces acting on the region boundary.

19.3.3 Symmetry of the stress tensor

As noted at the start of this section, geophysical fluids have their angular momentum affected by torques arising from the moment of body forces acting throughout the fluid region, plus the moment of contact forces acting on the region boundary. There is a third term in equation (19.23) that does not fit into either category, and it is given by the volume integral

$$-\int_{\mathcal{R}(\mathbf{v})} \epsilon_{mnp} \mathbb{T}_{pn} dV = \int_{\mathcal{R}(\mathbf{v})} \epsilon_{mpn} \mathbb{T}_{pn} dV \equiv \int_{\mathcal{R}(\mathbf{v})} T_m^\times dV, \quad (19.28)$$

where we defined

$$T_m^\times = \epsilon_{mpn} \mathbb{T}_{pn}. \quad (19.29)$$

This term contributes a volume source to angular momentum and yet it is *not* associated with body forces. We might refer to it as a *torque density* (torque source per volume). As already noted, such torque sources are *not* relevant for geophysical fluids, in which case we conclude that geophysical fluids are affected only by symmetric stress tensors

$$\mathbb{T}_{mn} = \mathbb{T}_{nm} \implies \epsilon_{mnp} \mathbb{T}_{np} = 0. \quad (19.30)$$

Symmetry of the stress tensor is a central property of the stresses acting on most fluids, including geophysical fluids. We thus only consider symmetric stress tensors throughout this book.

To further support the above conclusion concerning a symmetric stress tensor, consider a particular component of the torque density, such as the vertical

$$T_3^\times = \epsilon_{3pn} \mathbb{T}_{pn} = \mathbb{T}_{12} - \mathbb{T}_{21}, \quad (19.31)$$

with the corresponding torque applied to a fluid element given by

$$T_3^\times \delta V = (\mathbb{T}_{12} - \mathbb{T}_{21}) \delta x \delta y \delta z. \quad (19.32)$$

What sort of angular acceleration is induced by this torque when computed relative to the fluid element center? To answer this question, assume the fluid element is moving as a rigid body so that we can compute its angular acceleration by dividing the torque by the moment of inertia for the fluid element. The moment of inertia depends on the shape of the element, which is unspecified. Even so, we can estimate the moment of inertia computed relative to a vertical axis through the center of the element

$$I_3 = \alpha [(\delta x)^2 + (\delta y)^2] \rho \delta x \delta y \delta z, \quad (19.33)$$

where α is a dimensionless geometric factor. Dividing the torque (19.32) by the moment of inertia thus leads to an estimate of the angular acceleration

$$\text{angular acceleration} \approx \frac{T_3^\times}{I_3} \approx \frac{\mathbb{T}_{12} - \mathbb{T}_{21}}{[(\delta x)^2 + (\delta y)^2] \rho \alpha}. \quad (19.34)$$

Now consider the continuum limit, found as δx and δy are reduced to zero. In the absence of an unspecified counteracting torque, a finite angular acceleration in the continuum limit (where $(\delta x)^2 + (\delta y)^2 \rightarrow 0$) is ensured only if the stress tensor is symmetric so that the numerator vanishes.

19.4 Forces and torques in a static fluid

In this section we return to the study of a static fluid in a gravitational field originally considered in Section 18.3. The exact solution is known as *exact hydrostatic balance*, which distinguishes it from the approximate hydrostatic balance appropriate for moving geophysical fluids under certain scaling regimes (Section 23.3). For a static fluid, all forces and all torques sum to zero at any point. Similarly, the integrated forces and integrated torques acting on any finite fluid region also vanish. The static fluid, although trivial dynamically, offers useful practice in applying the formalism of continuum mechanics to a system where we know the answer. Furthermore, there are interesting and important applications of these ideas, such as in the building of dams and underwater structures, both of which we certainly hope will remain static!

19.4.1 Force balance

The force balance in an exact hydrostatic fluid was addressed in Section 18.3 where we deduced the following relation between the pressure gradient and geopotential gradient

$$\nabla p = -\rho \nabla \Phi. \quad (19.35)$$

This equality holds at every point within the fluid, and as such it is a strong form of the hydrostatic balance.² Integrating over a finite fluid region, \mathcal{R} , and using the divergence theorem for scalar fields, (2.67), renders the finite volume or weak form of hydrostatic balance

$$\int_{\mathcal{R}} \rho \nabla \Phi \, dV = - \int_{\mathcal{R}} \nabla p \, dV = - \oint_{\partial \mathcal{R}} p \hat{\mathbf{n}} \, dS. \quad (19.36)$$

Expanding the above relations for the special case of $\Phi = g z$ leads to the differential statements

$$0 = \hat{\mathbf{x}} \cdot \nabla p = \hat{\mathbf{y}} \cdot \nabla p \quad \text{and} \quad \rho g = -\hat{\mathbf{z}} \cdot \nabla p, \quad (19.37)$$

with the first two equations implying that the exact hydrostatic pressure is only a function of z . The corresponding weak form of hydrostatic balance reads

$$g M = - \int_{\mathcal{R}} \nabla p \cdot \hat{\mathbf{z}} \, dV = - \oint_{\partial \mathcal{R}} p (\hat{\mathbf{n}} \cdot \hat{\mathbf{z}}) \, dS, \quad (19.38)$$

where $M = \int_{\mathcal{R}} \rho \, dV$ is the mass in the fluid region, and the weak form of horizontal balances are

$$0 = \int_{\mathcal{R}} \nabla p \cdot \hat{\mathbf{x}} \, dV = \oint_{\partial \mathcal{R}} p (\hat{\mathbf{n}} \cdot \hat{\mathbf{x}}) \, dS \quad (19.39a)$$

$$0 = \int_{\mathcal{R}} \nabla p \cdot \hat{\mathbf{y}} \, dV = \oint_{\partial \mathcal{R}} p (\hat{\mathbf{n}} \cdot \hat{\mathbf{y}}) \, dS. \quad (19.39b)$$

We emphasize that these finite volume balances hold for any arbitrarily shaped fluid region.

19.4.2 Pressure force balance for a homogeneous fluid

To further our understanding of the pressure force balances in a static fluid, consider a constant density static ocean sitting under a massless atmosphere, in which case the hydrostatic pressure is

$$p = -\rho g z, \quad (19.40)$$

where $z < 0$ for the ocean. Now examine the pressure forces acting on the three sides of the triangle in Figure 19.5. This geometry is simple enough to explicitly compute the pressure forces, thus confirming the general properties in equations (19.38), (19.39a), and (19.39b).

The outward normal vectors along the three triangle faces are given by

$$\hat{\mathbf{n}}_A = +\hat{\mathbf{y}} \quad \text{and} \quad \hat{\mathbf{n}}_B = -\hat{\mathbf{z}} \quad \text{and} \quad \hat{\mathbf{n}}_C = \hat{\mathbf{z}} \cos \varphi - \hat{\mathbf{y}} \sin \varphi, \quad (19.41)$$

where

$$\tan \varphi = \frac{z_2 - z_1}{y_2 - y_1} = \frac{\Delta z}{\Delta y} \quad (19.42)$$

²Recall the discussion of weak and strong formulation in Section 13.1. In brief, the weak formulation provides integral relations whereas the strong formulation provides differential relations.

is the slope of the hypotenuse relative to the horizontal. The integrated pressure force along the vertical face is thus given by

$$\mathbf{F}_A^{\text{press}} = - \int p \hat{\mathbf{n}}_A d\mathcal{S} = \hat{\mathbf{y}} \Delta x \int_{z_1}^{z_2} \rho g z dz = \hat{\mathbf{y}} (\rho g/2) (z_2 + z_1) \Delta z \Delta x, \quad (19.43)$$

where Δx is the thickness of the triangle in the $\hat{\mathbf{x}}$ direction into the page. Note that $\mathbf{F}_A^{\text{press}}$ points in the $-\hat{\mathbf{y}}$ direction since $z_2 + z_1 < 0$. Likewise, the integrated pressure force along the horizontal face is given by

$$\mathbf{F}_B^{\text{press}} = - \int p \hat{\mathbf{n}}_B d\mathcal{S} = -\hat{\mathbf{z}} \Delta x \int_{y_1}^{y_2} \rho g z_1 dy = -\hat{\mathbf{z}} \rho g z_1 \Delta y \Delta x, \quad (19.44)$$

which points upward since $z_1 < 0$.

The integrated pressure force along the sloped hypotenuse face, C , requires a bit of trigonometry. For this purpose we make use of the formalism from Section 15.4.3, in which the vertical position along the hypotenuse is written

$$z = \eta(y) = z_2 - (y_2 - y) \tan \varphi, \quad (19.45)$$

so that the horizontal projection of the surface area is given by equation (20.4)

$$d\mathcal{S} = |\nabla(z - \eta)| dx dy = \frac{dx dy}{|\cos \varphi|}. \quad (19.46)$$

Hence, the integrated pressure force on the hypotenuse is given by

$$\mathbf{F}_C^{\text{press}} = - \int p \hat{\mathbf{n}}_C d\mathcal{S} \quad (19.47a)$$

$$= \frac{\hat{\mathbf{n}}_C}{\cos \varphi} \Delta x \int_{y_1}^{y_2} \rho g \eta(y) dy \quad (19.47b)$$

$$= (\hat{\mathbf{z}} - \hat{\mathbf{y}} \tan \varphi) \rho g \Delta x [z_2 \Delta y - y_2 \Delta z - (1/2)(y_1 + y_2) \Delta y \tan \varphi] \quad (19.47c)$$

$$= (\hat{\mathbf{z}} - \hat{\mathbf{y}} \tan \varphi) \rho g \Delta x [z_2 \Delta y - y_2 \Delta z + (y_1 + y_2) \Delta z / 2] \quad (19.47d)$$

$$= (\hat{\mathbf{z}} - \hat{\mathbf{y}} \tan \varphi) \rho g \Delta x \Delta y (z_1 + z_2) / 2. \quad (19.47e)$$

Bringing these results together renders the net pressure forces in the two directions

$$\hat{\mathbf{y}} \cdot (\mathbf{F}_A^{\text{press}} + \mathbf{F}_B^{\text{press}} + \mathbf{F}_C^{\text{press}}) = (\rho g \Delta x / 2) [(z_1 + z_2) \Delta z - \tan \varphi (z_1 + z_2) \Delta y] = 0 \quad (19.48a)$$

$$\hat{\mathbf{z}} \cdot (\mathbf{F}_A^{\text{press}} + \mathbf{F}_B^{\text{press}} + \mathbf{F}_C^{\text{press}}) = \rho g \Delta x \Delta y \Delta z / 2 = M g, \quad (19.48b)$$

where the mass of the triangle is given by

$$M = \rho \Delta x \Delta y \Delta z / 2. \quad (19.49)$$

We thus see that the area integrated horizontal pressure forces vanish, whereas the area integrated vertical pressure force balances the weight of the fluid. Again, these results are expected given the general expressions (19.38), (19.39a), and (19.39b) of force balance. Even so, being able to explicitly compute the pressure forces acting around a region, and to confirm the general force balances, is a useful means to become familiar with hydrostatic pressure.

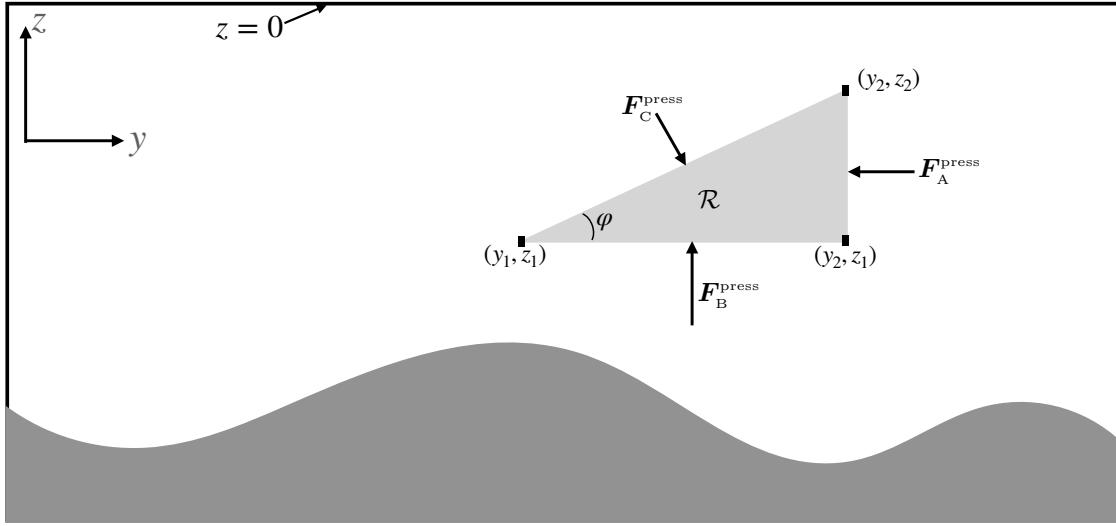


FIGURE 19.5: A right triangle region of fluid in a static ocean where $z < 0$. The positions for the three corners are shown as (y_1, z_1) , (y_2, z_1) , and (y_2, z_2) , along with the pressure forces acting on the three sides. In exact hydrostatic balance, the area integrated pressure force acting over the triangle boundary vanishes, $-\oint_{\partial\mathcal{R}} p \hat{n} d\mathcal{S} = 0$. If the density of the fluid is assumed constant, then we can analytically compute the force balance as detailed in Section 19.4.2.

19.4.3 Torque balance

Torques arise in the presence of force *couplets*, which in turn lead to time changes in the angular momentum. In our discussion of the stress tensor in Section 19.3, we saw that a symmetric stress tensor removes volume sources of torque; i.e., there are no internal sourced force couplets. So the only means to impart a nonzero torque is for force couplets to arise from body forces (forces originating outside of the fluid region) and from contact forces that act between fluid elements within the region. The aim of this section is to show that for a static fluid with a static angular momentum, then the net torque vanishes both at a point and on an arbitrary fluid region.

Strong form: zero torques acting on a fluid element

The torque is the moment of a force computed about a chosen origin. For a static geophysical fluid, the torque acting on a fluid element is given by the moment of the pressure force plus the moment of the effective gravity force (central gravity plus planetary centrifugal)

$$\mathbf{x} \wedge \mathbf{f} \rho \delta V = \mathbf{x} \wedge (-\rho^{-1} \nabla p - \nabla \Phi) \rho \delta V. \quad (19.50)$$

As seen in Section 19.4.1, the pressure and effective gravitational forces exactly balance at each point with a static fluid so that $\nabla p = -\rho \nabla \Phi$. Hence, there can be no torques at each point since there are no net forces at each point.

Weak form: zero torques acting on a finite fluid region

To show that the torque vanishes for a finite fluid region, we can merely integrate the fluid element result (19.50) over the finite region. Since the integral of zero is still zero, there are no torques on the region. An alternative approach makes use of the weak formulation by following the discussion

in Section 19.3.2. In this approach, we start by writing the time change in angular momentum acting on a static fluid region

$$\frac{d}{dt} \int_{\mathcal{R}} \mathbf{L} = \int_{\mathcal{R}} [\mathbf{x} \wedge (-\rho \nabla \Phi)] dV + \oint_{\partial \mathcal{R}} [\mathbf{x} \wedge (-\hat{\mathbf{n}} p)] d\mathcal{S}. \quad (19.51)$$

Note that although the discussion in Section 19.3.2 focused on a Lagrangian region, $\mathcal{R}(\mathbf{v})$, there is no distinction here between Lagrangian and Eulerian since the fluid is static.

The pressure contribution in equation (19.51) is written in its contact force form, which is appropriate for a weak formulation. However, to compare its contribution to the torque with that from effective gravity requires us to convert the area integral to a volume integral. For that purpose we use Cartesian tensor notation and expose full details for clarity

$$\oint_{\partial \mathcal{R}} (\hat{\mathbf{n}} \wedge \mathbf{x})_a p d\mathcal{S} = \epsilon_{abc} \oint_{\partial \mathcal{R}} \hat{n}_b x_c p d\mathcal{S} \quad \text{permutation symbol (Section 1.4.1)} \quad (19.52a)$$

$$= \epsilon_{abc} \int_{\mathcal{R}} \partial_b (x_c p) dV \quad \text{divergence theorem} \quad (19.52b)$$

$$= \epsilon_{abc} \int_{\mathcal{R}} (\delta_{bc} p + x_c \partial_b p) dV \quad \text{product rule} \quad (19.52c)$$

$$= \epsilon_{abc} \int_{\mathcal{R}} x_c \partial_b p dV \quad \epsilon_{abc} \delta_{bc} = 0 \quad (19.52d)$$

$$= -\epsilon_{acb} \int_{\mathcal{R}} x_c \partial_b p dV \quad \epsilon_{abc} = -\epsilon_{acb} \quad (19.52e)$$

$$= - \int_{\mathcal{R}} (\mathbf{x} \wedge \nabla p)_a dV \quad \text{vector cross product notation.} \quad (19.52f)$$

This result then brings the angular momentum equation (19.51) to the form

$$\frac{d}{dt} \int_{\mathcal{R}} \mathbf{L} = \int_{\mathcal{R}} [\mathbf{x} \wedge (-\rho \nabla \Phi - \nabla p)] dV. \quad (19.53)$$

At this point we can invoke the strong form force balance in equation (19.35), thus revealing that the integrand on the right hand side vanishes at each point in the fluid. However, this approach is no different than starting from the strong formulation of the torques in equation (19.50) and integrating over a finite region. An alternative approach, remaining fully within the weak formulation, states that if the region's angular momentum remains constant, then that defines a region experiencing zero net torque. This approach is the same as taken for the force balance, whereby we say that a fluid region experiencing no acceleration is one that has zero net forces acting on it. Hence, for a region with constant angular momentum we are led to the finite volume (weak form) torque balance

$$\int_{\mathcal{R}} [\mathbf{x} \wedge (\rho \nabla \Phi + \nabla p)] dV = 0 \implies \int_{\mathcal{R}} \mathbf{x} \wedge \rho \nabla \Phi dV = - \int_{\partial \mathcal{R}} \mathbf{x} \wedge \hat{\mathbf{n}} p d\mathcal{S}, \quad (19.54)$$

with this balance the direct analog for torques of the weak form of the force balances given by equations (19.38), (19.39a), and (19.39b).

19.5 Flux-form Eulerian momentum equation

We often find it useful to consider Cauchy's form of the momentum equation (19.20) in its flux-form Eulerian expression. Making use of Cartesian tensors, we expand the material time derivative

acting on the velocity and introduce the mass conservation equation (15.6) so that

$$\rho \frac{D\mathbf{v}}{Dt} = \rho [\partial_t \mathbf{v} + (\mathbf{v} \cdot \nabla) \mathbf{v}] \quad (19.55a)$$

$$= \rho [\partial_t \mathbf{v} + (\mathbf{v} \cdot \nabla) \mathbf{v}] + \mathbf{v} (\partial_t \rho + \nabla \cdot (\rho \mathbf{v})) \quad (19.55b)$$

$$= \partial_t (\rho \mathbf{v}) + \nabla \cdot [\rho \mathbf{v} \otimes \mathbf{v}], \quad (19.55c)$$

where $\mathbf{v} \otimes \mathbf{v}$ is the outer product of the velocity vector and it has Cartesian tensor components written as

$$(\mathbf{v} \otimes \mathbf{v})_{mn} = v_m v_n. \quad (19.56)$$

Consequently, the momentum equation (19.20) takes on the flux-form Eulerian expression

$$\partial_t (\rho \mathbf{v}) + \nabla \cdot [\rho \mathbf{v} \otimes \mathbf{v}] = \rho \mathbf{f} + \nabla \cdot \mathbb{T}. \quad (19.57)$$

Alternatively, we can move the advection of momentum term onto the right hand side so that

$$\partial_t (\rho \mathbf{v}) = \rho \mathbf{f} + \nabla \cdot [\mathbb{T} - \rho \mathbf{v} \otimes \mathbf{v}], \quad (19.58)$$

which takes on the component form

$$\partial_t (\rho v_m) = \rho f_m + \partial_n [\mathbb{T}_{mn} - \rho v_m v_n]. \quad (19.59)$$

In this form we see that momentum advection can be interpreted as a stress that modifies the linear momentum per volume at a point in space. We refer to the stress,

$$\mathbb{T}_{mn}^{\text{kinetic}} = -\rho (\mathbf{v} \otimes \mathbf{v})_{mn} = -\rho v_m v_n, \quad (19.60)$$

as the *mechanical stress* or *kinetic stress*, which arises from the mechanical interactions between moving fluid elements. The turbulent contribution to the mechanical stress is known as the *Reynolds stress*.

For a rotating fluid in a gravity field we set the body force to

$$\rho \mathbf{f} = -2 \rho \boldsymbol{\Omega} \wedge \mathbf{v} - \rho \nabla \Phi, \quad (19.61)$$

and the divergence of the stress tensor is

$$\nabla \cdot \mathbb{T} = -\nabla p + \rho \mathbf{F}, \quad (19.62)$$

with \mathbf{F} the frictional acceleration (see Section 19.7). In this case, Cauchy's equation (19.57) takes on the form

$$\frac{\partial(\rho \mathbf{v})}{\partial t} + \nabla \cdot (\rho \mathbf{v} \otimes \mathbf{v}) + 2 \rho \boldsymbol{\Omega} \wedge \mathbf{v} = -\nabla p - \rho \nabla \Phi + \rho \mathbf{F}. \quad (19.63)$$

19.6 Linear momentum for arbitrary regions

Consider the budget of linear momentum for an arbitrary region, \mathcal{R} , moving in an arbitrary manner within the fluid. For this purpose we make use of the Leibniz-Reynolds Transport Theorem (16.34)

$$\frac{d}{dt} \left[\int_{\mathcal{R}} \varphi dV \right] = \int_{\mathcal{R}} \frac{\partial \varphi}{\partial t} dV + \oint_{\partial \mathcal{R}} \varphi \mathbf{v}^{(b)} \cdot \hat{\mathbf{n}} d\mathcal{S}, \quad (19.64)$$

where $\mathbf{v}^{(b)}$ is the velocity of the region boundary, $\partial\mathcal{R}$, with $\hat{\mathbf{n}}$ the outward normal along the boundary. Applying this result to a component of the linear momentum per volume, $\varphi = \rho v_m$, and making use of the flux-form Eulerian momentum equation (19.59) leads to

$$\frac{d}{dt} \left[\int_{\mathcal{R}} \rho v_m dV \right] = \int_{\mathcal{R}} \partial_t (\rho v_m) dV + \oint_{\partial\mathcal{R}} (\rho v_m) \mathbf{v}^{(b)} \cdot \hat{\mathbf{n}} d\mathcal{S} \quad (19.65a)$$

$$= \int_{\mathcal{R}} [\rho f_m + \partial_n (\mathbb{T}_{mn} - \rho v_m v_n)] dV + \oint_{\partial\mathcal{R}} (\rho v_m) \mathbf{v}^{(b)} \cdot \hat{\mathbf{n}} d\mathcal{S} \quad (19.65b)$$

$$= \int_{\mathcal{R}} \rho f_m dV + \oint_{\partial\mathcal{R}} (\mathbb{T}_{mn} - \rho v_m v_n) \hat{n}_n d\mathcal{S} + \oint_{\partial\mathcal{R}} (\rho v_m) \mathbf{v}^{(b)} \cdot \hat{\mathbf{n}} d\mathcal{S} \quad (19.65c)$$

$$= \int_{\mathcal{R}} \rho f_m dV + \oint_{\partial\mathcal{R}} [\mathbb{T}_{mn} + \rho v_m (v_n^{(b)} - v_n)] \hat{n}_n d\mathcal{S}. \quad (19.65d)$$

We can write this relation in the more elegant geometric form as

$$\frac{d}{dt} \left[\int_{\mathcal{R}} \rho \mathbf{v} dV \right] = \int_{\mathcal{R}} \rho \mathbf{f} dV + \oint_{\partial\mathcal{R}} [\mathbb{T} + \rho \mathbf{v} \otimes (\mathbf{v}^{(b)} - \mathbf{v})] \cdot \hat{\mathbf{n}} d\mathcal{S}. \quad (19.66)$$

As a relation between geometric objects (vectors and tensors), the momentum budget (19.66) is independent of coordinate representation. We conclude that the evolution of linear momentum over an arbitrary region is affected by the volume integrated body force acting over the region, plus the impacts from stresses acting on the region boundary. Notably, the stresses have a contribution from the advection of linear momentum across the region boundary, with advection computed relative to motion of the boundary. In Section 19.9 we specialize the budget (19.66) to an infinitesimally thin interface. That analysis is then used to develop stress conditions for a surface within a single fluid media, and the stress condition at the boundary between two fluids.

We refer to a *Lagrangian region* as one that moves with the barycentric velocity so that $\mathbf{v}^{(b)} = \mathbf{v}$, in which case the mechanical stress is eliminated from the finite volume momentum budget (19.66). In fact, to eliminate the mechanical stress only requires the normal components of the boundary velocity to equal that of a fluid element, $(\mathbf{v} - \mathbf{v}^{(b)}) \cdot \hat{\mathbf{n}} = 0$. In either case we distinguish a Lagrangian region by writing $\mathcal{R}(\mathbf{v})$ to emphasize that the region moves with the barycentric fluid velocity, \mathbf{v} . For this case the linear momentum is only affected by body forces as well as stresses contained in the stress tensor

$$(\mathbf{v} - \mathbf{v}^{(b)}) \cdot \hat{\mathbf{n}} = 0 \implies \frac{d}{dt} \left[\int_{\mathcal{R}(\mathbf{v})} \rho \mathbf{v} dV \right] = \int_{\mathcal{R}(\mathbf{v})} \rho \mathbf{f} dV + \oint_{\partial\mathcal{R}(\mathbf{v})} \mathbb{T} \cdot \hat{\mathbf{n}} d\mathcal{S}. \quad (19.67)$$

This relation is Reynold's transport theorem (Section 16.3.5) as applied to linear momentum.

19.7 Constitutive relation between stress and rate of strain

Thus far we have offered a rather general treatment of stress, developing its properties according to the conservation of linear momentum and angular momentum. We now develop *constitutive relations*, which relate stress to properties of the fluid as well as properties of the fluid flow.

19.7.1 Thermodynamic pressure and mechanical pressure

Consider a fluid in which the stress on an area element is always normal to the area element and is independent of the orientation. This fluid is in hydrostatic balance and the corresponding stress

tensor and stress vector are written

$$\mathbb{T}_{mn} = -p \delta_{mn} \iff \mathbb{T} \cdot \hat{\mathbf{n}} = -p \hat{\mathbf{n}}, \quad (19.68)$$

where p is the hydrostatic pressure field. Since the pressure introduced here arises from purely mechanical considerations, we refer to it as the *mechanical pressure*. For a compressible fluid at rest, we can identify the mechanical pressure with the *thermodynamic pressure* encountered in our study of equilibrium thermodynamics (Chapter 21). Furthermore, if we assume that local thermodynamic equilibrium is maintained for fluid elements within a moving fluid, then we are motivated to continue making this identification between mechanical pressure and thermodynamical pressure (see Section 4.5 of [Kundu et al. \(2016\)](#) or Section 1.10 of [Salmon \(1998\)](#)). However, we note that there is no fully deductive theory supporting this equality of pressures. The reader in search of a deductive theory likely will need to penetrate nonequilibrium statistical mechanics, which is outside our scope.

When the fluid flow is incompressible (Chapter 17), we lose the equality between the mechanical pressure and thermodynamical pressure, even when the fluid is at rest. The reason is that an incompressible flow is unable to do pressure work on a fluid element since the fluid element cannot change its volume. Hence, for incompressible flow there is no connection between pressure and changes to internal energy as per the first law of thermodynamics (Section 21.2). An incompressible flow only has access to the mechanical pressure as revealed through the measurement of stresses. Furthermore, the mechanical pressure instantaneously conforms to the needs of incompressibility throughout the fluid (see Section 24.3 and Section 34.5). Correspondingly, energetic consistency requires us to make use of the geopotential in the equation of state for *in situ* density in a Boussinesq fluid (in which the velocity is non-divergent), rather than the thermodynamic pressure. We explore this point in Section 24.7.

19.7.2 Couette flow and the viscous tensor

Couette flow arises when fluid is placed between two long and straight concentric cylinders that can rotate. Relative motion between the two cylinders leads to fluid motion. For example, if the inner surface rotates, then fluid next to the cylinder wall will move with the cylinder. Any normal stresses on the fluid imparted by the cylinders are directed toward the cylinder axis and so cannot render any tangential motion. This elegant experiment proves that fluid motion can be induced by purely tangential stresses. Furthermore, the tangential stress imparted by the inner cylinder transfers through the fluid to the outer cylinder. Indeed, if the inner cylinder rotates at a constant rate, then eventually the whole fluid-cylinder system rotates as a solid body. Couette flow thus exhibits how real fluids can support tangential stresses in response to tangential strains, thus providing a clear distinction from a perfect fluid where only normal stresses (i.e., pressure) are supported.

As evidenced by the Couette flow, a moving fluid has a more complex stress relation than a static fluid. In particular, the presence of tangential stresses in the fluid provides evidence for an additional piece to the stress tensor that we write as

$$\mathbb{T}_{mn} = -p \delta_{mn} + \tau_{mn} \iff \mathbb{T} = -p \mathbb{I} + \tau. \quad (19.69)$$

The pressure term remains isotropic as for a fluid at rest, thus imparting normal stresses. The additional tensor, τ , is referred to as the *friction tensor* or *viscous tensor*. The viscous tensor captures the irreversible exchanges of momentum between moving (relative to one another) fluid elements, such as in Couette flow, with the irreversible momentum exchange supported by fluid viscosity. Viscosity is assumed to be identically zero in a perfect fluid, so that a perfect fluid can only support normal stresses from pressure even when the perfect fluid has relative motion.

As noted above, we assume the viscous tensor vanishes when there is zero relative motion within the fluid.³ The physical idea is that fluid strains are needed to generate friction between fluid elements to support the transfer of momentum through the presence of viscosity. The determination of frictional stresses from kinematic properties (such as strain) requires a constitutive relation. The constitutive relation commonly used for geophysical fluids follows that for a *Newtonian fluid*, which is a particular type of *Stokesian fluid* whose frictional stresses are assumed to be linearly proportional to the rate of strain.

The diagonal stresses, \mathbb{T}_{11} , \mathbb{T}_{22} , and \mathbb{T}_{33} , are known as the *direct stresses* or *normal stresses*, whereas the off-diagonal stresses are *shear stresses*. The sum of the direct stresses forms the trace of the stress tensor and is given by

$$\mathbb{T}_{qq} = \mathbb{T}_{11} + \mathbb{T}_{22} + \mathbb{T}_{33} = -3p + \tau_{qq}. \quad (19.70)$$

If $\tau_{qq} = 0$ then it is known as the *deviatoric friction tensor*. As argued in Section 19.7.5, a deviatoric friction tensor is consistent with the assumption of equal mechanical and thermodynamical pressures.

19.7.3 Guidance from Galilean invariance

Consider a fluid in uniform motion in free space. Boosting the reference frame allows us to move to a reference frame where the fluid is static. Through Galilean invariance (Section 13.6) we expect the dynamics to remain unchanged. Since we assume friction vanishes when the fluid is static (as per a Stokesian fluid), Galilean invariance implies that the frictional stresses vanish when the fluid undergoes uniform motion in any direction.

Uniform motion of fluid elements is reflected in zero velocity gradients, which offers a key insight into how friction depends on strains. Namely, these considerations suggest that the friction tensor is a function of gradients in the velocity field, $\partial_m v_n$. Furthermore, as the stress tensor must be symmetric, the simplest expression for the friction tensor is one that is linearly proportional to the rate of strain tensor introduced in Section 14.2.5. Fluids that satisfy this linear stress-rate of strain constitutive relation are known as *Newtonian fluids*. Furthermore, this constitutive relation takes the same mathematical form as *Hooke's Law* used in the study of elastic materials.

19.7.4 A comment on Rayleigh drag

Rayleigh drag is a particular form of friction that makes use of the acceleration

$$\mathbf{F}_{\text{Rayleigh}} = -\gamma \mathbf{v}, \quad (19.71)$$

where $\gamma > 0$ is an inverse time scale. Rayleigh drag does *not* satisfy the Galilean principle since it decelerates all flows, even uniformly moving flows, towards rest where rest is defined by the laboratory frame. Furthermore, Rayleigh drag is not equal to the divergence of a frictional stress tensor, and so it does not arise from a contact stress. Even so, it has found some use for rudimentary purposes, particularly when aiming to derive analytic expressions for how friction acts on flows in a bulk sense. We provide one example in our discussion of Ekman dynamics in Section 28.2.3.

19.7.5 Constitutive relation for Newtonian fluids

There are many details involved with deriving the Newtonian fluid constitutive relation, with discussions provided in Chapter 5 of [Aris \(1962\)](#) and Section 4.5 of [Kundu et al. \(2016\)](#) for general

³This assumption follows those for a Stokes fluid as discussed in Section 5.21 of [Aris \(1962\)](#).

fluids, and Chapter 17 and 18 of [Griffies \(2004\)](#) for stratified fluids with particular focus on the ocean. We here offer a taste of these considerations by starting with the constitutive relation

$$\tau_{mn} = \rho (2\nu \mathbb{S}_{mn} + \lambda \nabla \cdot \mathbf{v} \delta_{mn}), \quad (19.72)$$

with \mathbb{S}_{mn} the components to the rate of strain tensor introduced in Section 14.2.7 and whose trace equals to the velocity divergence

$$\mathbb{S}_{qq} = \nabla \cdot \mathbf{v}. \quad (19.73)$$

The first contribution to the frictional stress (19.72) includes the strain tensor multiplied by the *first kinematic viscosity*, $\nu > 0$ (dimensions of squared length per time). The second contribution arises just from flow divergence as scaled by a *second kinematic viscosity*, λ . The sum

$$\nu_{\text{bulk}} = \rho (\lambda + 2\nu/3) \quad (19.74)$$

is known as the *bulk viscosity*, which, as discussed in the following, will be set to zero. Finally, one sometimes finds it more convenient to work with the *dynamic viscosity*

$$\mu_{\text{vsc}} = \rho \nu. \quad (19.75)$$

Note that we set ρ to the constant Boussinesq reference density, ρ_0 (Chapter 24), when working with a Boussinesq fluid.

Deviatoric friction tensor

As noted in Section 19.7.1, the pressure appearing in the stress tensor is a mechanical pressure that equals to minus one-third the trace of the stress tensor when the fluid is at rest

$$\mathbb{T}_{qq}^{\text{static}} = -3p. \quad (19.76)$$

We assume that the friction stress tensor does *not* alter this trace, so that the friction tensor has zero trace and is known as a *deviatoric friction tensor*⁴

$$\tau_{qq} = 0 = 3\nu_{\text{bulk}} \nabla \cdot \mathbf{v} \implies \lambda = -2\nu/3, \quad (19.77)$$

so that the total stress tensor is given by

$$\mathbb{T}_{mn} = -\delta_{mn} p + 2\mu_{\text{vsc}} \mathbb{S}_{mn}^{\text{dev}} \quad \text{with} \quad \mathbb{S}_{mn}^{\text{dev}} = \mathbb{S}_{mn} - \delta_{mn} \mathbb{S}_{qq}/3, \quad (19.78)$$

where \mathbb{S}^{dev} is the *deviatoric rate of strain tensor*. We next offer arguments for why the friction tensor used for geophysical flows should have zero trace.

Equality of the mechanical and thermodynamic pressures

The frictional stress tensor (19.72) is not the precise form typically used in geophysical fluid modeling. Instead, the velocity divergence term is generally dropped even for compressible flows, and the viscosity is anisotropic and more generally takes the form of a fourth order tensor (see Chapter 17 and 18 of [Griffies \(2004\)](#) for the ocean). Furthermore, what is generally respected in most geophysical applications is the deviatoric nature of the friction tensor. That property is maintained

⁴A second order tensor in 3-dimensions, \mathbb{D} , has a *deviator* with components given by $\mathbb{D}_{mn}^{\text{dev}} = \mathbb{D}_{mn} - (1/3) \delta_{mn} \mathbb{D}_{qq}$. By construction, the trace of the deviator vanishes: $\mathbb{D}_{qq}^{\text{dev}} = 0$.

since it is consistent with our assumption in Section 19.7.1 that the mechanical pressure equals to the thermodynamic pressure.

To see this equality between the pressures, introduce the mechanical pressure, p^{mech} , according to the trace of the stress tensor

$$\mathbb{T}_{qq} = -3 p^{\text{mech}}. \quad (19.79)$$

That is, mechanical pressure is minus one-third the trace of the stress tensor whether the fluid is at rest or in motion. We can, in principle, measure this pressure by measuring the stresses. If we now return to the general form of the stress tensor

$$\mathbb{T}_{mn} = -\delta_{mn} p + \rho (\lambda \nabla \cdot \mathbf{v} \delta_{mn} + 2 \nu \mathbb{S}_{mn}), \quad (19.80)$$

with p here given by the thermodynamic pressure, then the trace is

$$\mathbb{T}_{qq} = -3 p + \rho (3 \lambda + 2 \nu) \nabla \cdot \mathbf{v}. \quad (19.81)$$

Setting the two traces (19.79) and (19.81) equal then leads to

$$p - p^{\text{mech}} = \nu_{\text{bulk}} \nabla \cdot \mathbf{v}. \quad (19.82)$$

In principle, one can measure the difference in pressures, but in practice this measurement is rather difficult. Stokes assumed $p - p^{\text{mech}} = 0$ by taking a zero bulk viscosity, and he used arguments from kinetic theory of gases to support that choice. This choice is generally taken for geophysical fluid applications, largely based on the assumption of local thermodynamic equilibrium mentioned in Section 19.7.1.

Local thermodynamical equilibrium is not a good assumption in supersonic flows, in which case $p - p^{\text{mech}} \neq 0$ and the bulk viscosity is nonzero. Correspondingly, the divergence term and the second kinematic viscosity, $\lambda \neq -2\nu/3$, are important. Additionally, the second viscosity is important when concerned with the damping of acoustic waves. Neither topics are considered in this book so that we have no further concern for the second viscosity.

Frictional force per volume

The viscous force per volume is given by the divergence of the frictional stress tensor

$$\rho F_n = \partial_m \mathbb{T}_{mn} = 2 \partial_m (\mu_{\text{vsc}} \mathbb{S}_{mn}^{\text{dev}}). \quad (19.83)$$

For a Boussinesq fluid this friction simplifies to

$$\rho_0 F_n = 2 \rho_0 \partial_m (\nu \mathbb{S}_{mn}), \quad (19.84)$$

where we set the dynamic viscosity to $\mu_{\text{vsc}} = \rho_0 \nu$, with ρ_0 the constant Boussinesq reference density. Finally, for the case of a constant kinematic viscosity, such as for molecular viscous effects, we have the Boussinesq result reducing to the Laplacian form

$$\mathbf{F} = \nu \nabla^2 \mathbf{v}. \quad (19.85)$$

This form is commonly used in scale analysis, such as when determining the Reynolds number in Section 19.8.

19.7.6 Navier-Stokes and Euler equations

The *Navier-Stokes* equation is a special form of the momentum equation found by assuming a Newtonian fluid constitutive relation. In this case the Navier-Stokes momentum equation (18.16), in the presence of rotation and gravity, takes on the form

$$\rho \frac{D\mathbf{v}}{Dt} + 2\rho \boldsymbol{\Omega} \wedge \mathbf{v} = -\rho \nabla \Phi - \nabla p + \nabla \cdot (2\mu_{vsc} \mathbb{S}^{dev}), \quad (19.86)$$

where we set the friction tensor equal to $\tau = 2\mu_{vsc} \mathbb{S}^{dev}$. Quite often when examining the mathematical properties of the Navier-Stokes equation, one assumes the flow to be incompressible, in which case $\mathbb{S}^{dev} = \mathbb{S}$ since $\mathbb{S}_{qq} = \nabla \cdot \mathbf{v} = 0$. A further simplification occurs by assuming a constant density, ρ , and constant kinematic viscosity, ν , in which case the friction tensor reduces to the Laplacian form 19.85 so that the Navier-Stokes equation becomes

$$\frac{D\mathbf{v}}{Dt} + 2\boldsymbol{\Omega} \wedge \mathbf{v} = -\nabla \Phi - \rho^{-1} \nabla p + \nu \nabla^2 \mathbf{v}. \quad (19.87)$$

This form, or even simpler when ignoring rotation and gravity, is commonly studied by mathematicians concerned with existence and uniqueness properties of solutions (e.g., see [Doering and Gibbon \(2004\)](#)). When assuming the fluid to be perfect, so that there are no viscous forces, the momentum equation is referred to as the *Euler equation*⁵

$$\frac{D\mathbf{v}}{Dt} + 2\boldsymbol{\Omega} \wedge \mathbf{v} = -\nabla \Phi - \rho^{-1} \nabla p. \quad (19.88)$$

It is tempting to consider the Euler equations to be a continuous limit of the Navier-Stokes equation as the viscosity goes to zero. However, there is a key distinction between the two equations. Namely, the Navier-Stokes equations admit solutions that display statistically equilibrated turbulent motions, whereby energy cascades to the small scales through vortex stretching in three dimensional flows. This energy is ultimately dissipated by viscosity at the small spatial scales, and this mechanism holds no matter how small the viscosity, so long as it is nonzero. In contrast, for the Euler equations, with identically zero viscosity, energy cannot be dissipated at the small scales so that an equilibrium turbulent flow is not available.

19.7.7 Laplacian friction in terms of vorticity and divergence

The Laplacian friction operator with a constant viscosity is afforded the following decomposition

$$\nu^{-1} F_n = \partial_m (\partial_m v_n) \quad (19.89a)$$

$$= \partial_m (\partial_m v_n - \partial_n v_m + \partial_n v_m) \quad (19.89b)$$

$$= \partial_m (\partial_m v_n - \partial_n v_m) + \partial_n (\partial_m v_m) \quad (19.89c)$$

$$= -2 \partial_m \mathbb{A}_{mn} + \partial_n \nabla \cdot \mathbf{v} \quad (19.89d)$$

$$= \epsilon_{nmp} \partial_m \omega_p + \partial_n \nabla \cdot \mathbf{v} \quad (19.89e)$$

$$= (\nabla \wedge \boldsymbol{\omega})_n + \partial_n \nabla \cdot \mathbf{v}. \quad (19.89f)$$

In the fourth equality we introduced the rotation tensor (14.42)

$$\mathbb{A}_{mn} = (1/2) (\partial_n v_m - \partial_m v_n), \quad (19.90)$$

⁵Note that some authors refer to the Euler equations only in the case of a perfect fluid that has no body force, so that both rotation and gravitation vanish.

which is related to the vorticity, $\boldsymbol{\omega} = \nabla \wedge \mathbf{v}$, via equation (14.44)

$$\mathbb{A}_{mn} = -\epsilon_{mnp} \omega_p / 2. \quad (19.91)$$

These manipulations have served to decompose the Laplacian viscous acceleration, with a constant viscosity, into the two terms

$$\mathbf{F} = \nu [\nabla \wedge \boldsymbol{\omega} + \nabla(\nabla \cdot \mathbf{v})]. \quad (19.92)$$

The Laplacian friction acceleration is thus due to the curl of the vorticity plus gradients in the velocity divergence. Many geophysical flows are dominated by vorticity since the divergence is quite small. Indeed, the Boussinesq fluids discussed in Chapter 24 have $\nabla \cdot \mathbf{v} = 0$, in which case frictional acceleration arises only when the vorticity has a nonzero curl. Correspondingly, irrotational flows (where $\boldsymbol{\omega} = 0$) that are also non-divergent have zero Laplacian frictional acceleration.

19.7.8 Frictional stresses in a sheared flow

As a means to connect the above ideas in this section to the Couette flow discussed in Section 19.7.2, consider an incompressible velocity that only has a zonal component with a vertical shear (Figure 19.6)

$$\mathbf{v} = u(z) \hat{x}. \quad (19.93)$$

In this case the only nonzero components to the rate of strain tensor are due to the vertical shear, $\mathbb{S}_{13} = \mathbb{S}_{31} = \partial_z u / 2$. Now consider a horizontal area whose outward normal is parallel to the \hat{z} direction. The frictional force acting on that area is given by the area integral of the frictional stress

$$\mathbf{F}_{\text{area}} = \int \tau \cdot \hat{\mathbf{n}} \, d\mathcal{S} = \int \tau \cdot \hat{z} \, dx \, dy = \rho_0 \frac{\hat{x}}{2} \nu A \frac{\partial u}{\partial z}, \quad (19.94)$$

where $A = \int dx \, dy$ is the horizontal area, and where we used the constant reference density, ρ_0 , for a Boussinesq fluid. Hence, the zonal stress arises from the nonzero vertical shear.

Momentum is deposited in regions where there is a divergence in the stress, in which case momentum is transferred from regions of high vertical shear to low vertical shear. At a point, the momentum is affected by the divergence of the viscous stress at that point. For $\mathbf{v} = u(z) \hat{x}$ we have

$$\left[\frac{\partial(\rho v_m)}{\partial t} \right]_{\text{viscous}} = \partial_n \tau_{nm} \implies \left[\frac{\partial(\rho u)}{\partial t} \right]_{\text{viscous}} = \partial_z (\mu_{\text{vsc}} \partial_z u), \quad (19.95)$$

so that zonal momentum is preferentially deposited to or removed from regions with high vertical curvature in the zonal velocity. Spatial variations in the dynamic viscosity, $\mu_{\text{vsc}} = \rho_0 \nu$, also contribute to friction.

19.7.9 The net stress tensor

Combining the frictional stress tensor with pressure and kinetic stress yields the flux-form momentum equation (19.63)

$$\frac{\partial(\rho \mathbf{v})}{\partial t} + 2\rho \boldsymbol{\Omega} \wedge \mathbf{v} + \rho \nabla \Phi = \nabla \cdot \mathbb{T}^{\text{net}}, \quad (19.96)$$

where we introduced the net stress tensor

$$\mathbb{T}^{\text{net}} = -p \mathbb{I} - \rho \mathbf{v} \otimes \mathbf{v} + \tau = \begin{bmatrix} -p - \rho u^2 + \tau_{11} & -\rho u v + \tau_{12} & -\rho u w + \tau_{13} \\ -\rho u v + \tau_{12} & -p - \rho v^2 + \tau_{22} & -\rho v w + \tau_{23} \\ -\rho u w + \tau_{13} & -\rho v w + \tau_{23} & -p - \rho w^2 + \tau_{33} \end{bmatrix}. \quad (19.97)$$

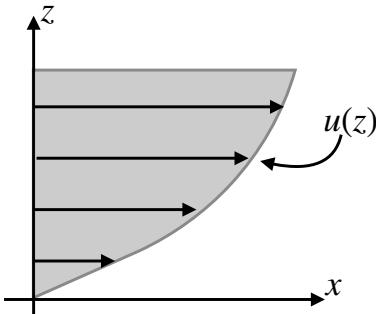


FIGURE 19.6: Sample profile of zonal velocity possessing a vertical shear: $\mathbf{v} = u(z) \hat{x}$ and with a no-slip boundary condition at $z = 0$. The resulting zonal frictional stress arises from the nonzero vertical shear and viscosity.

The left hand side of the momentum equation (19.96) includes the local time tendency plus the body forces from Coriolis and effective gravity. The right hand side is the divergence of the net stress tensor, with this tensor combining the pressure stress, kinetic stress, and frictional stress. Varieties of the net stress tensor appear in subsequent chapters of this book, with details dependent on the chosen approximations.

19.7.10 Comments and further study

There are more elaborate relations between the frictional stress and rate of strain tensors than those considered here. The most general form for a Newtonian fluid introduces a fourth-order viscosity tensor as in Section 4.5 of [Kundu et al. \(2016\)](#) and Chapter 17 in [Griffies \(2004\)](#). We also suggest the treatment of stress in Chapter 5 of [Aris \(1962\)](#).

Geophysical fluids such as air and water are generally well treated using Newtonian constitutive relations. However, there are some geophysical turbulence theories that propose a non-Newtonian constitutive relation for part of their closures, whereby the constitutive relation makes use of products of the rate of strain tensor for computing stress. [Anstey and Zanna \(2017\)](#) offer a compelling approach with a subgrid scale stress tensor that is non-Newtonian and furthermore contains a non-zero trace, thus resulting in a modification to the mechanical pressure. Additional nonlinear relations can arise when the viscous tensor is a function of the flow, such as with the Smagorinsky scheme commonly used for Large Eddy Simulations (LES) (see [Smagorinsky \(1993\)](#) or Chapter 18 of [Griffies \(2004\)](#)).

19.8 Reynolds number and flow regimes

How important is friction relative to other terms in the momentum equation? In particular, how does it compare to the material acceleration? We consider that question in the context of the non-rotating and constant density Navier-Stokes equations with a constant viscosity

$$\frac{\partial \mathbf{v}}{\partial t} + (\mathbf{v} \cdot \nabla) \mathbf{v} = -\rho^{-1} \nabla p + \nu \nabla^2 \mathbf{v}, \quad (19.98)$$

with ρ the constant density.

19.8.1 Non-dimensional Navier-Stokes

We non-dimensionalize the Navier-Stokes equation (19.98) to garner an understanding of relative magnitudes of the various terms. For that purpose, introduce the dimensional scales and corresponding non-dimensional fields

$$L = \text{length scale} \quad U = \text{velocity scale} \quad P = \text{pressure scale} \quad T = \text{time scale}, \quad (19.99)$$

so that equation (19.98) takes the form

$$\frac{\partial \hat{\mathbf{v}}}{\partial \hat{t}} + \frac{V T}{L} (\hat{\mathbf{v}} \cdot \hat{\nabla}) \hat{\mathbf{v}} = -\frac{T P}{\rho V L} \hat{\nabla} \hat{p} + \frac{T \nu}{L^2} \hat{\nabla}^2 \hat{\mathbf{v}}, \quad (19.100)$$

where the hat fields are non-dimensional and defined according to

$$\nabla = L \hat{\nabla} \quad \partial_t = T^{-1} \partial_{\hat{t}} \quad \mathbf{v} = U \hat{\mathbf{v}} \quad p = P \hat{p} \quad (19.101)$$

19.8.2 Ratio of inertial to frictional accelerations

We are concerned with three dimensional flows with only a single length and velocity scale, L and U . For the time scale we assume that it is determined by fluid particle time scale; i.e., the advective time

$$T \sim L/U. \quad (19.102)$$

Furthermore, we assume that the scale of mechanical pressure is comparable to the stress induced by the flow itself

$$P \sim \rho U^2. \quad (19.103)$$

These assumed scales for time and pressure bring the non-dimensional Navier-Stokes equation (19.100) into the rather tidy form

$$\frac{\partial \hat{\mathbf{v}}}{\partial \hat{t}} + (\hat{\mathbf{v}} \cdot \hat{\nabla}) \hat{\mathbf{v}} = -\hat{\nabla} \hat{p} + \frac{1}{\text{Re}} \hat{\nabla}^2 \hat{\mathbf{v}}. \quad (19.104)$$

Flow regimes of the non-dimensional Navier-Stokes equation are specified by the non-dimensional number, $\text{Re} = L U / \nu$, which is the *Reynolds number*. By definition, the Reynolds number the ratio of scales for material (inertial) acceleration to frictional acceleration

$$\text{Re} = \frac{\text{inertial accelerations}}{\text{frictional accelerations}} = \frac{U/T}{\nu U/L^2} = \frac{L^2/T}{\nu} = \frac{L U}{\nu}. \quad (19.105)$$

19.8.3 Reynolds numbers for some example flows

Laboratory experiments with flow around and within various objects indicates the following regimes of flow as a function of the Reynolds number:

$$\text{Re} \sim \begin{cases} \leq 10^2 & \text{laminar} \\ 10^2 - 10^3 & \text{quasi-periodic flow} \\ 10^3 - 10^4 & \text{transition to turbulence} \\ \geq 10^4 & \text{fully turbulent.} \end{cases} \quad (19.106)$$

These numbers are fuzzy given dependence on the geometry of the objects placed in the flow and their characteristic length scale. What is more general concerns the behavior of the flow, with a

transition from laminar to turbulent typically occurring as the flow moves from relatively low to high Reynolds number.

For a given molecular kinematic viscosity, the Reynolds number is dependent on the velocity and length scales. Let us consider some examples. First, put your finger into a flowing stream of water, such as in a gentle mountain creek. Let the length scale for the finger be 10^{-2} m and the stream flow at a speed of $U \approx 0.1 - 1$ m s $^{-1}$. With the kinematic viscosity of water given by (page 75 of [Gill \(1982\)](#))

$$\nu_{\text{water}} = 10^{-6} \text{ m}^2 \text{ s}^{-1}, \quad (19.107)$$

we see that our finger poking into the mountain stream is associated with a flow Reynolds number on the order of

$$\text{Re}_{\text{finger in stream}} = 10^3 - 10^4. \quad (19.108)$$

Hence, mountain stream flow around a finger is at the lower end of the turbulent regime. Hence, we might expect to see slightly turbulent whirls and eddies downstream from our finger.

Now consider an oceanographic length scale given by a Gulf Stream ring (see Figure 26.1) in which $L \approx 10^5$ m. Assuming the flow speed is on the same order as the mountain stream (good assumption) leads to a huge Reynolds number for Gulf Stream flow

$$\text{Re}_{\text{Gulf Stream}} = 10^{10} - 10^{11}. \quad (19.109)$$

For the atmosphere, we take $L = 10^6$ m for a typical atmospheric weather system, $U = 10$ m s $^{-1}$ for the speed, and

$$\nu_{\text{air}} = 1.4 \times 10^{-5} \text{ m}^2 \text{ s}^{-1}, \quad (19.110)$$

for the kinematic viscosity of air at standard pressure (page 75 of [Gill \(1982\)](#)). Given the larger length and velocity scales, the Reynolds number for large-scale atmospheric circulation features is

$$\text{Re}_{\text{weather system}} = 10^{12}. \quad (19.111)$$

These values for the Reynolds number are huge relative to typical values found in engineering flows. They signal the minor role that molecular friction plays in large-scale geophysical fluid flows. Even so, molecular friction is the process leading to mechanical energy dissipation at the small scales. A fundamental feature of large Reynolds number flow is the presence of turbulent motions. Turbulent flows are highly nonlinear and affect a transfer of mechanical energy across length and time scales. This cascade leads to the dissipation of mechanical energy at the small scales. It is at the small scales that flow curvature can be large enough for the relatively tiny values of molecular viscosity to dissipate the energy, thus preventing an *ultraviolet catastrophe*; i.e., preventing the unbounded pile up of mechanical energy at the smallest scales.⁶

The ocean and atmosphere exhibit a huge variety of turbulent regimes, from the macroturbulence of quasi-geostrophic eddies to the microturbulence of boundary layers. Turbulence is not directly considered in this book. However, certain of its implications are identified in various places given that it is so basic to the ocean and atmosphere flows. [Vallis \(2017\)](#) offers a pedagogical entry point for the physics and maths of geophysical turbulence.

⁶Ultraviolet refers to the high wavenumber end of the flow spectrum, as in the violet part of the visible electromagnetic spectrum.

19.9 Stress on an interface

In this section we study the stress acting on an interface. This analysis applies to an arbitrary surface within a single media as well as for the boundary interface separating a liquid and a gas (air-sea boundary) or between a fluid and a rigid boundary (air-land or ocean-land). We ignore the effects from surface tension discussed in Section 19.10 since we are interested in length scales larger than the capillary-gravity waves where surface tension is important, with equation (19.141) indicating that these wavelengths are roughly $\lambda_{\text{cap-grav}} \approx 17 \text{ cm}$.

19.9.1 General formulation

Formulation of the stress boundary conditions follows from applying the finite volume momentum equation (19.66) to a cylindrical region straddling a moving interface such as that shown in Figure 19.7. The sides of the cylinder have thickness h and the top and bottom have area δS . In the limit that the cylinder thickness goes to zero, the volume integrals in equation (19.66) vanish under the assumption of a smooth velocity field on both sides of the interface as well as smooth body forces. We are thus left with the constraint that the area integrated contact forces must vanish when integrated around the cylinder boundary

$$\oint_{\partial\text{cylinder}} [\mathbb{T} + \rho \mathbf{v} \otimes (\mathbf{v}^{(b)} - \mathbf{v})] \cdot \hat{\mathbf{n}} dS = \oint_{\partial\text{cylinder}} [-p \mathbb{I} + \tau + \rho \mathbf{v} \otimes (\mathbf{v}^{(b)} - \mathbf{v})] \cdot \hat{\mathbf{n}} dS = 0. \quad (19.112)$$

The end-caps on the cylinder vanish as $h \rightarrow 0$, in which case we have no constraint based on the stresses acting on the end-caps. Instead, the $h \rightarrow 0$ limit leads us to conclude that the contact force on one side of the interface is equal and opposite to that on the other side. This condition is a direct statement of the Newton's third law as manifest via the local equilibrium of stresses discussed in Section 19.1.2. For the stresses acting on the interface in Figure 19.7 we have

$$[-p_A \mathbb{I} + \tau_A + \rho_A \mathbf{v}_A \otimes (\mathbf{v}^{(b)} - \mathbf{v}_A)] \cdot \hat{\mathbf{n}}_A + [-p_B \mathbb{I} + \tau_B + \rho_B \mathbf{v}_B \otimes (\mathbf{v}^{(b)} - \mathbf{v}_B)] \cdot \hat{\mathbf{n}}_B = 0. \quad (19.113)$$

Setting $\hat{\mathbf{n}} = \hat{\mathbf{n}}_B = -\hat{\mathbf{n}}_A$ leads to

$$[-p_A \mathbb{I} + \tau_A + \rho_A \mathbf{v}_A \otimes (\mathbf{v}^{(b)} - \mathbf{v}_A)] \cdot \hat{\mathbf{n}} = [-p_B \mathbb{I} + \tau_B + \rho_B \mathbf{v}_B \otimes (\mathbf{v}^{(b)} - \mathbf{v}_B)] \cdot \hat{\mathbf{n}}, \quad (19.114)$$

which is an expanded expression of Newton's third law given in equation (19.5). Recall that we are ignoring surface tension, which means there is no pressure jump across the interface (see Section 19.10). Hence, setting $p_A = p_B$ allows us to cancel pressure thus leaving an interface stress condition involving just the frictional stress and kinetic stress

$$[\tau_A + \rho_A \mathbf{v}_A \otimes (\mathbf{v}^{(b)} - \mathbf{v}_A)] \cdot \hat{\mathbf{n}} = [\tau_B + \rho_B \mathbf{v}_B \otimes (\mathbf{v}^{(b)} - \mathbf{v}_B)] \cdot \hat{\mathbf{n}}. \quad (19.115)$$

We now consider a variety of examples to illustrate this condition.

19.9.2 Solid material boundary

Consider a solid material boundary through which no matter crosses. Let region B be the solid side of the interface and region A the fluid side (region A is either the ocean or atmosphere). The material nature of the boundary means that no matter crosses it, in which case (see Section 15.4.2)

$$(\mathbf{v}_A - \mathbf{v}^{(b)}) \cdot \hat{\mathbf{n}} = (\mathbf{v}_B - \mathbf{v}^{(b)}) \cdot \hat{\mathbf{n}} = 0. \quad (19.116)$$

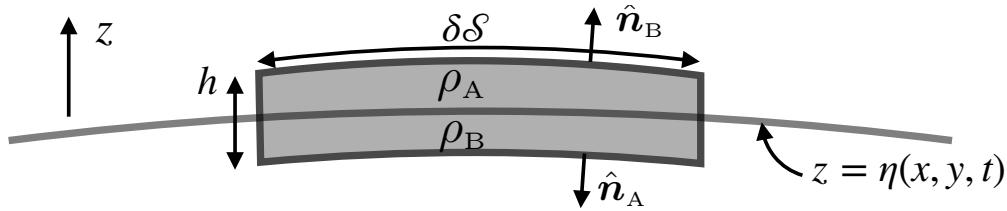


FIGURE 19.7: An infinitesimal cylindrical region used in formulating the stress boundary condition at an interface. The interface can be one that separates two fluid regions with densities ρ_A and ρ_B . It can also represent the boundary between a fluid (region A) and solid (region B). The interface generally moves with velocity $\mathbf{v}^{(b)}$. We orient the interface through the outward normals according to $\hat{\mathbf{n}} = \hat{\mathbf{n}}_B = -\hat{\mathbf{n}}_A$. The outward normal for region A points into region B whereas the outward normal for region B points into region A. For this particular interface, the normal direction has a nonzero projection in the vertical, $\hat{\mathbf{n}} \cdot \hat{\mathbf{z}} \neq 0$, thus allowing us to define the interface vertical position according to $z = \eta(x, y, t)$. This interface represents an idealized geometry useful to formulate the stress condition at the boundary between fluid media, such as the air-sea interface, fluid-land interface, or interior fluid interface (e.g., buoyancy surface). The single geometric assumption is that there are no overturning motions so that $\hat{\mathbf{n}} \cdot \hat{\mathbf{z}} \neq 0$, with this assumption based on convenience. The stress condition is general and so does not require this assumption.

A nonzero $\mathbf{v}^{(b)}$ corresponds here to a moving solid boundary, such as the region next to the grounding line of an ice-shelf. More commonly, in geophysical fluid applications we have $\mathbf{v}^{(b)} = 0$ for solid boundaries. Since there is no contribution from the kinetic stress, the stress condition (19.115) reduces to

$$\tau_A \cdot \hat{\mathbf{n}} = \tau_B \cdot \hat{\mathbf{n}}. \quad (19.117)$$

This identity is consistent with

$$\tau_A \cdot \hat{\mathbf{n}} = \boldsymbol{\tau}^{\text{friction } A}(\mathbf{x}, t, \hat{\mathbf{n}}) = -\boldsymbol{\tau}^{\text{friction } A}(\mathbf{x}, t, -\hat{\mathbf{n}}) = \tau_B \cdot \hat{\mathbf{n}} = \boldsymbol{\tau}^{\text{friction } B}(\mathbf{x}, t, \hat{\mathbf{n}}), \quad (19.118)$$

which expresses Newton's third law in the form of equation (19.5). Hence, the frictional force imparted by the land on the fluid is equal and opposite to that imparted by the fluid on the land.

19.9.3 No-slip boundary condition

At solid boundaries, the kinematic boundary condition from Section 15.4.1 sets the normal component of the velocity to zero

$$\mathbf{v} \cdot \hat{\mathbf{n}} = 0 \quad \text{kinematic no-flux condition on static material boundary.} \quad (19.119)$$

However, no place have we specified the tangential component of the velocity along a solid boundary. What is it?

Laboratory experiments over the 19th and 20th centuries indicate that there is no relative motion of molecules at solid-fluid interfaces.⁷ That is, a fluid at the solid-fluid interface has a velocity matching that of the solid so that the fluid sticks to the solid boundary as depicted in Figure 19.8. The no-slip boundary condition means that both the normal and tangential components of the fluid velocity vanish next to static solid boundaries

$$\mathbf{v} \cdot \hat{\mathbf{n}} = \mathbf{v} \cdot \hat{\mathbf{t}} = 0 \quad \text{no-slip condition on static solid boundaries.} \quad (19.120)$$

⁷As discussed in the historical essay by [Anderson \(2005\)](#), it was the work of Prandtl in 1905 that first exposed the fundamental nature of the no-slip boundary condition, and its role in establishing boundary layers around solid bodies immersed in a fluid flow.

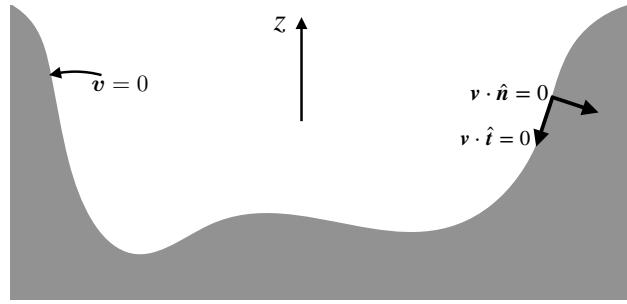


FIGURE 19.8: The no-slip boundary condition means that fluid exhibits zero relative motion at the solid-fluid boundary, $\mathbf{v} \cdot \hat{\mathbf{t}} = 0$. That is, the fluid sticks to the solid boundary. The kinematic no-normal flow boundary condition, $\mathbf{v} \cdot \hat{\mathbf{n}} = 0$, plus the dynamic no-slip boundary condition together mean that both the normal and tangential components of the velocity vanish at solid boundaries. For both the normal and tangential components of the velocity to vanish at a solid boundary requires the velocity to fully vanish at this boundary.

The no-slip boundary condition gives rise to an exchange of momentum between the solid and fluid, with this exchange mediated by viscosity. This boundary condition is the origin of the tangential stresses found in the Couette flow discussed in Section 19.7.2. In the absence of viscous friction, as per an inviscid perfect fluid, the no-slip boundary condition cannot be imposed since doing so would mathematically over-specify the flow. Consequently, for inviscid fluids the tangential component of the velocity remains unspecified at solid boundaries.

19.9.4 Lagrangian interface

Consider a Lagrangian interface within the fluid, with this interface defined so that

$$(\mathbf{v}_A - \mathbf{v}^{(b)}) \cdot \hat{\mathbf{n}} = (\mathbf{v}_B - \mathbf{v}^{(b)}) \cdot \hat{\mathbf{n}} = 0. \quad (19.121)$$

Again the kinetic stress contribution to equation (19.115) vanishes. We thus have the frictional stress condition (19.117) and a Newton's third law interpretation (19.118), yet now the frictional transfer takes place between two regions of the same fluid.

19.9.5 Permeable interface

Now allow for the interface to be permeable and introduce the dia-surface transport defined by equation (44.41)

$$(\mathbf{v}_A - \mathbf{v}^{(b)}) \cdot \hat{\mathbf{n}} = (\mathbf{v}_B - \mathbf{v}^{(b)}) \cdot \hat{\mathbf{n}} = w^{\text{dia}}, \quad (19.122)$$

where we assumed continuous dia-surface transport across the interface; i.e., what leaves one side enters the other. The kinetic stress thus adds to the frictional contribution in the stress boundary condition (19.115), with the kinetic stress providing a transfer of momentum across the interface through the dia-surface advection of fluid elements

$$\tau_A \cdot \hat{\mathbf{n}} + \rho_A \mathbf{v}_A w^{\text{dia}} = \tau_B \cdot \hat{\mathbf{n}} + \rho_B \mathbf{v}_B w^{\text{dia}}. \quad (19.123)$$

We consider two cases.

Single continuous fluid media

If the interface is within a single continuous fluid media, then $\rho_A = \rho_B$ and $\mathbf{v}_A = \mathbf{v}_B$, in which case the advection of momentum from region A to B is equal to that from region B to A . Consequently, the frictional condition (19.117) again holds.

Air-sea boundary interface

Consider now the air-sea boundary where region B is the ocean and region A the atmosphere. Introduce the dia-surface mass flux according to equation (15.63)

$$\rho_A w^{\text{dia}} = \rho_B w^{\text{dia}} = -\mathcal{Q}_m, \quad (19.124)$$

where \mathcal{Q}_m is the mass per time per surface area crossing the boundary. The minus sign is implied by the convention that $\mathcal{Q}_m > 0$ means that mass enters the ocean side of the interface and leaves the atmosphere side. Even though the densities and velocities on the two sides of the interface are unequal, mass conservation means that the mass transport entering the ocean must leave the atmosphere, and vice versa. Consequently, the stress boundary condition (19.123) takes the form

$$\tau_{\text{atm}} \cdot \hat{n} - \mathbf{v}_{\text{atm}} \mathcal{Q}_m = \tau_{\text{ocn}} \cdot \hat{n} - \mathbf{v}_{\text{ocn}} \mathcal{Q}_m \implies (\tau_{\text{atm}} - \tau_{\text{ocn}}) \cdot \hat{n} = (\mathbf{v}_{\text{atm}} - \mathbf{v}_{\text{ocn}}) \mathcal{Q}_m. \quad (19.125)$$

There is generally a jump in the frictional stress in the presence of mass transport across the air-sea interface. However, when introducing the dia-surface transport (19.122) we see that $\mathbf{v}_A \cdot \hat{n} = \mathbf{v}_B \cdot \hat{n}$ at the interface. So a nonzero $\mathbf{v}_{\text{atm}} - \mathbf{v}_{\text{ocn}}$ arises just from differences in the tangential velocities at the air-sea interface. If those velocities are equal so that $(\mathbf{v}_{\text{atm}} - \mathbf{v}_{\text{ocn}}) \mathcal{Q}_m = 0$, then we are again left with the friction boundary condition (19.117).

19.9.6 Summary comments

There are three terms in the general expression for the stress boundary condition (19.114), with contributions from pressure, friction, and kinetic stress. In the absence of surface tension, pressure is continuous at the interface; i.e., its value is the same on both sides of the interface. In the absence of mass transport across the interface, then we find a continuous kinetic stress at the interface that then leads to a continuous frictional stress. However, mass transport crossing the interface leads to a jump in the friction for those case where the tangential velocity has a jump across the interface. Assuming no such jump then leads again to a continuous frictional stress.

19.9.7 Comments on boundary layers

A fundamental advance in the relevance of fluid mechanics for describing observed flows came from the 20th century work of Prandtl and others who noted the central role of viscosity, even the tiny molecular values, in forming boundary layers when fluids flow next to rigid bodies.⁸ Prandtl's work focused on flows around airplane wings, thus supporting the development of aerodynamics as a scientific and engineering discipline. The key ideas transfer to geophysical flows where boundary layers form in the atmosphere and ocean as these fluids interact with the solid earth. Boundary layers also form where the atmosphere and ocean interact with one another.

A key facet of geophysical boundary layers concerns the dominance of turbulence in producing an eddy viscosity that is many orders larger than molecular viscosity. Indeed, molecular viscosity plays a role only in a very small region (the *laminar subregion*) immediately adjacent to the boundary, where as the bulk of the boundary layer is dominated by turbulence. The study of boundary layers is well outside our scope, as it is largely a topic of turbulence. However, in Chapter 28 we dive into the details of geophysical boundary layers that are affected by pressure, Coriolis, and

⁸For a historical treatment of boundary layer theory see [Anderson \(2005\)](#), or for a pedagogical study see [Tennekes and Lumley \(1972\)](#). The associated mathematical methods of singular perturbation theory and matched asymptotic expansions (e.g., [Dyke \(1975\)](#)) offer an example of how the study of physical systems can spawn the development of new mathematical methods.

turbulence induced friction. The leading role of rotation distinguishes geophysical boundary layers from engineering applications. The associated *Ekman boundary layers* are crucial for understanding circulation and transport in both the atmosphere and ocean.

19.10 Surface tension

Surface tension is present on surfaces that separate two immiscible liquids or between a liquid and gas. It has many consequences, such as allowing certain insects to walk on water even though their body density is greater than water, and for the predominantly spherical shape of rain drops. A molecular dynamics understanding of surface tension involves tools from physical chemistry that are outside of our scope. Instead, we develop heuristics sufficient to determine when one needs to be concerned with surface tension in the study of geophysical fluid mechanics. As shown in this section, surface tension can be safely ignored when concerned with length scales on the order of meters or larger.

As an introduction to surface tension, consider that atmospheric pressure at the earth's surface is roughly $p_{atm} = 10^5 \text{ N m}^{-2}$. As we saw in Section 19.7, pressure acts normal to a surface regardless the surface orientation. So fill a container of water whose weight per horizontal area is less than the atmospheric pressure, $\rho g h < p_{atm}$ and turn the container upside-down as in Figure 19.9. Does the water spill from the container? Common experience with drinking glasses indicate that water will spill. But what about containers with a very small cross-sectional area such as the pipettes used in chemistry laboratories? Pipettes, or more generally capillary tubes, hold the liquid regardless the orientation. They do so since their cross-sectional area is small enough to allow forces from surface tension to overcome gravitational instabilities acting at the liquid-gas interface. In the remainder of this section we discuss elements of surface tension with the goal to develop intuition as well as to determine the length scales where it becomes important.

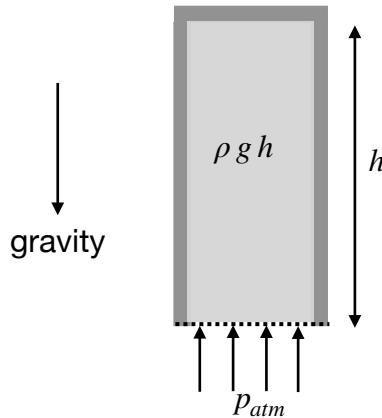


FIGURE 19.9: A container of water with density ρ and height h is placed upside-down. Atmospheric pressure, p_{atm} , will support water with thickness $h < p_{atm}/(\rho g) \approx 10 \text{ m}$ if the cross-sectional area of the container is small enough to allow for surface tension to overcome the gravitationally unstable waves that otherwise allow water to spill from the container. The liquid-gas interface supports both gravity waves (as in Section 32.8.3) and capillary waves. If the wavelength is small enough then surface tension suppresses the growth of unstable gravity waves so that the liquid remains within the *capillary tube*. However, for longer waves allowed by increasing the cross-sectional area, then any fluctuation will allow the gravitational instability to overcome surface tension, thus breaking the interface and releasing water.

19.10.1 Force balance on an air-water interface

Consider two fluids with distinct densities. Air and water provide one example of special importance to understanding physics at the ocean-atmosphere boundary. Another example concerns two immiscible layers of water within the ocean or two layers of air within the atmosphere. For molecules well within either of the fluid regions, the intermolecular forces are statistically isotropic. In contrast, intermolecular forces are not isotropic for molecules within a mean free path distance from the interface.⁹ Attractive (cohesive) intermolecular (van der Waals) forces dominate within a liquid whereas gas molecules generally feel more repulsive forces. Hence, a liquid molecule within the liquid-gas interface preferentially experiences an attractive force towards the liquid side of the interface, as depicted in Figure 19.10. Surface tension arises from the cohesive force per area acting between molecules in a direction that parallels the interface, with surface tension acting to resist perturbations to the interface shape.

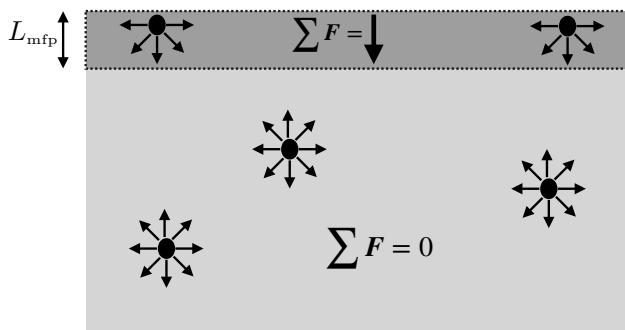


FIGURE 19.10: Surface tension at a liquid-gas interface arises from the anisotropic cohesive forces acting on liquid molecules within a mean-free-path distance, L_{mfp} , from the interface, which contrasts to the isotropic cohesive forces acting away from the interface. The net intermolecular force vanishes for interior molecules, whereas the net force acts inward on molecules at the interface. Surface tension refers to the cohesive force per area acting between molecules in a direction that is parallel to the interface.

Anisotropic attractive intermolecular forces cause the interface between the two fluids to behave as a stretched membrane that experiences a tensile force resisting any stretching of the interface. The magnitude of the tensile force per unit length is the *surface tension*, γ (units N m^{-1}). The surface tension is a property of the two fluids, including their temperature, as well as any impurities that might be included on the interface; e.g., oil on the surface of water effects properties of the capillary waves found on the air-sea interface. In the following we focus on the liquid-gas example to be specific and to expose issues that arise in studies of the air-sea interface. For a liquid-gas interface surrounding a liquid drop, the tensile force acts to curve the interface towards the liquid into a spherical shape.

The tensile force along a line segment is directed normal to the line and tangent to the interface

$$\mathbf{f}_{\text{interface}} = -\gamma \hat{\mathbf{n}} \wedge \delta \mathbf{x}, \quad (19.126)$$

where $\hat{\mathbf{n}}$ is a normal vector pointing towards the center of the curved interface, and $\delta \mathbf{x}$ is a line element oriented so that the normal $\hat{\mathbf{n}}$ points to the left facing in the direction of the line increment. Figure 19.11 depicts the surface tensile forces acting on the surface of a spherical bubble of water. Note that it is sometimes useful to consider the product γdS as the work (units of $\text{N m}^{-1} = \text{Joule}$)

⁹ As discussed in Section 12.2, the mean free path is a statistical measure of the distance a molecule moves before hitting another molecule.

required to create an area, dS , on the interface. We make use of this energetic perspective in Section 19.10.2.

To develop an expression for the pressure jump across the liquid-gas interface, consider a spherical droplet of radius R shown in Figure 19.11 and focus on the circular cross-section cut through the center of the sphere. The net tensile force acting on the circumference of the circle is

$$\mathbf{F}_{\text{circle}} = \oint_{\text{circle}} \mathbf{f}_{\text{interface}} = - \oint_{\text{circle}} \gamma \hat{\mathbf{n}} \wedge \delta \mathbf{x} = -2\pi R \gamma \hat{\mathbf{z}}. \quad (19.127)$$

Equilibrium of the spherical droplet is realized by a pressure jump across the circular cross-sectional area

$$\pi R^2 (p_{\text{in}} - p_{\text{out}}) = 2\pi R \gamma \implies (p_{\text{in}} - p_{\text{out}}) = 2\gamma/R. \quad (19.128)$$

Hence, the pressure jump is determined by the surface tension (a property of the two fluids) and the curvature of the sphere, R , which is also the radius of curvature for the sphere. Pressure is higher inside of the sphere, with this pressure required to balance the pressure outside the sphere plus the surface tension. Notably, equilibrium for smaller bubbles requires a larger pressure difference than for larger bubbles.

The pressure jump is known as the *capillary pressure*. It arises from surface tension and curvature of the interface. The relation (19.128) is a special case of the Young-Laplace formula, specialized here to a sphere.

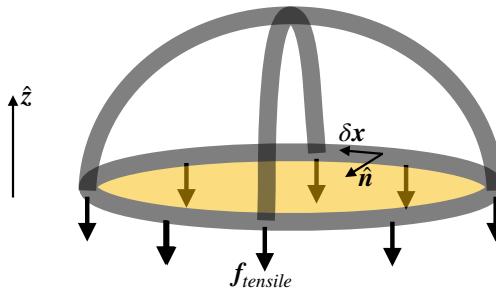


FIGURE 19.11: Surface tension on a spherical water droplet, with water on the inside of the sphere and air on the outside. The tensile forces act parallel to the spherical interface between the water and air. When cutting a circular cross-section as shown here, the surface tensile force acts downward. In equilibrium, the net tensile forces acting downward along the circumference of the hemisphere ($2\pi R \gamma$) are balanced by a pressure jump across the droplet, with the interior pressure larger than the exterior pressure. Focusing on the circular cross-section, this area remains static so long as $2\pi R \gamma = \pi R^2 (p_{\text{in}} - p_{\text{out}})$, leading to a pressure jump across the droplet interface $p_{\text{in}} - p_{\text{out}} = 2\gamma/R$.

19.10.2 Young-Laplace formula

We garner added insight into the physics of surface tension by considering the energetics required to enable a virtual displacement of a surface through a pressure field along with the work required to change the area of the surface. The resulting equation for the pressure jump across the surface is referred to as the *Young-Laplace formula*, which expresses the pressure jump in terms of the surface tension and the principle radii of curvature for the surface.

Consider a horizontal surface depicted in Figure 19.12 that represents the interface separating fluid-A from fluid-B, with $\hat{\mathbf{n}}$ a unit normal vector oriented from fluid-A to fluid-B. Now consider a virtual displacement of each point along the interface by an infinitesimal distance, δh , with $\hat{\mathbf{n}} \delta h$ connecting points on the initial position of the interface to the displaced position, where $\delta h > 0$ if

the displacement is directed towards fluid-B and $\delta h < 0$ if directed towards fluid-A. The (signed) volume swept out by an infinitesimal area dA is given by $\delta h dA$, with this volume realized by applying the pressure work to the surface

$$W_{\text{volume}} = (p_B - p_A) \delta h dA. \quad (19.129)$$

For example, if $p_B > p_A$ and the displacement is into fluid-B ($\delta h > 0$), then $W_{\text{volume}} > 0$, thus indicating the need to apply positive work to the surface to move it into the fluid region with higher pressure. Conversely, the required pressure work is negative if displacing the interface into a region with lower pressure.

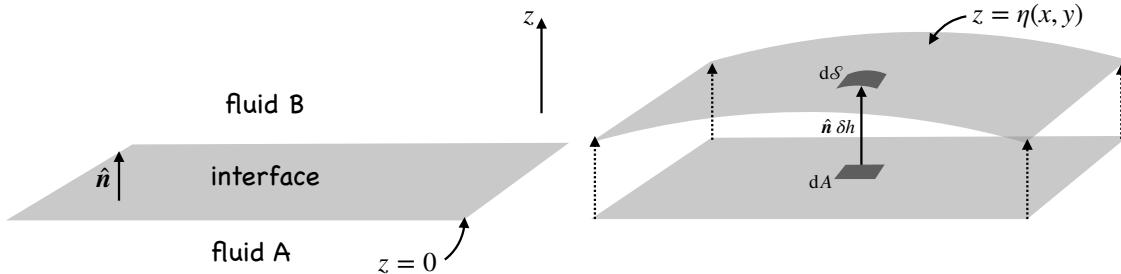


FIGURE 19.12: Left panel: initial position of an interface separating two fluid regions, fluid-A and fluid-B. Right panel: infinitesimal displacement of the interface sweeps out a volume in space. To determine the volume, extend a unit normal vector, \hat{n} , from the initial interface position and pointing towards fluid-B. Let δh be the distance along that normal to the new position, with $\delta h > 0$ if the displacement moves towards fluid-B and $\delta h < 0$ for displacements pointing to fluid-A. We assume that displacements at each interface point can move independently of adjacent points, so that the interface area generally changes.

In the presence of surface tension, work must overcome the surface area energy in order to change to the interface area

$$W_{\text{area}} = \gamma \delta A, \quad (19.130)$$

where δA is the change in area of an infinitesimal element on the interface

$$\delta A = dS - dA \quad (19.131a)$$

$$= dA \left[\sqrt{1 + (\nabla \delta h)^2} - 1 \right] \quad (19.131b)$$

$$\approx dA (\nabla \delta h)^2 / 2. \quad (19.131c)$$

To reach this result we made use of equation (5.30) that relates the area of an infinitesimal element on a curved surface to the area of its horizontal projection (see Section 5.3.1). We next make use of the surface curvature detailed in Section 5.3.2, where equation (5.35) shows that the vertical displacement is given, for small displacements, by

$$\delta h \approx \frac{1}{2} R_1^{-1} (\mathbf{x} \cdot \mathbf{e}_1)^2 + \frac{1}{2} R_2^{-1} (\mathbf{x} \cdot \mathbf{e}_2)^2. \quad (19.132)$$

R_1^{-1}, R_2^{-1} are the eigenvalues and $\mathbf{e}_1, \mathbf{e}_2$ are the corresponding eigenvectors of the matrix of second partial derivatives of $\delta h(x, y)$, whereas the inverse eigenvalues, R_1, R_2 , are the radii of curvature of the displaced surface. Orienting the Cartesian axes along the eigenvector directions renders

$$(\nabla \delta h)^2 \approx (x/R_1)^2 + (y/R_2)^2 = \delta h \left[\frac{1}{R_1} + \frac{1}{R_2} \right], \quad (19.133)$$

where we set

$$\delta h/R_1 = (x/R_1)^2 \quad \text{and} \quad \delta h/R_2 = (y/R_2)^2. \quad (19.134)$$

We are thus led to the area difference

$$\delta A \approx dA \delta h \left[\frac{1}{R_1} + \frac{1}{R_2} \right]. \quad (19.135)$$

Note that $\delta A > 0$ whether displacing the surface into a concave or convex direction, since the sign of δh accounts for the sign of the radii of curvature.

The total work for the interface displacement is given by the sum of the area work and volume work

$$W_{\text{area}} + W_{\text{volume}} = dA \delta h [\gamma (R_1^{-1} + R_2^{-2}) + p_B - p_A] \quad (19.136)$$

and equilibrium results if the work vanishes

$$p_A - p_B = \gamma (R_1^{-1} + R_2^{-2}). \quad (19.137)$$

This equation is the Young-Laplace formula, which reduces to equation (19.128) if $R_1 = R_2$ as for a sphere. It says that there is a pressure jump, known as the *capillary pressure*, across an interface as given by the surface tension times the sum of the inverse principle radii of curvature. Pressure is higher on the concave side of the interface, such as fluid-A depicted in Figure 19.12 or the inside of a bubble/droplet.

19.10.3 Some oceanographic examples

We close this section with examples relevant to the ocean.

Soluble gas bubbles inside water

The previous considerations hold whether there is liquid or gas inside the spherical droplet/bubble. As an example, consider a spherical gas bubble of radius $R = 10^{-6}$ m inside water and make use of the air-water surface tension $\gamma = 0.072$ N m⁻¹

$$p_{\text{in}} - p_{\text{out}} = 2\gamma/R \approx 144 \times 10^3 \text{ N m}^{-2} = 1.42 p_{\text{atm}}, \quad (19.138)$$

where $p_{\text{atm}} = 101 \times 10^3$ N m⁻² is standard atmospheric pressure. If the gas inside the bubble is water soluble, then the enhanced pressure inside the bubble will induce more gas to dissolve in the water, which in turn will cause the bubble to shrink and thus increase the pressure inside the bubble. Small bubbles of soluble gases can thus be squeezed towards zero radius by the effects of surface tension induced pressure.

Length scale for capillary waves

Capillary waves exist along the air-sea interface due to the restorative effects from surface tension. When present within a gravity field, the capillary waves appear along with gravity waves. We see capillary waves when there is a very slight breeze on the ocean surface. Capillary waves also arise when a tiny stone is thrown into a still pond, whereas gravity waves dominate when a larger stone is used. This phenomenology arises from the following considerations of the dispersion relation.

The dispersion relation for capillary-gravity waves (e.g., Section 54 of *Fetter and Walecka (1980)*) is given by

$$\omega^2 = k g \left[1 + \frac{k^2 \gamma}{g \rho} \right], \quad (19.139)$$

where ρ is the density of water, $k = 2\pi/\lambda$ is the wave number, and ω is the radial frequency. The non-dimensional parameter $k^2 \gamma / (g \rho)$ provides a regime boundary where capillary waves are important ($k^2 \gamma > g \rho$) and negligible ($k^2 \gamma < g \rho$). To deduce a correspondingly length scale we introduce the wavelength

$$\lambda_{\text{cap-grav}} = 2\pi \sqrt{\frac{\gamma}{\rho g}}, \quad (19.140)$$

with $\lambda < \lambda_{\text{cap-grav}}$ the capillary wave regime and $\lambda > \lambda_{\text{cap-grav}}$ the gravity wave regime. Using $\gamma = 0.072 \text{ N m}^{-1}$ and $\rho = 1000 \text{ kg m}^{-3}$ leads to

$$\lambda_{\text{cap-grav}} = 2\pi \sqrt{\frac{\gamma}{\rho g}} \approx 0.017 \text{ m} = 17 \text{ cm}. \quad (19.141)$$

Since this book is mostly concerned with length scales larger than $\lambda_{\text{cap-grav}}$, we generally ignore the dynamics of capillary waves for our study of geophysical fluid mechanics.

19.10.4 Further study

Although we have no further concern for surface tension in this book, its study forms an important aspect of air-sea interaction physics. There are many places to continue its study, with the following offering treatments similar to the physical ideas given here.

This [30-minute video](#) from Prof. L. Trefethen provides a pedagogical summary of surface tension. The upside-down container of water in Figure 19.9 is based on a discussion of capillary-gravity waves in Section 3.1.3 of [Falkovich \(2011\)](#). Section 1.9 of [Batchelor \(1967\)](#) discusses how surface tension acts between two fluid media, with that discussion extended into his Section 3.3 to develop boundary conditions for velocity and stress. The bubble example in Section 19.10.3 is taken from Section 1.3 of [Kundu et al. \(2016\)](#). Section 4.10 of [Kundu et al. \(2016\)](#) provides a detailed accounting of the force balance at an interface, offering more details than found in [Batchelor \(1967\)](#). The energetic arguments used to derive the Young-Laplace formula follows Section 61 of [Landau and Lifshitz \(1987\)](#). Section 46 of [Fetter and Walecka \(1980\)](#) discuss the dynamics of membranes under tension, and Section 54 considers surface capillary-gravity waves.

19.11 Exercises

EXERCISE 19.1: FORCE BALANCE FOR A NON-ACCELERATING OCEAN

Consider an ocean basin, \mathcal{R} , as in Figure 19.2 with bottom interface separating the ocean fluid from the solid-earth, and upper interface separating the ocean fluid from the atmosphere, and where the atmosphere has a nonzero mass.¹⁰ Assume no matter crosses the ocean boundaries; i.e., no evaporation, precipitation, nor river runoff. However, assume there is a horizontal stress over the ocean surface with a stress vector $\boldsymbol{\tau}^{\text{surf}}$, thus imparting a net surface stress that drives motion. Allow for the ocean bottom to exchange momentum with the solid-earth through a bottom stress, $\boldsymbol{\tau}^{\text{bott}}$.

- What is the force balance for the full ocean domain if the center of mass experiences no acceleration? Express this force balance in words and in equations. The answer should be generally stated, with no need for specific details. Hint: consider Figure 19.2 and include the missing forces to this diagram. Mathematically express the force balance as an integral expression as per the weak formulation.

¹⁰In some applications it is suitable to assume a zero mass atmosphere. For this exercise, however, we do not make that assumption.

- (b) Express the vertical component of the force balance assuming the fluid is in an approximate hydrostatic balance, meaning that pressure and gravity forces balance.¹¹
- (c) Assume motion on a flat plane with a constant rotation rate $\Omega = \hat{z}\Omega = \hat{z}(f/2)$ as per an f -plane (Section 23.2). Also assume that the center of mass motion vanishes, $\int \rho \mathbf{v} dV = 0$. Discuss the resulting zonal and meridional force balance. Hint: one of the forces appearing in part (a) now vanishes.



¹¹In Section 23.3 we provide a discussion of when this approximation is appropriate for moving fluids. That discussion is not needed for the current exercise. Instead, we merely assume the vertical momentum balance is hydrostatic.

Pressure form stress

As introduced in our discussion of Cauchy's stress principle in Section 19.1, pressure form stress is the horizontal stress arising from pressure that acts on a sloped surface or interface. As a contact force per area, Newton's third law describes how form stress renders a transfer of pressure forces across interfaces, with pressure form stress affecting a vertical transfer of horizontal pressure forces. Hence, it provides an inviscid/reversible mechanism for the vertical transfer of horizontal momentum, thus complementing the vertical transfer associated with viscosity in the presence of tangential shear stresses (Section 19.7.2).

READER'S GUIDE TO THIS CHAPTER

In this chapter we study how form stresses act on the variety of interfaces encountered in geophysical fluids. We build from the study of stresses in Chapter 19, with an understanding of pressure form stress greatly enhancing our understanding of horizontal forces acting in geophysical fluids. The focus on horizontal forces in this chapter complements our studies in Chapter 25, whereby the net vertical acceleration from pressure and gravitational forces is repackaged into the buoyancy force.

20.1	Pressure form stresses at an interface	406
20.1.1	Concerning the sign of a form stress	407
20.1.2	Mathematical expression for form stress	407
20.1.3	Comments	408
20.2	Topographic and atmospheric form stresses on the ocean	409
20.2.1	Zonally symmetric ridge	409
20.2.2	Form stress transfer between the fluid and its boundaries	409
20.2.3	Decomposing topographic form stress	411
20.3	Interfacial form stress	412
20.3.1	Interfacial form stresses (IFS) transferred between layer	412
20.3.2	Zonally integrated interfacial form stress	414
20.3.3	Comments	414
20.4	Axial angular momentum budget for an ocean domain	414
20.4.1	Anticipating the budget	415
20.4.2	Axial angular momentum	415
20.4.3	Depth integrated budget	416
20.4.4	Atmospheric and topographic form stresses	416
20.4.5	Turbulent stresses at the surface and bottom	417
20.4.6	Summary budget for column integrated axial angular momentum . .	417
20.4.7	Zonal domain topology/geometry and zonal integration	418

20.4.8 Steady domain integrated balance	419
20.4.9 Remarks on a modified expression for the form stress	420
20.4.10 Steady zonal and depth integrated budget	420
20.4.11 Southern Ocean balances	421
20.4.12 Topographic form stress and ocean gyres	421
20.4.13 Further study	421

20.1 Pressure form stresses at an interface

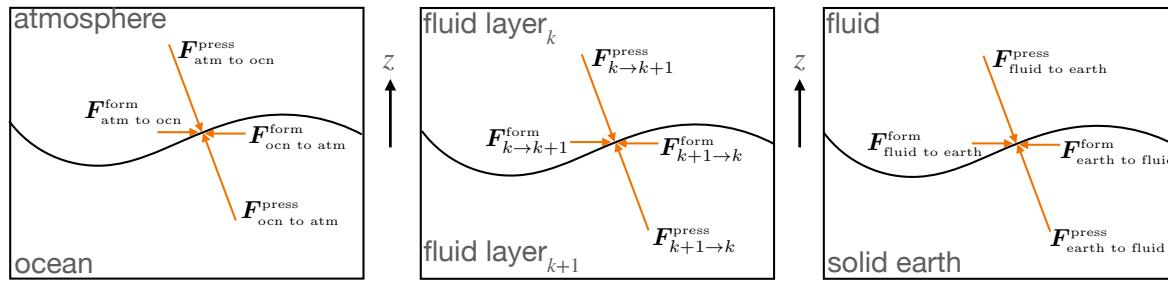


FIGURE 20.1: Illustrating the three interfaces of concern in geophysical fluid mechanics for the discussion of form stresses. Continuity of pressure at the interface, through Newtons' third law, means that the form stress on one side of the interface is equal and opposite to that acting on the other side. Left panel: a curved atmosphere-ocean interface leads to an atmospheric form stress acting on the ocean, $\mathbf{F}_{\text{atm to ocn}}^{\text{form}}$, and its equal and opposite oceanic form stress acting on the atmosphere, $\mathbf{F}_{\text{ocn to atm}}^{\text{form}} = -\mathbf{F}_{\text{atm to ocn}}^{\text{form}}$. Middle panel: a curved interior ocean interface (e.g., a buoyancy surface) leads to an interfacial form stress acting on the lower layer, $\mathbf{F}_{k \rightarrow k+1}^{\text{form}}$, and its equal and opposite interfacial form stress acting on the upper layer, $\mathbf{F}_{k+1 \rightarrow k}^{\text{form}} = -\mathbf{F}_{k \rightarrow k+1}^{\text{form}}$. Right panel: a curved fluid solid earth interface leads to a fluid form stress acting on the solid earth, $\mathbf{F}_{\text{fluid to earth}}^{\text{form}}$, and its equal and opposite oceanic form stress acting on the fluid, $\mathbf{F}_{\text{earth to fluid}}^{\text{form}} = -\mathbf{F}_{\text{fluid to earth}}^{\text{form}}$. The magnitude of the form stresses is a function of the slope of the interface (steeper slopes lead to larger magnitude) and the pressure acting at the interface.

As depicted in Figure 20.1, there are three surfaces or interfaces across which we commonly study form stresses in geophysical fluids.

- **ATMOSPHERE-OCEAN FORM STRESS:** A form stress occurs at the air-sea interface. From the perspective of the ocean, the nonzero sea level pressure from the atmosphere provides a pressure acting on the sloped upper ocean free surface, thus rendering an *atmospheric form stress* acting on the ocean. Through Newton's third law, this form stress is met by the equal in magnitude but oppositely directed *ocean form stress* acting on the atmosphere.
- **INTERIOR FLUID INTERFACIAL FORM STRESS:** A form stress occurs on an internal interface within the fluid, and we study such *interfacial form stresses* in Section 20.3. Although the interface is arbitrary, it is dynamically very interesting to study form stresses acting on buoyancy isosurfaces. The reason is that buoyancy interfaces are directly connected to the geostrophic motion studied in Chapter 26. In particular, in Section 26.7 we study form stresses associated with buoyancy interfaces found in geostrophic flows.
- **FLUID-TOPOGRAPHY FORM STRESS:** A form stress exists at a solid/fluid boundary, at which the ocean or atmosphere impart a pressure force on the solid earth. Through Newton's third law, the pressure force imparted by the fluid on the solid earth is met equally in magnitude but oppositely in direction by a force provided by the solid earth onto the fluid. The horizontal

projection of this force from the earth to the fluid is the *topographic form stress* and it is considered in Section 20.2.

20.1.1 Concerning the sign of a form stress

As a vector, pressure form stress has a direction and a magnitude, with three examples depicted in Figure 20.1. Even so, keeping track of the direction can be confusing if it is unclear who is the giver of the form stress and who is the receiver. To help in understanding the sign, imagine pushing against a rigid wall: you exert a force on the wall in one direction whereas, through Newton's third law, the wall exerts an equal and opposite force on you. One simply must specify the origin of the force in order to determine its sign. For example, as illustrated in Figure 20.2, is one concerned with the force applied by the ocean bottom pressure onto the earth (liquid ocean is giver and solid earth is receiver), or instead with the force from the earth applied onto the ocean fluid (earth is giver and ocean is receiver)? These forces have equal magnitude but opposite direction. Knowing the direction requires knowing the force giver and/or the force receiver.

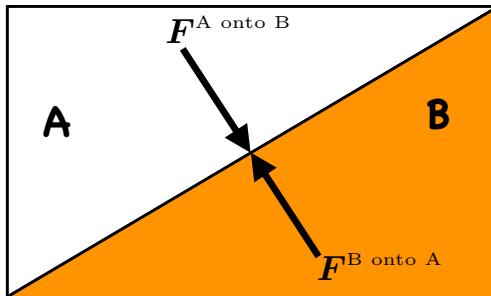


FIGURE 20.2: Contact forces, such as pressure, satisfy Newton's third law. Hence, the contact force at a point imparted by region A onto region B, $\mathbf{F}^A \text{ onto } B$, is equal and opposite to the force imparted by region B onto region A so that $\mathbf{F}^A \text{ onto } B = -\mathbf{F}^B \text{ onto } A$. In the case where A=fluid (ocean or atmosphere) and B=solid earth, we generally refer to the horizontal portion of $\mathbf{F}^{\text{earth onto fluid}}$, per horizontal area, as the *topographic form stress*.

20.1.2 Mathematical expression for form stress

To expose the mathematics of form stress, consider a surface, \mathcal{S} , such as that shown in Figure 20.3. To develop a kinematic decomposition of the pressure force, we assume the surface has no vertical section, with this assumption commonly satisfied by surfaces of interest for geophysical flows.

Assuming the surface has no vertical sections allows us to write the vertical position of a point on the surface as¹

$$z = \eta(x, y, t). \quad (20.1)$$

The outward normal pointing away from the top side of the surface is given by

$$\hat{\mathbf{n}}_{\text{top}} = \frac{\nabla(z - \eta)}{|\nabla(z - \eta)|} = \frac{\hat{\mathbf{z}} - \nabla\eta}{\sqrt{1 + |\nabla\eta|^2}}. \quad (20.2)$$

Multiplying the pressure times the horizontal area element on the surface, $d\mathcal{S}$, leads to the net pressure force acting at a point on the top side of the surface

$$\mathbf{F}^{\text{press}} = -p \hat{\mathbf{n}}_{\text{top}} d\mathcal{S} = -p (\hat{\mathbf{z}} - \nabla\eta) dA = -p (-\partial_x \eta \hat{\mathbf{x}} - \partial_y \eta \hat{\mathbf{y}} + \hat{\mathbf{z}}) dA. \quad (20.3)$$

¹See the geometry discussion in Chapter 5 for more on the maths of such surfaces. Also see the discussion of generalized vertical coordinates in Chapters 43 and 44.

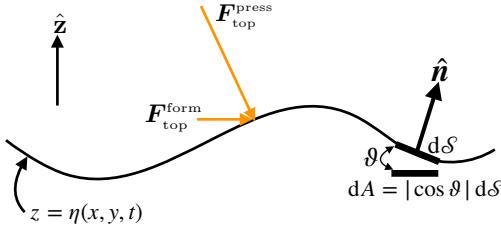


FIGURE 20.3: The pressure force acting on an arbitrary surface is given by $\mathbf{F}^{\text{press}} = -p \hat{\mathbf{n}} d\mathcal{S}$, where $d\mathcal{S}$ is the surface area element. We here depict the pressure acting on the top side of a surface, $\mathbf{F}_{\text{top}}^{\text{press}}$. Through Newton's third law, the pressure force vector acting on the top side of the interface is equal and opposite to the pressure force acting on the bottom side: $\mathbf{F}_{\text{top}}^{\text{press}} = -\mathbf{F}_{\text{bot}}^{\text{press}}$. The horizontal component of this force vector arises from the slope; i.e., the geometric *form* of the surface. We thus refer to the horizontal pressure force per area as the *form stress*, $\mathbf{F}_{\text{top}}^{\text{form}} = -\mathbf{F}_{\text{bot}}^{\text{form}}$. The area element on the surface, $d\mathcal{S}$, has a horizontal projection given by $dA = dx dy = \cos \vartheta d\mathcal{S}$, with the angle assumed to be within the range $-\pi/2 < \vartheta < \pi/2$ so that the surface is nowhere vertical.

In this equation we used the identity²

$$d\mathcal{S} = |\nabla(z - \eta)| dA = \sqrt{1 + |\nabla\eta|^2} dA, \quad (20.4)$$

with

$$dA = dx dy \quad (20.5)$$

the horizontal projection of the surface area element (see Figure 20.3). We identify the form stress acting on the top side of this interface as

$$\text{pressure form stress acting on top side of interface} \equiv p \nabla\eta. \quad (20.6)$$

The name follows since the stress is determined by the “form” of the surface as measured by its slope, $\nabla\eta$. We can thus write the pressure force acting on the top side of the surface as the sum of a vertical pressure force plus a horizontal pressure form stress

$$\mathbf{F}_{\text{top}}^{\text{press}} = \hat{z} [\hat{z} \cdot \mathbf{F}_{\text{top}}^{\text{press}}] + \mathbf{F}_{\text{top}}^{\text{form}} = p (-\hat{z} + \nabla\eta) dA \quad \text{pressure force on top of interface.} \quad (20.7)$$

Newton's third law, as manifested by Cauchy's Stress principle (Section 19.1) says that there is a local mechanical equilibrium of pressure contact forces within a fluid. Additionally, as seen in our discussion of stress on an interface in Section 19.9, this local equilibrium holds for pressure forces acting on interfaces separating two fluids, such as the atmosphere and ocean, as well as a fluid and the solid earth. Thus, the contact pressure force acting on the bottom side of the interface is equal in magnitude but oppositely directed to the contact force acting on the top side (see Section 19.7.2)

$$\mathbf{F}_{\text{bot}}^{\text{press}} = \hat{z} [\hat{z} \cdot \mathbf{F}_{\text{bot}}^{\text{press}}] + \mathbf{F}_{\text{bot}}^{\text{form}} = p (+\hat{z} - \nabla\eta) dA \quad \text{pressure force on bottom of interface.} \quad (20.8)$$

20.1.3 Comments

Consider a container filled with water at rest. The horizontal pressure forces acting on the container sides are pressure form stresses between the water and the container. As discussed in Section 19.4, without motion we know that the form stresses balance over the whole of the fluid-container boundary, whereas horizontal motion occurs if the form stresses are out of balance. Quite generally,

²The identity (20.4) follows from trigonometry summarized in Figure 20.3. See further details in the kinematic boundary conditions of Section 15.4 and the analogous dia-surface transport in Section 44.4.

when concerned with fluid motion, we are interested in processes that lead to unbalanced form stresses. For example, when studying bottom topographic form stresses in the ocean, the bulk of the form stress acts to support the ocean water within the ocean basin. The dynamically active portion of the topographic form stress, associated with the fluid motion, is a small residual of the total form stress. Careful analysis is required to diagnose dynamically relevant patterns, with [Molemaker et al. \(2015\)](#) and [Gula et al. \(2015\)](#) presenting one method, and we explore their method as part of Exercise 36.10.

Notably, form stress, particularly interfacial form stress, can appear mysterious in some presentations. Part of the reason is that it sometimes appears seemingly without prior motivation as part of mathematical manipulations of the momentum equation. We illustrate these manipulations in Sections 20.2, 20.3, and 20.4, yet aim to offer sufficient physical motivation to help guide the maths. Another reason for the mystery is that the signs ascribed to form stress are often not clearly specified, with such ambiguities motivating the somewhat pedantic discussion in Section 20.1.1.

20.2 Topographic and atmospheric form stresses on the ocean

In this section we focus on form stress arising from the shape of the solid earth interface with the atmosphere and ocean; i.e., the *fluid-topographic form stress*. As we are normally interested in the form stress applied to the fluid, we focus on the *topographic form stress*. We also encounter the form stress associated with undulations in the ocean free surface and the atmospheric pressure at that interface, with the *atmospheric form stress* the stress imparted to the ocean from the atmosphere.

20.2.1 Zonally symmetric ridge

In Figure 20.4 we depict an idealized ridge with an example oceanic pressure field to illustrate the nature of topographic form stress acting on a fluid. Rather than assuming exact hydrostatic equilibrium as in Figure 19.5, with zero horizontal pressure gradients, we here consider pressure to be higher to the west of the ridge than to the east. Since the ridge is assumed to be symmetric in the zonal direction, we conclude that the topographic form stress, which acts just at the fluid-solid interface, is higher on the west side of the ridge than on the east. In turn, the net topographic form stress acting on the fluid is to the west, whereas the net oceanic form stress acting on the solid earth is to the east. We encounter this situation in Section 20.4.11 when studying the force balances for steady circulation in a zonally periodic channel.

20.2.2 Form stress transfer between the fluid and its boundaries

We now illustrate how topographic form stress commonly appears in the study of momentum balances acting on a fluid. For definiteness, consider a column of ocean fluid extending from the bottom at $z = \eta_b(x, y)$ to the free surface at $z = \eta(x, y, t)$, and focus on the zonal force balance such as depicted in Figure 20.4. In computing the acceleration acting on this column at a particular horizontal position, we need to determine the depth integrated zonal pressure gradient

$$\text{depth integrated zonal pressure gradient} = - \int_{\eta_b}^{\eta} \frac{\partial p}{\partial x} dz. \quad (20.9)$$

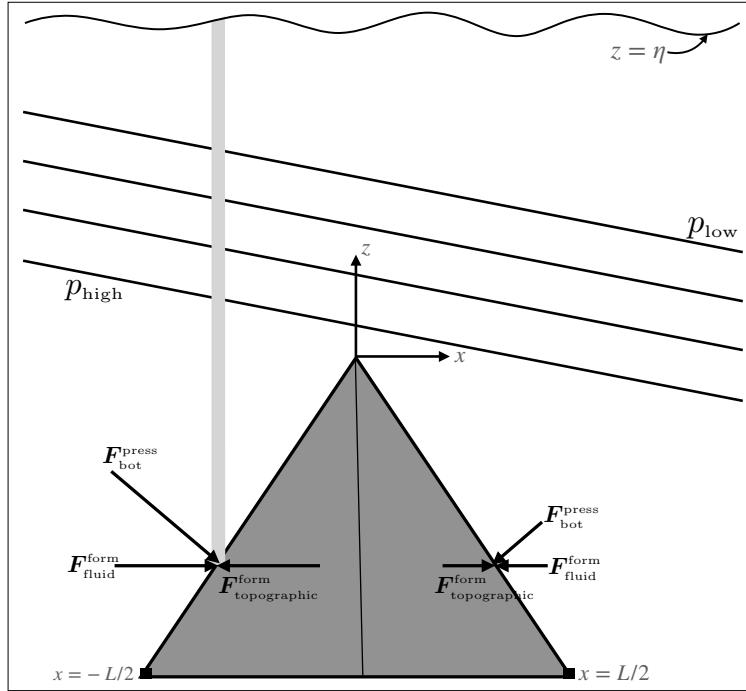


FIGURE 20.4: Depicting contact forces acting at a fluid-solid interface. The topographic form stress acts on the fluid and the equal in magnitude but oppositely directed fluid form stress acts on the solid earth. In this illustration the topography is assumed to be a ridge in the shape of an equilateral triangle. We also assume there is higher pressure to the west of the ridge than to the east as per the zonal balance discussed in Figure 20.8 for a Southern Ocean ridge. Hence, the topographic form stress has a larger magnitude on the western side of the ridge than on the eastern side. When integrated over the full ridge, there will be a net westward topographic form stress acting on the fluid and a net eastward fluid form stress acting on the solid earth. If the ridge was free to move, it would move to the east. The thin gray column extends from the solid earth bottom to the ocean free surface. As this column sits on the western side of the ridge, topographic form stress provides a westward acceleration at the column bottom. The net acceleration of the column is determined by integrating the contact forces around the column boundary and body forces throughout the column interior. We study the axial angular momentum budget for a fluid column in Section 20.4, with form stresses appearing in that budget.

We expose the contact force version of the pressure force by making use of Leibniz's Rule (Section 15.5) to write

$$-\int_{\eta_b}^{\eta} \frac{\partial p}{\partial x} dz = \underbrace{-\frac{\partial}{\partial x} \int_{\eta_b}^{\eta} p dz}_{\text{zonal deriv depth integrated pressure}} + \underbrace{\frac{\partial \eta}{\partial x} p_a}_{\text{atmospheric form stress}} - \underbrace{\frac{\partial \eta_b}{\partial x} p_b}_{\text{topographic form stress}}, \quad (20.10)$$

where we introduced the atmospheric pressure, p_a , and bottom pressure, p_b , which act at the surface and bottom interfaces, respectively. The decomposition identifies the following three pressure contributions to the pressure force acting on the fluid column.

- **ZONAL DERIVATIVE OF THE COLUMN INTEGRATED PRESSURE:** The first term arises from the zonal derivative of pressure across the vertical sides of the column

$$\text{zonal derivative of layer integrated pressure} = -\frac{\partial}{\partial x} \int_{\eta_b}^{\eta} p dz. \quad (20.11)$$

This term leads to a net eastward acceleration if the depth integrated pressure is higher to the west than the east.

- ATMOSPHERIC FORM STRESS AT THE FREE SURFACE: In the presence of a sloping free surface interface, $\partial\eta/\partial x \neq 0$, the atmospheric pressure, p_a , imparts an atmospheric form stress onto the ocean

$$\text{zonal atmospheric form stress acting on ocean} = \frac{\partial\eta}{\partial x} p_a. \quad (20.12)$$

For example, if the free surface slopes up to the east, $\partial\eta/\partial x > 0$, then the atmosphere provides a positive (eastward) zonal form stress onto the ocean. In turn, through Newton's third law, the ocean provides a westward zonal form stress to the atmosphere.

- TOPOGRAPHIC FORM STRESS ON OCEAN: The bottom pressure, p_b , present at $z = \eta_b$ imparts an oceanic form stress to the solid earth

$$\text{zonal oceanic form stress acting on solid earth} = \frac{\partial\eta_b}{\partial x} p_b. \quad (20.13)$$

In turn, through Newton's third law, the topographic form stress acting on the ocean is equal in magnitude but oppositely directed

$$\text{zonal topographic form stress acting on ocean} = -\frac{\partial\eta_b}{\partial x} p_b. \quad (20.14)$$

For example, if the bottom rises to the east, so that $\partial\eta_b/\partial x > 0$, then the oceanic form stress acting on the solid earth is eastward whereas the topographic form stress acting on the ocean is westward. As a check, we verify that the signs of these form stresses are consistent with those in Figure 20.4.

20.2.3 Decomposing topographic form stress

Assuming the fluid maintains an approximate hydrostatic balance, and focusing on the oceanic case, allows us to decompose the bottom pressure into the following terms

$$p_b = p_a + g \int_{\eta_b}^{\eta} \rho dz \quad (20.15a)$$

$$= \underbrace{g \rho_0 [\eta + p_a/(g \rho_0)]}_{\text{external}} - \underbrace{g \rho_0 \eta_b}_{\text{topog}} + \underbrace{g \int_{\eta_b}^{\eta} (\rho - \rho_0) dz}_{\text{internal}} \quad (20.15b)$$

$$\equiv p_{\text{ext}} + p_{\text{topog}} + p_{\text{int}}. \quad (20.15c)$$

We refer to the contribution from applied surface pressure plus surface height undulations as *external*, whereas those arising from density deviations relative to a constant reference density are termed *internal*. There is a final contribution from bottom topography itself. Multiplying this pressure by the bottom topography slope renders an expression for the various contributions to topographic form stress

$$-p_b \nabla \eta_b = -p_{\text{ext}} \nabla \eta_b - p_{\text{topog}} \nabla \eta_b - p_{\text{int}} \nabla \eta_b. \quad (20.16)$$

The topographic term is static whereas the other two terms are time dependent. External contributions arise from undulations in the free surface as well as the applied pressure. This contribution fluctuates due to motions occurring on the relatively rapid time scales associated with external gravity waves or atmospheric pressure fluctuations such as through synoptic weather patterns. Internal contributions arise from the relatively slow internal movements of density surfaces, such as from internal gravity waves or even slower motions due to advection and diffusion. The study from [McCabe et al. \(2006\)](#) pursue this decomposition of the topographic form stress as part of their analysis of flow around a headland.

20.3 Interfacial form stress

In this section we focus on the form stress acting at an interface within the fluid itself, which is known as the *interfacial form stress*. As part of this discussion we expose some of the common manipulations found when considering finite volume integrated momentum budgets, whereby we decompose the horizontal pressure gradient acceleration acting on an infinitesimal column of fluid within the layer, as depicted in Figure 20.5. These manipulations are analogous to those considered in Section 20.2 for the topographic and atmospheric form stresses.

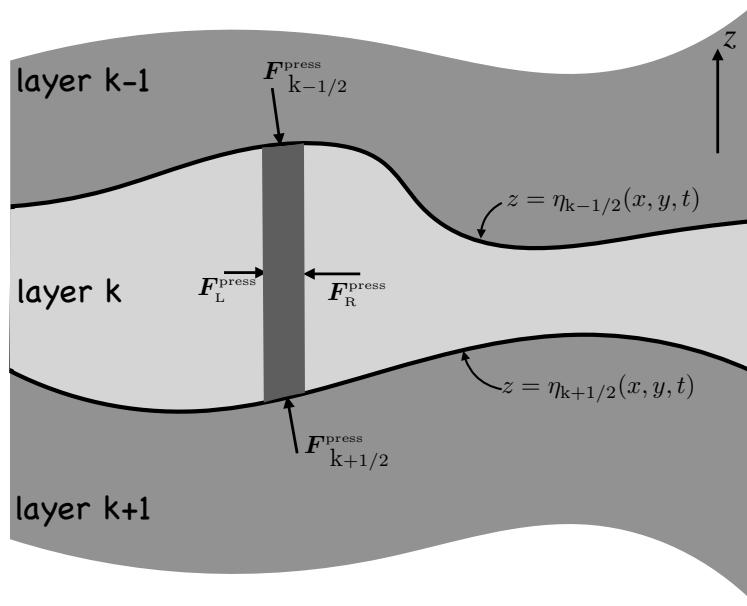


FIGURE 20.5: A schematic of the contact pressure force per area acting on the boundaries of a vertical column region within a fluid layer. The horizontal cross-sectional area of the column is depth independent. The interface at the lower boundary is at the vertical position $z = \eta_{k+1/2}$, and the upper interface is at $z = \eta_{k-1/2}$. In accordance with Newton's third law, pressures are continuous across each of the $\eta_{k\pm 1/2}$ layer interfaces so that the pressure forces are equal in magnitude yet oppositely directed on the opposite sides to the interfaces. The boundaries of the dark gray columnar region feel a contact pressure force acting inward. The left side of the column experiences a pressure p_L ; the right side experiences p_R ; the upper interface has a pressure $p_{k-1/2}$ acting between the layer $k - 1$ and layer k , and the lower interface has a pressure $p_{k+1/2}$ acting between the layer $k + 1$ and layer k . The interfacial form stress (IFS) is the name given to the horizontal pressure stress acting on the upper and lower layer interfaces. Through Newton's third law, the IFS imparted to layer k at the $z = \eta_{k-1/2}$ interface is equal and opposite to the IFS imparted to layer $k - 1$ at this same interface. The same holds for the IFS at the $k + 1/2$ interface. It is common to define the layers according to buoyancy (see Section 26.7) given its direct connection to pressure and dynamics. Even so, the ideas of pressure contact forces are generic and thus hold for layers defined arbitrarily.

20.3.1 Interfacial form stresses (IFS) transferred between layer

When studying the momentum of a column of fluid within a chosen layer, we need to compute the depth integrated zonal pressure gradient over a layer at a particular horizontal point

$$\text{layer integrated zonal pressure gradient} = - \int_{\eta_{k+1/2}}^{\eta_{k-1/2}} \frac{\partial p}{\partial x} dz, \quad (20.17)$$

where $z = \eta_{k-1/2}(x, y, t)$ is the vertical position for the top of the fluid layer and $z = \eta_{k+1/2}(x, y, t)$ is the vertical position for the bottom interface. If the layer integrated pressure

gradient points downgradient to the east, then pressure accelerates the column to the east.

Although the depth integrated pressure gradient expression (20.17) is straightforward to understand, we also find it useful to consider the complementary perspective by studying the contact force version of the pressure acceleration. Proceeding as in Section 20.2 for topographic and atmospheric form stresses, we make use of Leibniz' Rule for a fluid layer

$$-\int_{\eta_{k+1/2}}^{\eta_{k-1/2}} \frac{\partial p}{\partial x} dz = \underbrace{-\frac{\partial}{\partial x} \int_{\eta_{k+1/2}}^{\eta_{k-1/2}} p dz}_{\text{zonal deriv layer integrated pressure}} + \underbrace{\frac{\partial \eta_{k-1/2}}{\partial x} p_{k-1/2}}_{\text{IFS at } k-1/2 \text{ interface}} - \underbrace{\frac{\partial \eta_{k+1/2}}{\partial x} p_{k+1/2}}_{\text{IFS at } k+1/2 \text{ interface}}, \quad (20.18)$$

where we introduced the pressures acting at a point on the interfaces

$$p_{k-1/2} = p(x, y, z = \eta_{k-1/2}, t) \quad \text{and} \quad p_{k+1/2} = p(x, y, z = \eta_{k+1/2}, t). \quad (20.19)$$

The decomposition identifies the following three pressure contributions, analogous to the decomposition in Section 20.2 for the topographic and atmospheric form stresses.

- **ZONAL DERIVATIVE OF THE COLUMN INTEGRATED PRESSURE:** The first term arises from the zonal derivative of pressure across the vertical sides of the column within the layer

$$\text{zonal derivative of layer integrated pressure} = -\frac{\partial}{\partial x} \int_{\eta_{k+1/2}}^{\eta_{k-1/2}} p dz. \quad (20.20)$$

- **INTERFACIAL FORM STRESS AT UPPER INTERFACE:** The pressure at the $z = \eta_{k-1/2}$ interface is given by $p_{k-1/2}$. In the presence of a sloping interface, $\partial \eta_{k-1/2} / \partial x \neq 0$, this pressure imparts the following interfacial form stress (IFS) to layer-k:

$$\text{IFS on layer-k from the } \eta_{k-1/2} \text{ interface} = \frac{\partial \eta_{k-1/2}}{\partial x} p_{k-1/2}. \quad (20.21)$$

For example, if the upper layer interface slopes up to the east, $\partial \eta_{k-1/2} / \partial x > 0$, then the interfacial form stress provides a positive (eastward) zonal force to layer-k. In turn, through Newton's third law, the layer above, labelled k – 1, feels an interfacial form stress accelerating this layer to the west.

- **INTERFACIAL FORM STRESS AT LOWER INTERFACE:** The pressure, $p_{k+1/2}$ present at the $z = \eta_{k+1/2}$ interface imparts an interfacial form stress to layer-k given by

$$\text{IFS on layer-k from } \eta_{k+1/2} \text{ interface} = -\frac{\partial \eta_{k+1/2}}{\partial x} p_{k+1/2}. \quad (20.22)$$

For example, if the layer slopes down to the east, $\partial \eta_{k+1/2} / \partial x < 0$, then the interfacial form stress accelerates layer-k to the east. In turn, through Newton's third law, the interfacial form stress acts to accelerate the layer below, labelled k + 1, to the west.

Now apply the above to a column of ocean fluid, and extending the integration to include the full ocean column from the free upper surface to the solid earth bottom. Note that all the intermediate interfacial form stresses vanish in the depth integral, with this cancellation a result of Newton's third law. Hence, accumulation of the interfacial form stresses throughout the ocean column leaves only the interfacial form stress at the top and at the bottom, with those boundary form stresses arising from mechanical interactions with the atmosphere and solid earth as discussed in Section 20.2. This result was already encountered in a more general context of contact forces in Section 19.1. It will also be found in our analysis of the depth integrated axial angular momentum budget in Section 20.4.

20.3.2 Zonally integrated interfacial form stress

Besides studying the force acting on a column at a particular horizontal position, it is interesting to study the net zonal force acting on the layer. For pressure, we thus need to consider the zonal integral of the layer integrated zonal pressure gradient

$$-\int \left[\int_{\eta_{k+1/2}}^{\eta_{k-1/2}} \frac{\partial p}{\partial x} dz \right] dx = \int \left[-\frac{\partial}{\partial x} \int_{\eta_{k+1/2}}^{\eta_{k-1/2}} p dz + \frac{\partial \eta_{k-1/2}}{\partial x} p_{k-1/2} - \frac{\partial \eta_{k+1/2}}{\partial x} p_{k+1/2} \right] dx. \quad (20.23)$$

If the domain is zonally periodic or is bounded by sloping shorelines (see Figure 20.7 discussed in Section 20.4.7), then the first term vanishes so that the zonally integrated pressure acting on the layer arises just from the interfacial form stresses

$$-\int \left[\int_{\eta_{k+1/2}}^{\eta_{k-1/2}} \frac{\partial p}{\partial x} dz \right] dx = \int \left[\frac{\partial \eta_{k-1/2}}{\partial x} p_{k-1/2} - \frac{\partial \eta_{k+1/2}}{\partial x} p_{k+1/2} \right] dx. \quad (20.24)$$

This zonal integral is only affected by zonal anomalies for the layer vertical positions and pressures

$$-\int \left[\int_{\eta_{k+1/2}}^{\eta_{k-1/2}} \frac{\partial p}{\partial x} dz \right] dx = \int \left[\frac{\partial \eta'_{k-1/2}}{\partial x} p'_{k-1/2} - \frac{\partial \eta'_{k+1/2}}{\partial x} p'_{k+1/2} \right] dx \quad (20.25a)$$

$$= \int \left[-\eta'_{k-1/2} \frac{\partial p'_{k-1/2}}{\partial x} + \eta'_{k+1/2} \frac{\partial p'_{k+1/2}}{\partial x} \right] dx, \quad (20.25b)$$

where primes denote deviations from the zonal mean. We spend some time in Section 20.4.10 to prove that it is only the anomalies that matter for the zonally periodic channel, or for domains with sloping shorelines. Furthermore, note that for the second equality we introduced the alternative expressions for the form stresses afforded by zonal periodicity or zonal sloped shorelines. We offer some cautionary remarks in Section 20.4.9 regarding this second equality.

20.3.3 Comments

Interfacial form stress acts on any arbitrary surface drawn in a fluid. Interfaces defined by buoyancy surfaces make the connection between the general concepts presented here to geostrophic mechanics, and they do so given the connection between buoyancy slopes and thermal wind (Section 26.4.4). Most studies of interfacial form stress are thus concerned with isopycnal interfacial form stress, with a discussion given in Section 26.7.

20.4 Axial angular momentum budget for an ocean domain

We here develop the column integrated budget for axial angular momentum over a finite region of the ocean, such as shown in Figure 20.6. We then further specialize the budget by zonally integrating. The analysis exposes how a variety of physical processes affect torques that change the axial angular momentum of the ocean fluid, in particular those processes associated with form stress, surface mass transport, and boundary friction. The analysis thus brings together many of the concepts introduced in this chapter as well as the material on stress from Chapter 19. For some generality we make use of spherical coordinates, though doing so offers only a modest degree of extra details beyond Cartesian coordinates. Although here focused on the ocean, many of the

concepts and methods are directly relevant to a study of atmospheric axial angular momentum such as that lucidly discussed in Section 10.3 of [Holton \(1992\)](#).

Finally, recall that axial angular momentum is closely connected to the zonal linear momentum, with details for the connection provided in our study of geophysical particle dynamics in Sections 11.7 and 11.8. We here choose to study axial angular momentum since it has a simpler budget (equation (20.26) discussed below) than zonal linear momentum. The simpler conservation budget follows since axial angular momentum is directly connected to the axial symmetry of the rotating spherical planet (Sections 11.7 and 18.4).

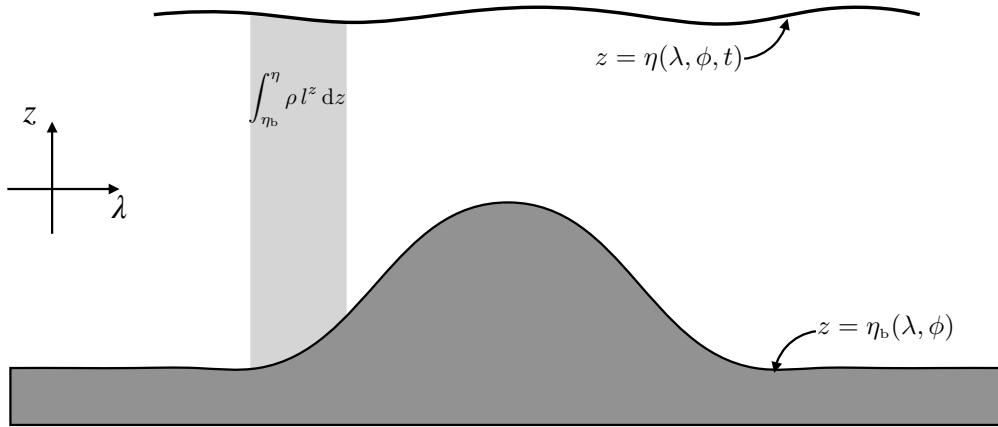


FIGURE 20.6: Schematic of the axial angular momentum for a fluid column, here depicted moving in an ocean with a topographic bump. In Section 20.4 we develop the budget for the depth and zonal integrated axial angular momentum in the ocean, where we see that the axial angular momentum is affected by a variety of boundary processes as well as interior transports and pressures.

20.4.1 Anticipating the budget

Before diving into the mathematical formulation, let us use some of the understanding gleaned from earlier sections of this chapter to anticipate the basic results. Doing so offers us a framework to guide the maths, and to double check that the maths indeed renders a physically sensible budget.

For this purpose, consider a column of fluid such as shown in Figure 20.6. The forces acting on that column arise from contact forces (pressure, kinetic, and frictional stresses) acting on the boundary (sides, top, and bottom), and body forces acting throughout the column (from effective gravity and Coriolis). There are further avenues for momentum to be transported across the ocean surface as part of the mass transported by rain, evaporation, and rivers. Each of these forces contribute a torque to the fluid column computed relative to the earth's rotational axis, thus modifying the axial angular momentum of the fluid column. In the following development, we mathematically express the variety of forces and corresponding torques, thus building up the axial angular momentum budget.

20.4.2 Axial angular momentum

The axial angular momentum budget for a fluid element follows that developed in Section 18.4, here written with zonal friction

$$\rho \frac{Dl^z}{Dt} = -\frac{\partial p}{\partial \lambda} + r_\perp \rho F^\lambda, \quad (20.26)$$

where

$$l^z = r_{\perp} (u + r_{\perp} \Omega) \quad (20.27)$$

is the axial angular momentum per unit mass, and

$$r_{\perp} = r \cos \phi \quad (20.28)$$

is the distance to the polar rotation axis (the moment arm). Use of the Eulerian form of mass conservation (equation (15.6)) leads to the Eulerian flux-form budget

$$\frac{\partial(\rho l^z)}{\partial t} + \nabla \cdot (\rho \mathbf{v} l^z) = -\frac{\partial p}{\partial \lambda} + r_{\perp} \rho F^{\lambda}, \quad (20.29)$$

with $l^z \rho dz$ the angular momentum per unit horizontal area. We use this form for the budget to develop the depth integrated axial angular momentum budget.

20.4.3 Depth integrated budget

Vertically integrating equation (20.29) over a column of ocean fluid renders a budget for the column-integrated axial angular momentum. This derivation requires Leibniz's Rule to reach the following identities that expose boundary contributions

$$\int_{\eta_b}^{\eta} \frac{\partial(\rho l^z)}{\partial t} dz = \frac{\partial}{\partial t} \left[\int_{\eta_b}^{\eta} \rho l^z dz \right] - \left[\rho l^z \frac{\partial \eta}{\partial t} \right]_{z=\eta} \quad (20.30a)$$

$$\int_{\eta_b}^{\eta} \nabla_z \cdot (\rho \mathbf{u} l^z) dz = \nabla_z \cdot \left[\int_{\eta_b}^{\eta} \rho \mathbf{u} l^z dz \right] - [\rho l^z \mathbf{u} \cdot \nabla \eta]_{z=\eta} + [\rho l^z \mathbf{u} \cdot \nabla \eta_b]_{z=\eta_b} \quad (20.30b)$$

$$\int_{\eta_b}^{\eta} \frac{\partial(\rho w l^z)}{\partial z} dz = [w \rho l^z]_{z=\eta} - [w \rho l^z]_{z=\eta_b}. \quad (20.30c)$$

The surface kinematic boundary condition (15.79) and bottom kinematic boundary condition (15.41) allow us to reach a reasonably tidy expression

$$\frac{\partial}{\partial t} \left[\int_{\eta_b}^{\eta} l^z \rho dz \right] = [l^z Q_m]_{z=\eta} - \nabla_z \cdot \left[\int_{\eta_b}^{\eta} l^z \mathbf{u} \rho dz \right] + \int_{\eta_b}^{\eta} \left[-\frac{\partial p}{\partial \lambda} + r_{\perp} \rho F^{\lambda} \right] dz. \quad (20.31)$$

The budget (20.31) says that the depth integrated axial angular momentum per horizontal area in a horizontally fixed fluid column has a time tendency (left hand side) arising from the convergence of horizontal advection of axial angular momentum plus torques due to surface boundary mass fluxes, depth integrated zonal pressure gradients, and depth integrated irreversible stresses. At this point we return to the discussion in Section 20.4.1 and verify that the mathematical expression of the budget meets our expectations based on prior understanding of the physical principles.

20.4.4 Atmospheric and topographic form stresses

We can further unpack the contribution from pressure in the budget (20.31) by making use of Leibniz's rule to write

$$\int_{\eta_b}^{\eta} \frac{\partial p}{\partial \lambda} dz = \frac{\partial P}{\partial \lambda} - p_a \frac{\partial \eta}{\partial \lambda} + p_b \frac{\partial \eta_b}{\partial \lambda}, \quad (20.32a)$$

where

$$P = \int_{\eta_b}^{\eta} p dz \quad (20.33)$$

is the depth-integrated pressure field, p_a is the pressure applied to the ocean at its surface, $z = \eta$, and p_b is the pressure at the ocean bottom, $z = \eta_b$. We studied this decomposition of the pressure force in Section 20.2. Again, we see that the depth integrated zonal pressure gradient has been decomposed into three terms: (i) zonal pressure differences integrated across the depth of the column, (ii) form stress imparted to the ocean from the atmospheric pressure, (iii) form stress imparted by the solid earth bottom topography onto the ocean.

20.4.5 Turbulent stresses at the surface and bottom

For turbulent stresses, we focus on the vertical transfer of zonal momentum arising from the vertical shear of horizontal stresses

$$\rho F^\lambda = \frac{\partial \tau^\lambda}{\partial z}, \quad (20.34)$$

where τ^λ is the zonal component to the stress vector.³ When integrated vertically over an ocean column, $\int_{\eta_b}^{\eta} \rho F^\lambda dz$, this friction arises from stresses acting in the ocean surface and bottom boundary/Ekman layers (Chapter 28), where the stress arises from turbulent motions that transfer momentum vertically through these layers.

To slightly simplify the contribution from friction, we assume that the axial moment-arm is well approximated by its value at the ocean surface

$$r_\perp = r \cos \phi = (z + R) \cos \phi \approx R \cos \phi = R_\perp, \quad (20.35)$$

with this assumption holding for the shallow fluid approximation built into the hydrostatic primitive equations discussed in Section 23.1. This assumption allows us to write the frictional contribution to the angular momentum budget (20.31) in the form

$$\int_{\eta_b}^{\eta} r_\perp \rho F^\lambda dz \approx R_\perp \int_{\eta_b}^{\eta} \rho F^\lambda dz = R_\perp (\tau_a^\lambda - \tau_b^\lambda). \quad (20.36)$$

The final expression introduced τ_a^λ , which is the zonal component to the stress acting on the ocean surface imparted through irreversible interactions between the ocean and the overlying atmosphere and/or ice. The signs are such that $\tau_a^\lambda > 0$ transfers an eastward momentum to the ocean such as via a westerly wind stress. Likewise, the stress τ_b^λ is the zonal stress at the ocean bottom imparted through irreversible interactions between the ocean and the solid-earth. The signs are such that $\tau_b^\lambda > 0$ reflects the transfer of eastward momentum from the ocean to the solid-earth, or conversely the transfer of westward momentum from the earth to the ocean. The net contribution from vertical friction is thus given by the moment arm, R_\perp , multiplied by the difference in boundary stresses.

20.4.6 Summary budget for column integrated axial angular momentum

Bringing all the pieces together leads to the depth integrated axial angular momentum budget

$$\frac{\partial}{\partial t} \left[\int_{\eta_b}^{\eta} l^z \rho dz \right] = -\nabla_z \cdot \left[\int_{\eta_b}^{\eta} l^z \mathbf{u} \rho dz \right] - \frac{\partial P}{\partial \lambda} + [l^z Q_m]_{z=\eta} + p_a \frac{\partial \eta}{\partial \lambda} - p_b \frac{\partial \eta_b}{\partial \lambda} + R_\perp (\tau_a^\lambda - \tau_b^\lambda). \quad (20.37)$$

Other than assuming the approximate form (20.35) for the moment-arm, and choosing a specific form of the frictional stress given by equation (20.34), this result is an exact budget for the axial angular momentum of a column of ocean fluid.

³There are other turbulent terms associated with interior Reynold stresses arising from horizontal shears. We omit these terms for the present analysis since they are generally smaller than stresses arising from vertical strains, and in particular from turbulent stresses associated with surface and bottom boundary processes.

Removing zonal means

We further isolate the processes contributing to the budget (20.37) by introducing the zonal mean operator

$$\bar{A} \equiv \frac{1}{L(\phi)} \int A d\lambda, \quad (20.38)$$

where

$$L(\phi) = (R \cos \phi) \Delta \lambda = R_{\perp} \Delta \lambda \quad (20.39)$$

is the zonal length of the domain as a function of latitude, ϕ , and $\Delta \lambda$ is the zonal extent of the domain in radians. For a domain that circles the planet, then $\Delta \lambda = 2\pi$. Other domains are possible when considering idealized theoretical models as in Figure 20.7. The corresponding zonal anomalies to the depth integrated pressure, sea surface height, and bottom topography are thus given by

$$P' = P - \bar{P} \quad \text{and} \quad \eta' = \eta - \bar{\eta} \quad \text{and} \quad \eta'_b = \eta_b - \bar{\eta}_b, \quad (20.40)$$

in which case equation (20.37) takes the form

$$\frac{\partial}{\partial t} \left[\int_{\eta_b}^{\eta} l^z \rho dz \right] = -\nabla_z \cdot \left[\int_{\eta_b}^{\eta} l^z \mathbf{u} \rho dz \right] - \frac{\partial P'}{\partial \lambda} + [l^z Q_m]_{z=\eta} + p_a \frac{\partial \eta'}{\partial \lambda} - p_b \frac{\partial \eta'_b}{\partial \lambda} + R_{\perp} (\tau_a^{\lambda} - \tau_b^{\lambda}). \quad (20.41)$$

Steady state balance

Steady state balances are of particular interest when studying the large-scale low frequency circulation. A steady state holds for the angular momentum budget (20.41) so long as following balance is maintained

$$\nabla_z \cdot \left[\int_{\eta_b}^{\eta} l^z \mathbf{u} \rho dz \right] = -\frac{\partial P'}{\partial \lambda} + [l^z Q_m]_{z=\eta} + p_a \frac{\partial \eta'}{\partial \lambda} - p_b \frac{\partial \eta'_b}{\partial \lambda} + R_{\perp} (\tau_a^{\lambda} - \tau_b^{\lambda}). \quad (20.42)$$

Consequently, a steady state is realized if the horizontal divergence of depth integrated axial angular momentum advection (left hand side) is balanced by torques created by the variety of physical processes on the right hand side. We further examine these physical processes by studying the zonally integrated budget.

20.4.7 Zonal domain topology/geometry and zonal integration

Zonal integration and zonal averaging offer a common means to filter the three dimensional equations, particularly in the atmosphere where zonal motions are much stronger than meridional given the earth's rotation and the differential solar heating of the planet. Additionally, zonal averaging is of particular interest in the Southern Ocean, where the Drake Passage latitudes offer a zonally unbounded domain for ocean circulation. Even for zonally bounded ocean domains where gyre circulations occur, it is of interest to zonally integrate across the domain to study balances leading to meridional motion across the chosen latitude.

We are thus motivated to integrate zonally across the full extent of the domain, with the resulting boundary contributions dependent on the geometry and topology of the domain. We illustrate three canonical domains in Figure 20.7. For the zonally periodic domain, the zonal integral of any zonal derivative vanishes so that

$$\oint \frac{\partial P}{\partial \lambda} d\lambda = 0, \quad (20.43)$$

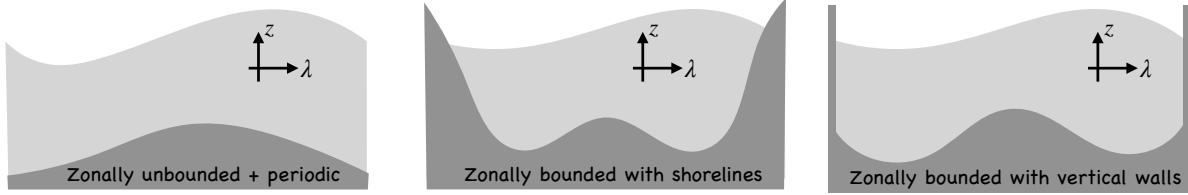


FIGURE 20.7: Three canonical zonal topologies/geometries considered in the study of fluid flow, particularly ocean flows. The light portion of each panel represents the fluid whereas the darker portion is the solid earth bottom topography. Left panel: zonally unbounded and periodic channel. Here, the topography, surface boundary forcing, and flow are zonally periodic. Middle panel: zonally bounded region where the zonal bounds occur along sloping shorelines at which the fluid thickness vanishes. The horizontal position of the vanishing thickness is time dependent since the fluid can move up and down the shoreline. Right panel: zonally bounded region where the fluid encounters a vertical sidewall so that the horizontal position of the fluid boundary is fixed. Fixed vertical sidewall boundaries are commonly found in numerical model simulations and yet they are uncommon in Nature.

where we write $\oint d\lambda$ for integration over a zonally periodic domain. Additionally, zonal derivatives vanish for a zonally bounded domain with sloping shorelines

$$\int_{\text{shorelines}} \frac{\partial P}{\partial \lambda} d\lambda = 0. \quad (20.44)$$

The reason this integral vanishes is that quantities such as P are depth integrated, and so $P = 0$ at the shoreline edge merely since the layer thickness vanishes at the shoreline edge. The same result holds for any other depth integrated quantity, such as the axial angular momentum.

It is only for the zonally bounded domain with vertical sidewalls (third panel in Figure 20.7) that we are unable to drop the zonal integral of zonal derivatives. We observe that vertical sidewalls are common for numerical models and many theories of the ocean circulation. However, vertical sidewalls are the exception in Nature. For this reason, in the following we focus on the more geophysically relevant periodic configuration and the sloping shoreline configuration.

20.4.8 Steady domain integrated balance

Consider the area integral of the steady state balance (20.42) over the full ocean domain that is either periodic and/or has sloping side boundaries. In this case the divergence of the angular momentum transport integrates to zero, so that we are left with the balance

$$\int_{\phi_s}^{\phi_n} \left(\int_{\lambda_w(\phi)}^{\lambda_e(\phi)} \left[[l^z Q_m]_{z=\eta} + p'_a \frac{\partial \eta'}{\partial \lambda} - p'_b \frac{\partial \eta'_b}{\partial \lambda} + R_\perp (\tau_a^\lambda - \tau_b^\lambda) \right] d\lambda \right) R^2 \cos \phi d\phi = 0. \quad (20.45)$$

In computing the area integral, we chose to first integrate over the longitudinal domain, $\lambda_w(\phi) \leq \lambda \leq \lambda_e(\phi)$, which is a function of latitude, and then to integrate over the full latitudinal domain, $\phi_s \leq \phi \leq \phi_n$. In most applications the surface mass term, $[l^z Q_m]_{z=\eta}$, is smaller than the other terms, in which case the balance is between the boundary form stresses and the boundary turbulent stresses.

In the above balance, we introduced the zonal anomalies for the applied surface pressure and the bottom pressure

$$p'_a(\lambda, \phi) = p_a(\lambda, \phi) - \bar{p}_a(\phi) \quad \text{and} \quad p'_b(\lambda, \phi) = p_b(\lambda, \phi) - \bar{p}_b(\phi). \quad (20.46)$$

We can introduce these anomalous fields since their zonal averages do not contribute to the budgets in either the periodic or sloping shoreline domains. To verify this property, note that

$$\int \bar{p}_a \frac{\partial \eta'}{\partial \lambda} d\lambda = \int \frac{\partial(\bar{p}_a \eta')}{\partial \lambda} d\lambda = 0. \quad (20.47)$$

For a periodic domain this term vanishes by inspection. For a zonally bounded domain with a sloping shoreline, it also vanishes since $\eta' = 0$ at the edge of the shoreline. Likewise, the bottom pressure term satisfies

$$\int \bar{p}_b \frac{\partial \eta'_b}{\partial \lambda} d\lambda = \int \frac{\partial(\bar{p}_b \eta'_b)}{\partial \lambda} d\lambda = 0, \quad (20.48)$$

which follows either by periodicity or since $\eta'_b = 0$ along the edge of a sloping shoreline. In conclusion, we see that it is only the zonal anomalies of the atmospheric and bottom pressures, and free surface and bottom topography, that impact the zonal mean zonal momentum balance (20.50) for the periodic and sloping shoreline domains.

20.4.9 Remarks on a modified expression for the form stress

We can further exploit the symmetries of the periodic and sloping shoreline domains by writing the form stresses in equation (20.50) in the alternative form

$$\int \left[p'_a \frac{\partial \eta'}{\partial \lambda} - p'_b \frac{\partial \eta'_b}{\partial \lambda} \right] d\lambda = \int \left[-\frac{\partial p'_a}{\partial \lambda} \eta' + \frac{\partial p'_b}{\partial \lambda} \eta'_b \right] d\lambda. \quad (20.49)$$

Diagnostics of zonally integrated form stress can be more convenient using one form or the other, depending on dataset or numerical model framework. We have a choice since the zonal integral is the same, and that freedom is afforded since the spatial integral removes local information that appears as a total zonal derivative. However, we offer two caveats in this regard.

- The identity (20.49) does not hold for the bounded domain with vertical sidewalls (third panel of Figure 20.7). If working in such a domain and if one chooses to study patterns based on the right hand side expression, then its zonal integral will not agree with that of the left hand side. Correspondingly, physical interpretations based on the right hand side are questionable.
- Although the zonal integrals in equation (20.49) agree for the periodic domain and sloping shoreline domain, there is no local identity between terms on the left hand side and right hand side. So if one wishes to make a statement about patterns of local form stresses, then it is necessary to return to the fundamental formulation of form stress appearing on the left hand side of equation (20.49).

20.4.10 Steady zonal and depth integrated budget

Now consider just a zonal integral, again over a domain that has a either periodic zonal topology or sloping shoreline geometry (Figure 20.7). In this case we are left with the zonal and depth integrated steady angular momentum budget

$$\frac{1}{R \cos \phi} \int \frac{\partial}{\partial \phi} \left[\int_{\eta_b}^{\eta} l^z v \rho dz \right] d\lambda = \int \left[[l^z Q_m]_{z=\eta} + p'_a \frac{\partial \eta'}{\partial \lambda} - p'_b \frac{\partial \eta'_b}{\partial \lambda} + R_\perp (\tau_a^\lambda - \tau_b^\lambda) \right] d\lambda. \quad (20.50)$$

The meridional divergence of the advective transport of angular momentum is balanced, in the steady state, by the boundary terms on the right hand side.

20.4.11 Southern Ocean balances

Under certain cases the primary balance in equation (20.50) is between the form stress and boundary turbulent stress, whereby

$$\oint \left[p'_a \frac{\partial \eta'}{\partial \lambda} + R_{\perp} \tau_a^{\lambda} \right] d\lambda \approx \oint \left[+p'_b \frac{\partial \eta'_b}{\partial \lambda} + R_{\perp} \tau_b^{\lambda} \right] d\lambda. \quad (20.51)$$

For much of the large-scale Southern Ocean circulation, the primary balance is even simpler: it is a balance between surface wind stress and topographic form stress

$$\oint \tau_a^{\lambda} d\lambda \approx \oint p'_b \frac{1}{R_{\perp}} \frac{\partial \eta'_b}{\partial \lambda} d\lambda = - \oint \eta'_b \frac{1}{R_{\perp}} \frac{\partial p'_b}{\partial \lambda} d\lambda, \quad (20.52)$$

with this balance exemplified in Figure 20.8. We now state in words what this balance means in the presence of a net eastward wind stress, $\oint \tau_a^{\lambda} d\lambda > 0$. The equivalent expressions on the right hand side allow complementary perspectives.

- For the first equality we see that balance is realized if on the upwind side of a topographic bump there is an anomalously high bottom pressure, with the opposite on the downwind side. Correspondingly, there is a net westward topographic form stress imparted by the solid earth onto the ocean that balances the eastward surface wind stress imparted by the atmosphere onto the ocean.
- The second equality in equation (20.52) reveals that for $\eta'_b > 0$ (topographic ridge), a steady angular momentum balance is maintained so long as the bottom pressure decreases across the ridge from west to east, just as depicted in Figure 20.8.

20.4.12 Topographic form stress and ocean gyres

Hughes and de Cueves (2001) are notable for having emphasized the importance of bottom topographic form stress, and the associated bottom pressure torque appearing in the vorticity equation (see Section 36.10). The key role for topographic form stress has long been appreciated for channel flows since the work of *Munk and Palmén* (1951) who studied the Southern Ocean force balances. However, *Hughes and de Cueves* (2001) showed that it is central even for steady ocean gyre circulations when allowing for sloping sides rather than vertical sides. Hence, sloping sides for the gyre domain allow for bottom form stress and bottom pressure torque to balance wind stress and wind stress curl in the steady momentum and steady vorticity balances, just as for zonal channel flows. As a result, topographic form stress and bottom pressure torques allow for a mostly inviscid balance in ocean gyres much like for the Southern Ocean. The critical role of sloping sides was not appreciated by the pioneering studies of ocean gyres from *Stommel* (1948) and *Munk* (1950), who used vertical sidewalls and relied on friction to close the vorticity budget. That is, vertical sides hide the role of topographic form stresses, thus requiring frictional side stresses to balance the wind stress. We return to these points in Section 35.9, in particular in Section 35.9.5, when studying gyre circulations in a shallow water fluid.

20.4.13 Further study

Elements of this section are based on *Hughes and de Cueves* (2001) and the analogous discussion of the global atmospheric axial angular momentum budget developed in Section 10.3 of *Holton* (1992).

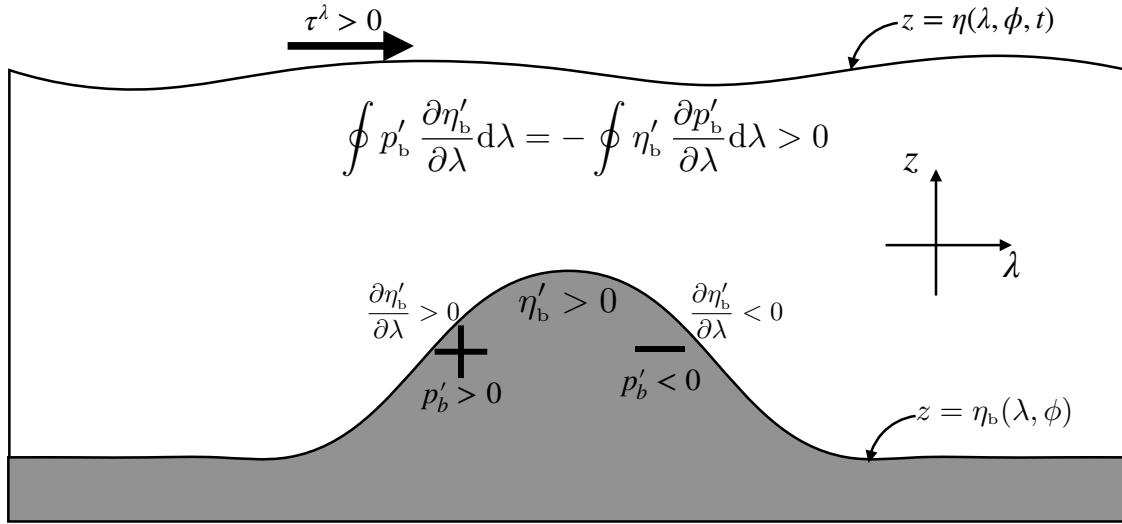


FIGURE 20.8: Depicting the balance between zonal wind stress and topographic form stress for a southern hemisphere zonally periodic ocean channel. For an eastward wind stress, the flow reaches a steady state if the upwind side of a topographic bump sees an anomalously high bottom pressure whereas on the downwind side it is anomalously low. The form stress imparted by the ocean onto the solid earth is eastward since the bottom pressure is higher in the west and lower in the east. Conversely, Newton's third law tells us that the topographic form stress imparted by the solid earth onto the ocean is westward. In this manner, the eastward force imparted by the atmosphere onto the ocean through wind stress is balanced by a westward topographic form stress imparted by the solid earth onto the ocean. Furthermore, the eastward bottom pressure gradient leads to a northward depth integrated geostrophic transport above the southern hemisphere ridge. Signs are swapped when flow moves over a depression, in which a westward bottom pressure gradient leads to southward geostrophic transport over the depression. We revisit these dynamical processes for the shallow water fluid in Figure 32.8.

Straub (1993) provides an analogous analysis with a focus on the Southern Ocean. The physical processes establishing the balances noted for the Southern Ocean remain under investigation, with further resources to the literature including Section 21.7 of *Vallis (2017)*, and the reviews by *Rintoul et al. (2001)*, *Rintoul and Naveira Garabato (2013)*, and *Rintoul (2018)*. We further revisit this balance in Section 32.6 for the shallow water system.



Equilibrium thermodynamics

Thermodynamics is a phenomenological discipline whose fundamentals lie in statistical mechanics. In this chapter we develop elements of equilibrium thermodynamics relevant to a static and single-phase multi-component fluid in or near equilibrium. We are compelled to examine thermodynamics since there is a coupling of thermodynamics to dynamics through the density of a fluid element. That is, mass density is a function of temperature, matter concentration, and pressure (Section 25.3). Hence, to understand fluid motion we need to understand how temperature evolves within the fluid. We are also interested in thermodynamics for its own right, in particular to understand how energy is transformed from one form to another and how energy flows through the fluid environment. Such questions are at the heart of ongoing research into human-induced climate warming.

READER'S GUIDE FOR THIS CHAPTER

This chapter provides a baseline understanding of equilibrium thermodynamics, with these ideas applied in Chapter 22 to moving fluid elements that are considered locally equilibrated systems. Our treatment follows the classic approach of Gibbs as presented in books from *Callen* (1985) and *Reif* (1965). Mathematically, we make use of partial differential calculus at the level of Chapter 2.

For some reflections on the enduring and subtle nature of thermodynamics, consider the following comments from two 20th century giants of physics.

A theory is the more impressive the greater the simplicity of its premises, the more different kinds of things it relates, and the more extended is its area of applicability. Therefore the deep impression which classical thermodynamics made upon me. It is the only physical theory of universal content concerning which I am convinced that, within the framework of the applicability of its basic concepts, it will never be overthrown. *Einstein* (1949)

Thermodynamics is a funny subject. The first time you go through it, you do not understand it at all. The second time you go through it, you think you understand it, except for one or two points. The third time you go through it, you know you do not understand it, but by that time you are so used to the subject it does not bother you anymore. *Attributed to Arnold Sommerfeld, unknown source*

21.1 Conceptual basics	425
21.1.1 Thermodynamic equilibrium versus steady state	425
21.1.2 Exchanges between thermodynamic systems	425
21.1.3 Quasi-static processes	426

21.1.4	Extensive and intensive properties	426
21.1.5	Thermodynamic configuration space has no metric	427
21.2	First law for materially closed systems	427
21.2.1	Basic expression of the first law of thermodynamics	428
21.2.2	The nature of working and heating	428
21.2.3	Mechanical work	429
21.2.4	Quasi-static transfer of thermal energy and entropy	429
21.2.5	Fundamental thermodynamic relation	430
21.2.6	Entropy and reversible/irreversible processes	430
21.3	First law for materially open systems	431
21.3.1	Constituent masses and matter concentrations	431
21.3.2	Scaling for homogeneous functions	432
21.3.3	Chemical work and the Gibbs-Duhem relation	433
21.3.4	Partial derivatives	433
21.3.5	Equations of state	434
21.3.6	Fundamental thermodynamic relation per unit mass	434
21.3.7	Seawater as a binary fluid	434
21.3.8	Further study	435
21.4	Thermodynamic potentials	435
21.4.1	Internal energy	436
21.4.2	Entropy	436
21.4.3	Enthalpy	437
21.4.4	Helmholtz free energy	438
21.4.5	Gibbs potential	438
21.4.6	Seawater chemical potential in terms of the Gibbs potential	439
21.5	Response functions	440
21.5.1	Heat capacities	440
21.5.2	Thermal expansion, haline contraction, and sound speed	441
21.6	Maxwell relations for homogeneous fluids	442
21.6.1	Maxwell relation from internal energy	442
21.6.2	Summary of the Maxwell relations	442
21.7	Adiabatic lapse rate	443
21.7.1	Isentropic rearrangement	443
21.7.2	Thermodynamic formulation	443
21.7.3	Adiabatic lapse rate for pressure changes	444
21.7.4	Adiabatic lapse rate for height changes	444
21.7.5	Further study	444
21.8	Potential temperature	445
21.8.1	Adiabatic temperature changes	445
21.8.2	Defining the potential temperature	445
21.8.3	Potential temperature and specific entropy	446
21.9	A simple ideal gas atmosphere	448
21.9.1	Equation of state	448
21.9.2	Internal energy	449
21.9.3	Heat capacity	449
21.9.4	Enthalpy	450
21.9.5	Thermal expansion coefficient	450
21.9.6	Fundamental thermodynamic relations	450
21.9.7	Isothermal compressibility	451
21.9.8	Sound speed	451
21.9.9	Adiabatic lapse rate	451
21.9.10	Geopotential thickness	452
21.9.11	Potential temperature	453
21.9.12	Further study	454

21.1 Conceptual basics

In the study of thermodynamics it is important to characterize how a physical system interacts with its surrounding environment through mechanical, thermal, and material interactions and exchanges. Tiny fluid elements, and their accumulation into finite fluid regions, constitute the physical systems we are concerned with here (see Section 13.2.4). Furthermore, we are concerned with systems that interact mechanically with their surroundings through pressure forces so that the systems are mechanically open. However, we find it useful to distinguish how a physical system interacts thermally and materially as well as the time scales over which these interactions take place.

21.1.1 Thermodynamic equilibrium versus steady state

Equilibrium thermodynamics is the study of systems in *thermodynamic equilibrium* and how they transit from one thermodynamic equilibrium to another through quasi-static processes. At a basic level, a system in thermodynamic equilibrium will remain in that state for all time, unless something acts to move it to another state. Note that “for all time” is a loaded term. More precisely, we mean “for a time extremely long compared to any time scale relevant to the physical system”, with such time scales considered in Section 21.1.3

A system in thermodynamic equilibrium is a steady state since there are no time changes (again, no time changes over time scales relevant to the system). However, a steady state need not be a thermodynamic equilibrium. For example, consider a region of fluid with nonzero heat fluxes yet with no heat flux convergence so the temperature of the region does not change. A temporally constant temperature is a signature of a steady state. However, the flow of heat is the sign of thermodynamic disequilibrium. The distinction is sometimes subtle always important.

21.1.2 Exchanges between thermodynamic systems

We distinguish the following exchanges between a thermodynamic system (e.g., a fluid element) and its surrounding environment.

- THERMALLY OPEN (DIABATIC) AND MATERIALLY OPEN: An open physical system exchanges matter, thermal energy, and mechanical forces with its surrounding environment. All real fluid elements are open in this manner.
- THERMALLY OPEN (DIABATIC) AND MATERIALLY CLOSED: We have occasion to consider a fluid element that is mechanically and thermally open yet materially closed. Such fluid elements maintain a fixed matter content yet exchange thermal and mechanical energy with their surrounding environment.
- THERMALLY CLOSED (ADIABATIC) AND MATERIALLY CLOSED: We sometimes consider a fluid element that is both materially and thermally closed and yet mechanically open. Such *material fluid parcels* (Section 13.2) maintain fixed matter and thermal properties yet move according to imbalances in mechanical forces. A *perfect fluid* is comprised of material fluid parcels.

Each of the above interactions is mechanically open, so that the fluid element feels the pressure of the surrounding environment. For pedagogical purposes we first study the thermodynamics of fluid elements that are thermally open yet materially closed and then extend to fluid elements

that are both thermally and materially open. As a somewhat overloaded terminology, “adiabatic” in fluid mechanics is often used for a system that is *both* thermally closed *and* materially closed. However, we maintain the distinction in our treatment to maintain consistency with the physics literature.

21.1.3 Quasi-static processes

In our study of thermodynamics for moving fluid systems in Chapter 22, we make use of the huge time scale separation between molecular degrees of freedom and macroscopic degrees of freedom. We are thus led to define a *quasi-static process* as one that proceeds via an arbitrary number of intermediate equilibrium states. In fact, as formally defined (e.g., Section 4.2 of [Callen \(1985\)](#)), a quasi-static process is a locus of equilibrium states that can only be approximated by realistic processes occurring through steps that are each arbitrarily close to thermodynamic equilibrium. By formulating the notion of a quasi-static process, we are afforded a precise definition for intensive properties such as its temperature, pressure, and chemical potential, whereas intensive properties are not well defined for systems out of thermodynamic equilibrium.

A quasi-static process is a relevant idealization for our study of macroscopic fluid motion. We recognize this point by considering the time scale separation between macroscopic and microscopic motions. Recall from Section 12.3.5 that the time between molecular collisions is on the order of $t_{\text{collision}} \approx 10^{-9}$ s, with this time scale setting the time scale for molecular equilibration. In contrast, typical macroscopic processes occur on time increments on the order of $\Delta\tau_{\text{macro}} \sim 1$ s. Hence, we can safely assume that any macroscopic process of interest for our study evolves through time increments that are much longer than the molecular collision time

$$\Delta\tau_{\text{macro}} \gg t_{\text{collision}} \approx 10^{-9} \text{ s.} \quad (21.1)$$

We are thus led to the following conceptual framework for our treatment of fluid mechanics. Namely, we assume that macroscopic fluid motion occurs through a series of thermodynamically equilibrated states; i.e., the fluid evolves quasi-statically. The justification for this foundational assumption is that the relevant time scales for macroscopic fluid motions are many orders of magnitude longer than the molecular time scales that bring about local thermodynamic equilibrium.

Some authors do not make a distinction between a quasi-static process and a reversible process. Our treatment follows [Callen \(1985\)](#) (see his sections 4.2 and 4.3) and [Reif \(1965\)](#) (see his sections 2.9 and 2.10), whereby reversible processes are subsets of quasi-static processes. Namely, a reversible process is a quasi-static process that occurs without net entropy change (see Section 21.2.6), and yet not all quasi-static processes are reversible.

21.1.4 Extensive and intensive properties

In describing thermodynamic systems we characterize properties according to whether they are *extensive* or *intensive*.¹ Extensive properties scale with the size of the system, such as mass, internal energy and entropy. Since we can count the amount of an extensive property, it is exchanged between thermodynamic systems and we can form budget equations as in Section 16.3. Correspondingly, extensive properties scale with a power 1 with the size of the system. In contrast, *Intensive* properties do not scale with the size of a system, with temperature and pressure the canonical examples. Intensive properties describe possible gradients between systems and at thermodynamic equilibrium the intensive variables are the same in all the connected systems. Intensive properties

¹We also study this characterization in Section 16.3.1 when formulating budget equations.

are scale invariant and thus scale with power 0 as the size of the system changes. Extensive and intensive properties come as conjugate pairs in thermodynamics in that intensive properties always multiply their conjugate extensive property (e.g., (p, V) and (T, S)) when appearing in the various forms of the first law of thermodynamics.

In this chapter, extensive properties are labeled with a superscript e (except for the mass and volume), with this label *not* a tensor index. In Section 21.3 we introduce the internal energy per mass and entropy per mass, converting the extensive properties to their specific form in which case we drop the e superscript. It is the specific form of extensive properties that provides a straightforward transfer to the study of a fluid continua.

21.1.5 Thermodynamic configuration space has no metric

In this chapter we are not concerned with points in geographical space. Rather, we characterize a physical system through specifying a suite of thermodynamic properties. These properties define a point within thermodynamic configuration space. Even though one often finds diagrams with orthogonal axes defined by values of thermodynamic properties, there is actually no notion of distance or angle between points in thermodynamic configuration space since there is no metric structure in this space. Mathematically, we say that thermodynamic configuration space comprises a *topological manifold*. A summary of the mathematical structure of equilibrium thermodynamics can be found in [this online tutorial from Salamon et al.](#) We have slightly more to say on this topic when studying the mathematics of watermass configuration space (also a topological manifold) in Section 52.2.1.

Even though we emphasize the absence of a metric structure in thermodynamic configuration space, there are some formulations that do introduce a metric. In so doing, these formulations transform the topological space to a Riemannian manifold. However, there are tradeoffs to doing so, with [Andresen et al. \(1988\)](#) offering a survey. Here, we follow the traditional approach of Gibbs as articulated in the books from [Callen \(1985\)](#) and [Reif \(1965\)](#).

21.2 First law for materially closed systems

As discussed in Chapter 12, there are a huge number of microscopic (molecular) degrees of freedom that are averaged over when describing a fluid as a continuous media. *Internal energy* embodies the energy of microscopic degrees of freedom not explicitly considered in a macroscopic continuum treatment. Internal energy arises from the translational kinetic energy of molecular motion, together with their internal degrees of freedom associated with rotation and vibration as well as intermolecular forces between molecules. When accounting for the total energy of a fluid system, we must include its internal energy due to microscopic degrees of freedom as well as the mechanical energy (kinetic and gravitational potential) arising from macroscopic degrees of freedom. The first law of thermodynamics accounts for changes to the internal energy of a thermodynamic system.

Geophysical fluids have multiple matter components that are exchanged between fluid elements. There are also multiple phases of matter that must be considered for realistic studies (e.g., water vapor in clouds, sea ice, snow, etc.). However, in this section we first treat the case of a homogeneous and single-phase fluid with fluid elements that are materially closed yet thermally open. In Section 21.3 we consider the case of a materially open system, whereas the case of multiple phases is beyond the scope of this book.

21.2.1 Basic expression of the first law of thermodynamics

The first law of thermodynamics for a materially closed system establishes a relationship between infinitesimal changes of internal energy of a physical system, the work done to or by the system, and the thermal energy transferred between the system and its surrounding environment. The first law takes on the mathematical form

$$dJ^e = dW + dQ \quad \Leftarrow \text{materially closed.} \quad (21.2)$$

In this equation, dJ^e is the infinitesimal increment (an exact differential) of the internal energy; dW is the change in internal energy due to work applied to the system (*working*); and dQ is the internal energy change due to thermal energy transferred to the system (*heating*). We only have occasion to study mechanical work in this book, though note that there are other forms such as those arising from electromagnetic forces.

The first law of thermodynamics is a statement of energy conservation of a physical system, where energy changes arise from working and heating applied to or by the system. The relevant energy is just the internal energy since we here ignore mechanical energy associated with moving fluids. But clearly that limitation is not suitable for a moving fluid, where mechanical energy exchanges are of direct interest. We thus take up the topic of mechanical energy in Chapter 22, where we add mechanical energy to internal energy to form the total energy of a moving fluid system (Section 22.5).

21.2.2 The nature of working and heating

Working and heating are both path-dependent thermodynamic processes that transform a system from one thermodynamic state to another. That is, working and heating represent *path functions*. They are mathematically represented by inexact differentials as denoted by the d symbol. It is remarkable that the first law in equation (21.5) shows that the sum of two inexact differentials equals to an exact differential. The exact differential is that of internal energy, which is a *state function* that is a property of the thermodynamic state of a system. It follows that if the internal energy change occurs in the absence of heating then the working process must occur as a path independent process. The converse holds if internal energy changes without any working.

Working and heating denote actions applied to a system (verbs) rather than properties of a system (nouns). They are energy *in transition* that arise at the boundary of a thermodynamic system. We raise this somewhat pedantic yet subtle point since the terms “work” and “heat” are often used instead of “working” and “heating”. Indeed, we will often make use of that language. However, such usage should be used with care as it can spuriously lead one to seek information concerning the “work content” or “heat content” of a physical system; i.e., to consider work and heat as state properties (i.e., nouns). Rather, in thermodynamics we only consider the work imparted to change a system’s energy (working), or likewise the thermal energy used to change a system (heating).

These conceptual points are particularly relevant when asking questions about the heat transported by a fluid (with dimensions energy per time and SI units of Watt = Joule per second). In the analysis of a heat budget, it is tempting to define the “heat content” of a fluid element or fluid region according to its temperature, mass, and heat capacity. But the notion of heat content spuriously conflates a thermodynamic process whereby a system moves from one state to another (heating) with a thermodynamic state property (e.g., enthalpy, which is a property of the state; see Section 21.4.3). Furthermore, any definition of heat content is ambiguous due to the arbitrariness of the temperature scale; i.e., heat content based on the Celsius scale is distinct from that based

on the Kelvin scale. Therefore, when working with heat transport, care should be exercised if also including the notion of heat content. One way to detect an error is to ask whether a particular conclusion is modified by changing the temperature scale. If so, then one should revisit assumptions of the analysis since the results might be unphysical.

21.2.3 Mechanical work

As forces do work on a physical system they change its energy, and we are concerned with work done on geophysical fluids by mechanical forces. One way to perform mechanical work is via changes to the volume of a fluid element through the action of pressure (a force per area). When such changes occur quasi-statically, then we are led to consider *pressure work*, which takes the mathematical form

$$dW = -p dV, \quad (21.3)$$

where p is the pressure applied to the boundaries of the fluid element. The assumption that the mechanical process is quasi-static allows us to unambiguously define pressure acting on the system, and thus to write equation (21.3) for the work. The negative sign arises since the compression of a fluid element into a smaller volume, $dV < 0$, requires positive mechanical work be applied to the fluid, $dW > 0$. The mathematical form of pressure work derives from the general form of mechanical work given by

$$\text{work} = \text{force} \times \text{distance} = \text{force}/\text{area} \times \text{distance} \times \text{area} = \text{force}/\text{area} \times \text{volume}. \quad (21.4)$$

We offer the following points in regards to this, and other, forms of mechanical work applied to a fluid element.

- Pressure is an intensive property that measures the *intensity* of a force (per area) that is conjugate to the extensive property, V . Pressure is also the *integrating factor* connecting the inexact (path dependent) differential dW to the exact differential dV (see Section 2.8 for the mathematics of inexact differentials).
- For a quasi-static process, pressure changes the internal energy of a fluid through the pressure work (21.3). As seen in Section 22.3.1, pressure also changes the kinetic energy of a moving fluid by changing the fluid speed. When combining the internal energy and mechanical energy budgets in Section 22.5, we see how pressure affects the total energy of a fluid element.
- The gravitational force performs work on a fluid that moves through a gravitational field. Gravitational work affects changes to the fluid's kinetic energy whereas it does not directly affect internal energy in geophysical fluids. Consequently, gravitational work is fully accounted for in the mechanical energy budget, with details provided in Section 22.3.
- We generally ignore surface tension when studying geophysical fluid motions on scales larger than a few centimeters. See Section 19.10 for a discussion of surface tension.

21.2.4 Quasi-static transfer of thermal energy and entropy

The internal energy of a thermodynamic system can change when the molecular degrees of freedom are energized by modifying the thermal properties of the system. For a materially closed system whose thermal properties change in a quasi-static manner, the thermal energy transfer is related to entropy changes via

$$dQ = T dS^e \quad \text{materially closed system.} \quad (21.5)$$

This equation is a statement of the second law of thermodynamics for a quasi-static and materially closed system. T is the thermodynamic temperature (measured relative to absolute zero) and it is an intensive variable whereas S^e is the extensive form of entropy. Entropy is an extensive state function so that T provides the integrating factor connecting the inexact differential $d\Omega$ to the exact differential dS^e . A nonzero $d\Omega$ in a geophysical fluid can arise from radiative fluxes external to the fluid element; internal sources from friction; and the exchange of thermal energy through the mixing of fluid properties.

21.2.5 Fundamental thermodynamic relation

We summarize the discussion of this section by writing the first law for quasi-static materially closed processes

$$d\mathcal{I}^e = -p dV + T dS^e \iff \text{quasi-static materially closed processes.} \quad (21.6)$$

This equation is known as the *Gibbs relation* or more commonly the *fundamental thermodynamic relation* for quasi-static materially closed processes. This relation suggests that we interpret minus the pressure as the amount of internal energy required to add one unit of volume to the system while holding entropy fixed. Likewise, temperature is the internal energy required to add one unit of entropy to the system while holding volume fixed.

It is notable that all differentials within the fundamental thermodynamic relation (21.6) are exact differentials of state functions. This property is a result of assuming the thermodynamic processes are quasi-static, in which case we can determine the integrating factors pressure and temperature to thus replace inexact differentials with exact. Even with the quasi-static restriction, this equation offers great utility in the formalism of thermodynamics, with its extension to materially open systems given in Section 21.3. Indeed, since we are only concerned with quasi-static changes to fluid elements in this book, the fundamental thermodynamic relation (21.6) provides the central expression of the first law of thermodynamics.

It follows from the first law expression in equation (21.6) that the internal energy is a natural function of volume and entropy

$$d\mathcal{I}^e = -p dV + T dS^e \implies \mathcal{I}^e = \mathcal{I}^e(V, S^e). \quad (21.7)$$

Conversely, the entropy for a materially closed system is naturally a function of volume and internal energy

$$T dS^e = d\mathcal{I}^e + p dV \implies S^e = S^e(V, \mathcal{I}^e). \quad (21.8)$$

These relations have many implications arising from the mathematical properties of partial derivatives. We delay our development of these properties until first considering the case of materially open and non-homogeneous systems that allow for matter exchange (Section 21.3).

21.2.6 Entropy and reversible/irreversible processes

We close this section by further characterizing thermodynamic processes according to whether they occur reversibly or irreversibly. A *reversible process* can, at each stage, go either forward or backward in time. In the absence of non-conservative forces (e.g., dissipation such as friction), Newton's dynamical laws are reversible. For example, one observes nothing unphysical about the motion of an ideal pendulum with time moving backward rather than forward.

From a thermodynamic perspective, a reversible process does not alter the net entropy of a physical system plus its surrounding environment. Reversible thermodynamic processes are quasi-static and yet not all quasi-static processes are reversible. For example, a process that involves

friction can evolve quasi-statically and yet frictional processes, as with any dissipative process, increases net entropy. When the net entropy changes, we say the process occurs *irreversibly*, with the second law of thermodynamics stating that the net entropy change must be positive. Quasi-static is a property of how a system changes, whereas reversibility is a statement about how both the system and its surrounding environment change.

The entropy increment for a quasi-static process is given by $dS^e = dQ/T$, with $T > 0$ (recall T is the Kelvin thermodynamic temperature) so that the entropy increment has the same sign as the heating increment. We here see how the idealization of a heat bath allows us to perform reversible heating; i.e., heating without change in net entropy for a thermodynamic system plus the heat bath. Heat baths are held at a fixed temperature, which is the idealization of the case when the surrounding environment is arbitrarily larger than the thermodynamic system under consideration. Now imagine exchanging heat between a thermodynamic system and a series of heat baths to progressively alter the system's temperature by differential increments dT . In each exchange of heat, the entropy of the system plus heat bath is constructed to remain unchanged since we are exchanging and equal magnitude of entropy between them

$$dS_{\text{net}} = dS_{\text{system}} + dS_{\text{bath}} = dQ/T - dQ/T = 0. \quad (21.9)$$

To reverse the process, we merely reverse the heat exchanges between the thermodynamic system and the heat baths.

As noted above, when any thermodynamic process occurs irreversibly there is a net increase in entropy of the universe, which is a statement of the second law of thermodynamics. In statistical mechanics, entropy is computed by counting the number of microstates accessible to any given macrostate. Reversible processes do not modify the number of accessible microstates so there is zero change in the entropy. In contrast, as codified by the second law, irreversible processes increase entropy by increasing the number of accessible microstates.

21.3 First law for materially open systems

We now extend the discussion of Section 21.2 to materially open physical systems whose changes occur through quasi-static processes. These considerations have applications to fluid elements whose matter content is modified by mixing with the surrounding environment.

21.3.1 Constituent masses and matter concentrations

Write M_n for the mass of a material constituent $n = 1, N$, with the total mass of the physical system written

$$M = \sum_{n=1}^N M_n. \quad (21.10)$$

Since we are interested in fluid elements, we restrict attention to constant mass physical systems, even if the system exchanges matter content with its surrounding environment. The corresponding matter concentration is written

$$C_n = M_m/M \implies \sum_{n=1}^N C_n = 1, \quad (21.11)$$

with the constant mass constraint $\sum_{n=1}^N C_n = 1$ meaning that only $N - 1$ of the concentrations are linearly independent. We also introduced such *tracer concentrations* in Section 16.1 when discussing the tracer equation.

21.3.2 Scaling for homogeneous functions

Both the internal energy, \mathcal{I}^e , and entropy, \mathcal{S}^e , are extensive properties of a fluid system. Hence, if we add or remove a piece of matter to the system these thermodynamic properties are correspondingly modified. The internal energy and entropy thus have their natural functional dependencies (21.7) and (21.8) extended to include the matter content

$$\mathcal{I}^e = \mathcal{I}^e(V, \mathcal{S}^e, M_n) \quad \text{and} \quad \mathcal{S}^e = \mathcal{S}^e(V, \mathcal{I}^e, M_n), \quad (21.12)$$

where the M_n argument is shorthand for $M_1, M_2 \dots M_N$ for N matter constituents.

Now scale the system by an arbitrary parameter λ . Under this scale operation, the extensive variables, \mathcal{I}^e , \mathcal{S}^e , as well as the volume, V , and masses, M_n , scale by the same scale factor. The relation (21.12) thus leads to

$$\mathcal{I}^e(\lambda V, \lambda \mathcal{S}^e, \lambda M_n) = \lambda \mathcal{I}^e(V, \mathcal{S}^e, M_n) \quad \text{and} \quad \mathcal{S}^e(\lambda V, \lambda \mathcal{I}^e, \lambda M_n) = \lambda \mathcal{S}^e(V, \mathcal{I}^e, M_n) \quad (21.13)$$

A function that scales in this way is termed a *homogeneous function of degree one*. Homogeneous function of degree one possess the following remarkable property, readily derived by differentiating both sides of equation (21.13) with respect to λ and then setting $\lambda = 1$

$$\mathcal{I}^e(V, \mathcal{S}^e, M_n) = V \left[\frac{\partial \mathcal{I}^e}{\partial V} \right]_{\mathcal{S}^e, M_n} + \mathcal{S}^e \left[\frac{\partial \mathcal{I}^e}{\partial \mathcal{S}^e} \right]_{V, M_n} + \sum_{n=1}^N M_n \left[\frac{\partial \mathcal{I}^e}{\partial M_n} \right]_{V, \mathcal{S}^e, M_{m \neq n}} \quad (21.14)$$

$$\mathcal{S}^e(V, \mathcal{I}^e, M_n) = V \left[\frac{\partial \mathcal{S}^e}{\partial V} \right]_{\mathcal{I}^e, M_n} + \mathcal{I}^e \left[\frac{\partial \mathcal{S}^e}{\partial \mathcal{I}^e} \right]_{V, M_n} + \sum_{n=1}^N M_n \left[\frac{\partial \mathcal{S}^e}{\partial M_n} \right]_{V, \mathcal{I}^e, M_{m \neq n}}. \quad (21.15)$$

Making use of the fundamental thermodynamic relations (21.7) and (21.8) leads to

$$\mathcal{I}^e = -V p + \mathcal{S}^e T + \sum_{n=1}^N M_n \left[\frac{\partial \mathcal{I}^e}{\partial M_n} \right]_{V, \mathcal{S}^e, M_{m \neq n}} \quad (21.16)$$

$$T \mathcal{S}^e = V p + \mathcal{I}^e + \sum_{n=1}^N M_n \left[\frac{\partial \mathcal{S}^e}{\partial M_n} \right]_{V, \mathcal{I}^e, M_{m \neq n}}. \quad (21.17)$$

Self-consistency means that

$$\sum_{n=1}^N M_n \left[\frac{\partial \mathcal{I}^e}{\partial M_n} \right]_{V, \mathcal{S}^e} = - \sum_{n=1}^N M_n \left[\frac{\partial \mathcal{S}^e}{\partial M_n} \right]_{V, \mathcal{I}^e, M_{m \neq n}}, \quad (21.18)$$

which motivates defining the *chemical potential*

$$\mu_n \equiv \left[\frac{\partial \mathcal{I}^e}{\partial M_n} \right]_{V, \mathcal{S}^e, M_{m \neq n}} = - \left[\frac{\partial \mathcal{S}^e}{\partial M_n} \right]_{V, \mathcal{I}^e, M_{m \neq n}} \quad (21.19)$$

thus leading to

$$\mathcal{I}^e = T \mathcal{S}^e - p V + \sum_{n=1}^N \mu_n M_n \iff T \mathcal{S}^e = \mathcal{I}^e + p V - \sum_{n=1}^N \mu_n M_n. \quad (21.20)$$

By definition, the chemical potential, μ_n , measures the change in the internal energy, \mathcal{I}^e , when altering the mass, M_n , of the constituent n , while fixing the entropy, volume, and mass of the other components. Equivalently, it is minus the change in the entropy, \mathcal{S}^e , when altering the mass, M_n while fixing the volume, internal energy, and mass of the other components.

21.3.3 Chemical work and the Gibbs-Duhem relation

Changes to the matter composition of a system changes the internal energy through *chemical work*, written as $d\mathcal{C}$. If the changes to the matter composition occur quasi-statically then the chemical work is written

$$d\mathcal{C} = \sum_{n=1}^N \mu_n dM_n, \quad (21.21)$$

so that the chemical potential is the integrating factor connecting the inexact differential measuring the chemical work to the exact differential change in matter content. The chemical potential is the energy absorbed or released due to an infinitesimal change in the matter content. Inclusion of chemical work brings the first law to the form

$$d\mathcal{I}^e = dW + dQ + d\mathcal{C} \quad \text{materially open} \quad (21.22)$$

$$d\mathcal{I}^e = -p dV + T dS^e + \sum_{n=1}^N \mu_n dM_n \quad \text{quasi-static and materially open.} \quad (21.23)$$

Use of the quasi-static form of the first law (21.23) along with the differential of the result (21.20) leads to the *Gibbs-Duhem* relation

$$S^e dT - V dp + \sum_n M_n d\mu_n = 0. \quad (21.24)$$

In Exercise 21.1 we work through the derivation of Gibbs-Duhem in a bit more detail.

21.3.4 Partial derivatives

The fundamental thermodynamic relation (21.23) appears in terms of internal energy, with the corresponding partial derivative relations derived earlier

$$\left[\frac{\partial \mathcal{I}^e}{\partial S^e} \right]_{V, M_n} = T \quad (21.25)$$

$$\left[\frac{\partial \mathcal{I}^e}{\partial V} \right]_{S^e, M_n} = -p \quad (21.26)$$

$$\left[\frac{\partial \mathcal{I}^e}{\partial M_n} \right]_{S^e, V, M_{m \neq n}} = \mu_n. \quad (21.27)$$

Each equation relates an intensive property (right hand side) to the partial derivative of internal energy with respect to an extensive property. Furthermore, since each of the extensive properties is a homogeneous function of degree one, then the intensive properties are homogeneous functions of degree zero. That is, the intensive properties T , p , and μ_n do not scale with the size of the system. Rather, intensive properties are scale invariant.

Thermodynamics contains a plethora of partial derivatives, and we encounter many in later sections. Recall that partial derivatives are defined with the complement variables held fixed during the differentiation. Hence, so long as we are clear about functional dependence, extra subscripts such as those exposed in equations (21.25)-(21.27) are not needed for the partial derivatives. Nonetheless, traditional thermodynamic notation exposes all of the subscripts in order to remain explicit about the dependent and independent variables. Such notation, though clumsy, can be essential when in the midst of manipulations with thermodynamic potentials and their derivatives.

21.3.5 Equations of state

Equations (21.25)-(21.27) provide expressions for intensive properties, T , p , and μ_n , in terms of the partial derivatives of the internal energy in terms of extensive functions S^e , V , and M_m . We thus conclude that we can write T , p , and μ_n in the functional form

$$T = T(S^e, V, M_m) \quad \text{and} \quad p = p(S^e, V, M_m) \quad \text{and} \quad \mu_n = \mu_n(S^e, V, M_m). \quad (21.28)$$

These equations are known as *equations of state*. Knowledge of all the equations of state is equivalent to knowledge of the fundamental thermodynamic relation (21.23). We return to these equations in Section 21.4.1, where we encounter their form written in terms of the specific internal energy, specific volume, and matter concentration.

21.3.6 Fundamental thermodynamic relation per unit mass

For many purposes in fluid mechanics, it proves convenient to consider thermodynamic relations for a system of unit mass; i.e., “per unit mass”. Such relations are the natural starting point for moving from equilibrium thermodynamics to the quasi-equilibrium thermodynamics relevant for moving fluids (Chapter 22). For this purpose, we scale away the mass of the system by setting the scale factor $\lambda = M^{-1}$ in our discussion in Section 21.3.2 of how extensive properties scale. The result is the *specific* versions of the extensive properties

$$\mathcal{J}^e = M \mathcal{J} \quad (21.29a)$$

$$S^e = M S \quad (21.29b)$$

$$V = M \nu_s \quad (21.29c)$$

$$M_n = M C_n, \quad (21.29d)$$

where

$$\nu_s = 1/\rho \quad (21.30)$$

is the specific volume and the total mass, M , is held fixed. In the equality (21.29d), C_n is the mass fraction or concentration of species n in the fluid (Section 21.3.1). Substituting the specific quantities (21.29a)-(21.29d) into the fundamental thermodynamic relation (21.23) leads to the fundamental thermodynamic relation in terms of specific thermodynamic quantities

$$d\mathcal{J} = T dS - p d\nu_s + \sum_n \mu_n dC_n. \quad (21.31)$$

This is the form of the fundamental thermodynamic relation most commonly used in this book. Again, this relation holds for quasi-static processes that hold the total mass of the system fixed, thus making it relevant for our study of constant mass fluid elements in Chapter 22.

21.3.7 Seawater as a binary fluid

The atmosphere is a multi-component and multi-phase fluid that is well approximated as a mixture of water vapor and dry air. However, we do not consider moist atmospheric processes in this book nor do we consider phases changes. In contrast, there are many occasions in this book that require us to consider seawater as a binary fluid system of salt dissolved in fresh water so that their concentrations satisfy the constraint

$$C_{\text{salt}} + C_{\text{water}} = 1 \implies dC_{\text{water}} = -dC_{\text{salt}}. \quad (21.32)$$

We are thus able to write the Gibbs fundamental thermodynamic relation (21.31) in the form

$$d\mathcal{J} = T dS - p d\nu_s + \mu_{\text{water}} dC_{\text{water}} + \mu_{\text{salt}} dC_{\text{salt}} \quad (21.33a)$$

$$= T dS - p d\nu_s + \mu dC, \quad (21.33b)$$

where

$$C = C_{\text{salt}} \quad (21.34)$$

is the concentration of salt (mass of salt per mass of seawater), and

$$\mu = \mu_{\text{salt}} - \mu_{\text{water}} \quad (21.35)$$

is the *relative chemical potential*, often referred to as the *seawater chemical potential*. We return in Section 21.35 to discuss how the chemical potentials are computed according to partial derivatives of the Gibbs potential, which is the preferred method for the ocean and atmosphere.

The *absolute salinity* S , with units parts per thousand (gram per kilogram), is related to C_{salt} via

$$S = 1000 C_{\text{salt}}. \quad (21.36)$$

The range of absolute salinity in the ocean (roughly $0 \leq S \leq 40$) is more convenient than the range of C_{salt} , making salinity more commonly used in oceanography.

21.3.8 Further study

Appendix A of [Davis \(1994\)](#) and portions of Chapters 1 and 2 from [Olbers et al. \(2012\)](#) detail the full suite of thermodynamic relations for seawater. For our purposes it is sufficient to make use of the relation (21.33b) when considering a two-component fluid.

21.4 Thermodynamic potentials

Internal energy and entropy are referred to as state functions (functions only of the current state) as well as *thermodynamic potentials*, and they are related by equation (21.20), here written in its specific form

$$\mathcal{J} = T S - p \nu_s + \sum_{n=1}^N \mu_n C_n \iff T S = \mathcal{J} + p \nu_s - \sum_{n=1}^N \mu_n C_n. \quad (21.37)$$

Each thermodynamic potential is a natural function of certain other thermodynamic properties, as determined by the fundamental thermodynamic relation.

It is useful to have access to a suite of thermodynamic potentials (internal energy, entropy, enthalpy, Gibbs potential, Helmboltz free energy) that have different functional dependencies, which in turn yield distinct expressions for the fundamental equation of state. Thermodynamic potentials are related mathematically through a *Legendre transformation*. Motivation for their introduction comes from the distinct laboratory and environmental conditions whereby the controlling parameters may differ.

In this section we introduce the variety of thermodynamic potentials commonly used for fluid mechanics, and exhibit their fundamental functional dependencies. We emphasize the importance of committing to a single choice for the thermodynamic potential when manipulating thermodynamic equations. The reason to adhere to a single choice is that functional dependencies change when switching to a different thermodynamic potential, thus exposing oneself to confusion when swapping within a particular formulation.

21.4.1 Internal energy

Recall the fundamental thermodynamic relation (21.33b) written for a binary fluid

$$d\mathcal{J} = T dS - p d\nu_s + \mu dC. \quad (21.38)$$

Equation (21.38) identifies the specific internal energy, \mathcal{J} , as a natural function of specific entropy, S , specific volume, ν_s , and matter concentration, C

$$\mathcal{J} = \mathcal{J}(S, \nu_s, C). \quad (21.39)$$

Knowledge of the fundamental thermodynamic relation (21.38) allows us to derive a variety of thermodynamic relations via partial differentiation. For example, we can identify

$$\left[\frac{\partial \mathcal{J}}{\partial S} \right]_{\nu_s, C} = T \quad \text{and} \quad \left[\frac{\partial \mathcal{J}}{\partial \nu_s} \right]_{S, C} = -p \quad \text{and} \quad \left[\frac{\partial \mathcal{J}}{\partial C} \right]_{S, \nu_s} = \mu, \quad (21.40)$$

which are the specific forms of equations (21.25)-(21.27)

In Section 21.3.5 we introduced the notion of an equation of state, relating intensive properties to their corresponding extensive conjugates. Here we see that equations (21.40) provide a relation between T, p, μ as derivatives of a function, the internal energy, which is itself a function $\mathcal{J}(S, \nu_s, C)$. Hence, we may consider T, p, μ each as a function of the specific quantities (S, ν_s, C)

$$T = T(S, \nu_s, C) \quad \text{and} \quad p = p(S, \nu_s, C) \quad \text{and} \quad \mu = \mu(S, \nu_s, C), \quad (21.41)$$

which are equations of state in terms of specific quantities. In turn, the exact differentials of the intensive properties take on the form

$$dT = \left[\frac{\partial T}{\partial S} \right]_{\nu_s, C} dS + \left[\frac{\partial T}{\partial \nu_s} \right]_{C, S} d\nu_s + \left[\frac{\partial T}{\partial C} \right]_{S, \nu_s} dC \quad (21.42)$$

$$dp = \left[\frac{\partial p}{\partial S} \right]_{\nu_s, C} dS + \left[\frac{\partial p}{\partial \nu_s} \right]_{C, S} d\nu_s + \left[\frac{\partial p}{\partial C} \right]_{S, \nu_s} dC \quad (21.43)$$

$$d\mu = \left[\frac{\partial \mu}{\partial S} \right]_{\nu_s, C} dS + \left[\frac{\partial \mu}{\partial \nu_s} \right]_{C, S} d\nu_s + \left[\frac{\partial \mu}{\partial C} \right]_{S, \nu_s} dC. \quad (21.44)$$

21.4.2 Entropy

Rearrangement of the fundamental thermodynamic relation (21.38) leads to the exact differential for specific entropy

$$dS = \frac{1}{T} d\mathcal{J} + \frac{p}{T} d\nu_s - \frac{\mu}{T} dC. \quad (21.45)$$

In this form, specific entropy has the functional dependence

$$S = S(\mathcal{J}, \nu_s, C), \quad (21.46)$$

whose knowledge provides yet another form of the fundamental equation of state. This functional dependence, along with equation (21.45), lead to the following identities

$$\left[\frac{\partial S}{\partial \mathcal{J}} \right]_{\nu_s, C} = \frac{1}{T} \quad \text{and} \quad \left[\frac{\partial S}{\partial \nu_s} \right]_{\mathcal{J}, C} = \frac{p}{T} \quad \text{and} \quad \left[\frac{\partial S}{\partial C} \right]_{\mathcal{J}, \nu_s} = -\frac{\mu}{T}. \quad (21.47)$$

As for internal energy in Section 21.4.1, equation (21.47) provides a relation between T, p, μ as derivatives of a function, the entropy, which is itself a function $S(\mathcal{J}, \nu_s, C)$. Hence, we may consider T, p, μ as each a function of (\mathcal{J}, ν_s, C) to thus write the equations of state

$$T = T(\mathcal{J}, \nu_s, C) \quad \text{and} \quad p = p(\mathcal{J}, \nu_s, C) \quad \text{and} \quad \mu = \mu(\mathcal{J}, \nu_s, C). \quad (21.48)$$

21.4.3 Enthalpy

Thus far we have worked only with the fundamental thermodynamic relation (21.38). We now introduce the specific enthalpy

$$\mathcal{H} = \mathcal{I} + p \nu_s = T \mathcal{S} + \sum_{n=1}^N \mu_n C_n, \quad (21.49)$$

where the second equality made use of equation (21.37). Use of the fundamental thermodynamic relation (21.38) leads to the exact differential for enthalpy

$$d\mathcal{H} = d\mathcal{I} + d(p \nu_s) \quad (21.50a)$$

$$= T d\mathcal{S} - p d\nu_s + \mu dC + p d\nu_s + \nu_s dp \quad (21.50b)$$

$$= T d\mathcal{S} + \nu_s dp + \mu dC. \quad (21.50c)$$

Recalling that for quasi-static processes, $T d\mathcal{S}$ equals to the thermal energy added to a fluid element. Hence, for processes occurring at constant pressure and concentration, changes in enthalpy are determined by the thermal energy added to the system.

Equation (21.50c) provides the fundamental thermodynamic relation with enthalpy rather than internal energy. Consequently, the *Legendre transformation* (21.49) renders a functional dependence for enthalpy

$$\mathcal{H} = \mathcal{H}(\mathcal{S}, p, C), \quad (21.51)$$

which in turn leads to the following partial derivative identities

$$\left[\frac{\partial \mathcal{H}}{\partial \mathcal{S}} \right]_{p,C} = T \quad \text{and} \quad \left[\frac{\partial \mathcal{H}}{\partial p} \right]_{\mathcal{S},C} = \nu_s \quad \text{and} \quad \left[\frac{\partial \mathcal{H}}{\partial C} \right]_{\mathcal{S},p} = \mu. \quad (21.52)$$

As for internal energy in Section 21.4.1, equations (21.52) provide a relation between T, ν_s, μ as derivatives of a function, the enthalpy, which is itself a function $\mathcal{H}(\mathcal{S}, p, C)$. Hence, we may consider T, ν_s, μ as each a function of (\mathcal{S}, p, C) to thus render the following equations of state

$$T = T(\mathcal{S}, p, C) \quad \text{and} \quad \nu_s = \nu_s(\mathcal{S}, p, C) \quad \text{and} \quad \mu = \mu(\mathcal{S}, p, C). \quad (21.53)$$

Enthalpy's functional dependence (21.51) is more convenient for studies of geophysical fluids than that for internal energy, $\mathcal{I}(\mathcal{S}, \nu_s, C)$, or for entropy $\mathcal{S}(\mathcal{I}, \nu_s, C)$.

- In the laboratory or field, we generally have direct mechanical means for measuring pressure in a fluid, whereas specific volume requires indirect methods involving the equation of state for density discussed in Section 25.3.
- Correspondingly, the interaction between fluid elements typically occurs at near constant pressure. Hence, fluid elements exchange both their entropy and enthalpy when the exchange occurs as constant pressure.
- Specific entropy remains constant on a fluid element in the absence of mixing or other irreversible effects. Correspondingly, enthalpy remains constant for constant pressure motion without mixing. Conversely, in the presence of mixing at constant pressure, fluid elements mix their specific enthalpy, specific entropy, and tracer concentration.

21.4.4 Helmholtz free energy

The Helmholtz free energy is defined by the Legendre transformation

$$\mathcal{F} = \mathcal{I} - T\mathcal{S} = -p\nu_s + \sum_{n=1}^N \mu_n C_n, \quad (21.54)$$

where the second equality made use of equation (21.37). The exact differential of the Helmholtz free energy is given by

$$d\mathcal{F} = d\mathcal{I} - d(T\mathcal{S}) \quad (21.55a)$$

$$= d\mathcal{I} - T d\mathcal{S} - \mathcal{S} dT \quad (21.55b)$$

$$= -\mathcal{S} dT - p d\nu_s + \mu dC, \quad (21.55c)$$

where we used the fundamental thermodynamic relation (21.33b) for the final equality. Isothermal and constant concentration processes render the changes to the free energy equal to the pressure work applied to the system.

The Helmholtz free energy has the functional dependence

$$\mathcal{F} = \mathcal{F}(T, \nu_s, C), \quad (21.56)$$

which then leads to the partial derivatives identities

$$\left[\frac{\partial \mathcal{F}}{\partial T} \right]_{\nu_s, C} = -\mathcal{S} \quad \text{and} \quad \left[\frac{\partial \mathcal{F}}{\partial \nu_s} \right]_{T, C} = -p \quad \text{and} \quad \left[\frac{\partial \mathcal{F}}{\partial C} \right]_{T, \nu_s} = \mu. \quad (21.57)$$

As for internal energy in Section 21.4.1, equations (21.57) provide a relation between \mathcal{S}, p, μ as derivatives of a function, the Helmholtz free energy, which is itself a function $\mathcal{F}(T, \nu_s, C)$. Hence, we may consider \mathcal{S}, p, μ each as a function of (T, ν_s, C) to render the equations of state

$$\mathcal{S} = \mathcal{S}(T, \nu_s, C) \quad \text{and} \quad p = p(T, \nu_s, C) \quad \text{and} \quad \mu = \mu(T, \nu_s, C). \quad (21.58)$$

21.4.5 Gibbs potential

The Gibbs potential is defined by the Legendre transformation

$$\mathcal{G} = \mathcal{I} + p\nu_s - T\mathcal{S} = \mathcal{H} - T\mathcal{S} = \sum_{n=1}^N \mu_n C_n, \quad (21.59)$$

where the final equality made use of equation (21.37). The exact differential of the Gibbs potential is given by

$$d\mathcal{G} = d\mathcal{H} - d(T\mathcal{S}) \quad (21.60a)$$

$$= T d\mathcal{S} + \nu_s dp + \mu dC - T d\mathcal{S} - \mathcal{S} dT \quad (21.60b)$$

$$= -\mathcal{S} dT + \nu_s dp + \mu dC, \quad (21.60c)$$

where we made use of the fundamental thermodynamic relation (21.50c) written in terms of enthalpy. The Gibbs potential has the functional dependence

$$\mathcal{G} = \mathcal{G}(T, p, C), \quad (21.61)$$

which leads to the partial derivatives identities

$$\left[\frac{\partial \mathcal{G}}{\partial T} \right]_{p,C} = -\mathcal{S} \quad \text{and} \quad \left[\frac{\partial \mathcal{G}}{\partial p} \right]_{T,C} = \nu_s \quad \text{and} \quad \left[\frac{\partial \mathcal{G}}{\partial C} \right]_{T,p} = \mu. \quad (21.62)$$

As for internal energy in Section 21.4.1, equations (21.62) provide a relation between \mathcal{S} , ν_s , μ as derivatives of a function, the Gibbs potential, which is itself a function $\mathcal{G}(T, p, C)$. Hence, we may consider \mathcal{S} , ν_s , μ each as a function of (T, p, C) to render the equations of state

$$\mathcal{S} = \mathcal{S}(T, p, C) \quad \text{and} \quad \nu_s = \nu_s(T, p, C) \quad \text{and} \quad \mu = \mu(T, p, C). \quad (21.63)$$

The form of the fundamental dependies (21.61), and the associated equations of state (21.63), are often used in fluid mechanics and physical chemistry. The reason is that temperature, pressure, and concentration are readily measured in the laboratory and the environment. We can thus readily measure the partial derivatives of \mathcal{G} , and the functional dependence (21.63) provides a convenient means to express \mathcal{S} , ν_s , and μ (e.g., see the adiabatic lapse rate discussion in Section 21.7).

Given its convenient functional dependence, the Gibbs potential plays a central role in developing the thermodynamics of seawater as formulated by [Feistel \(1993\)](#) and codified by [IOC et al. \(2010\)](#). We thus endeavor to exhibit how quantities (e.g., response functions as in Section 21.5) can be computed based on knowledge of the Gibbs potential and its partial derivatives. For example, use of equation (21.62) renders the expression for the enthalpy

$$\mathcal{H} = \mathcal{G} + T \mathcal{S} = \mathcal{G} - T \left[\frac{\partial \mathcal{G}}{\partial T} \right]_{p,C}. \quad (21.64)$$

21.4.6 Seawater chemical potential in terms of the Gibbs potential

Throughout this section we have displayed equations for the chemical potential of a binary fluid in terms of the partial derivatives of the thermodynamic potentials, such as equation (21.62) using the Gibbs potential. Here we consider some details that lead to further understanding of the partial derivatives. We start by writing the chemical potential of fresh water and salt as contained within seawater in terms of the partial derivatives of the extensive Gibbs potential

$$\mu_{\text{water}} = \left[\frac{\partial \mathcal{G}^e}{\partial M_{\text{water}}} \right]_{T,p,M_{\text{salt}}} \quad \text{and} \quad \mu_{\text{salt}} = \left[\frac{\partial \mathcal{G}^e}{\partial M_{\text{salt}}} \right]_{T,p,M_{\text{water}}}. \quad (21.65)$$

The total mass of a sample of seawater is given by $M = M_{\text{water}} + M_{\text{salt}}$. Consequently, to compute these partial derivatives requires us to alter the mass of the sample as we hold the mass of one component fixed while varying the mass of the other component. This sort of partial derivative is less convenient for our purposes since we prefer to work with constant mass samples, such as we encounter with constant mass fluid elements. For that purpose we introduce the specific Gibbs potential, in which case the chemical potential of fresh water is

$$\mu_{\text{water}} = \left[\frac{\partial \mathcal{G}^e}{\partial M_{\text{water}}} \right]_{T,p,M_{\text{salt}}} = \left[\frac{\partial (M \mathcal{G})}{\partial M_{\text{water}}} \right]_{T,p,M_{\text{salt}}} = \mathcal{G} + M \left[\frac{\partial \mathcal{G}}{\partial M_{\text{water}}} \right]_{T,p,M_{\text{salt}}}. \quad (21.66)$$

The specific Gibbs potential is a natural function of T, p, C_n , and since $C_{\text{water}} + C_{\text{salt}} = 1$ we can write the Gibbs potential in terms of just one of the concentrations, typically chosen as C_{salt} . We are thus led to

$$\mu_{\text{water}} = \mathcal{G} + M \left[\frac{\partial \mathcal{G}}{\partial M_{\text{water}}} \right]_{T,p,M_{\text{salt}}} \quad (21.67a)$$

$$= \mathcal{G} + M \left[\frac{\partial \mathcal{G}}{\partial C_{\text{water}}} \right]_{T,p} \left[\frac{\partial C_{\text{water}}}{\partial M_{\text{water}}} \right]_{M_{\text{salt}}} \quad (21.67b)$$

where the concentration partial derivative is given by

$$\left[\frac{\partial C_{\text{water}}}{\partial M_{\text{water}}} \right]_{M_{\text{salt}}} = \left[\frac{\partial}{\partial M_{\text{water}}} \right]_{M_{\text{salt}}} \left[\frac{M_{\text{water}}}{M_{\text{water}} + M_{\text{salt}}} \right] = \frac{C_{\text{salt}}}{M}, \quad (21.68)$$

thus leading to the chemical potential of fresh water within seawater

$$\mu_{\text{water}} = \mathcal{G} + C_{\text{salt}} \left[\frac{\partial \mathcal{G}}{\partial C_{\text{water}}} \right]_{T,p} = \mathcal{G} - C_{\text{salt}} \left[\frac{\partial \mathcal{G}}{\partial C_{\text{salt}}} \right]_{T,p}. \quad (21.69)$$

We are thus able to work with the specific Gibbs function for a constant mass fluid element and compute its concentration partial derivative. Similar manipulations lead to the chemical potential for salt within seawater

$$\mu_{\text{salt}} = \mathcal{G} + C_{\text{water}} \left[\frac{\partial \mathcal{G}}{\partial C_{\text{salt}}} \right]_{T,p} = \mathcal{G} + (1 - C_{\text{salt}}) \left[\frac{\partial \mathcal{G}}{\partial C_{\text{salt}}} \right]_{T,p}. \quad (21.70)$$

We are thus led to the seawater chemical potential

$$\mu = \mu_{\text{salt}} - \mu_{\text{water}} = \left[\frac{\partial \mathcal{G}}{\partial C_{\text{salt}}} \right]_{T,p}, \quad (21.71)$$

which agrees with equation (21.62).

21.5 Response functions

Response functions measure the change in a thermodynamic property as the system is forced in some manner. We here introduce the heat capacity, thermal expansion coefficient, and haline contraction coefficient, which are three response functions commonly encountered in ocean and atmospheric fluid mechanics.

21.5.1 Heat capacities

The heat capacity measures the change in heat associated with a change in temperature at constant matter composition. There are two distinct heat capacities generally considered in fluid mechanics: one with specific volume held fixed and the other with pressure held fixed

$$c_v \equiv \frac{1}{M} \left[\frac{dQ}{dT} \right]_{\nu_s, C} \quad \text{SI units m}^2 \text{ s}^{-2} \text{ K}^{-1}. \quad (21.72)$$

$$c_p \equiv \frac{1}{M} \left[\frac{dQ}{dT} \right]_{p, C} \quad \text{SI units m}^2 \text{ s}^{-2} \text{ K}^{-1}. \quad (21.73)$$

If heating occurs quasi-statically, we can make use of the equation (21.5) relating heat and entropy, applied here in its specific form $M^{-1} dQ = T dS$. The result is a state function form of the specific heat capacities

$$c_v = T \left[\frac{\partial S}{\partial T} \right]_{\nu_s, C} = -T \left[\frac{\partial}{\partial T} \right]_{\nu_s, C} \left[\frac{\partial \mathcal{G}}{\partial T} \right]_{p, C} \quad (21.74)$$

$$c_p = T \left[\frac{\partial S}{\partial T} \right]_{p, C} = -T \left[\frac{\partial}{\partial T} \right]_{p, C} \left[\frac{\partial \mathcal{G}}{\partial T} \right]_{p, C} \quad (21.75)$$

where the second equalities in both of the above equations introduced the Gibbs potential according to equation (21.62). Furthermore, we can make use of the fundamental thermodynamic relation (21.31) with specific volume and matter concentration held fixed to yield

$$c_v = T \left[\frac{\partial S}{\partial T} \right]_{\nu_s, C} = \left[\frac{\partial J}{\partial T} \right]_{\nu_s, C}. \quad (21.76)$$

Likewise, making use of the fundamental thermodynamic relation (21.50c) written in terms of enthalpy leads to the constant pressure heat capacity

$$c_p = T \left[\frac{\partial S}{\partial T} \right]_{p, C} = \left[\frac{\partial J}{\partial T} \right]_{p, C} + p \left[\frac{\partial \nu_s}{\partial T} \right]_{p, C} = \left[\frac{\partial H}{\partial T} \right]_{p, C}. \quad (21.77)$$

21.5.2 Thermal expansion, haline contraction, and sound speed

Thermal expansion coefficient

The thermal expansion coefficient measures relative changes in density as temperature changes at constant pressure and concentration

$$\alpha_T = -\frac{1}{\rho} \left[\frac{\partial \rho}{\partial T} \right]_{p, C} = \frac{1}{\nu_s} \left[\frac{\partial \nu_s}{\partial T} \right]_{p, C} = \frac{1}{(\partial G/\partial p)_{T, C}} \frac{\partial}{\partial T} \left[\frac{\partial G}{\partial p} \right]_{T, C} \quad (21.78)$$

where the final equality introduced the Gibbs function according to equation (21.62). The minus sign in the definition is introduced since density typically reduces when temperature increases, so that for most substances $\alpha_T > 0$. Freshwater near its freezing point is an important counter-example, with $\alpha_T < 0$ allowing for solid ice to float on liquid water.

Haline contraction coefficient

A similar response function measures changes to density arising from changes in the salt concentration (salinity) in seawater

$$\beta_S = \frac{1}{\rho} \left[\frac{\partial \rho}{\partial S} \right]_{p, T} = -\frac{1}{\nu_s} \left[\frac{\partial \nu_s}{\partial S} \right]_{p, T} = -\frac{1}{(\partial G/\partial p)_{T, S}} \frac{\partial}{\partial S} \left[\frac{\partial G}{\partial p} \right]_{T, S} \quad (21.79)$$

where $S = 1000 C$ is the salinity (21.36). Seawater density typically increases (fluid element volume contracts) when salinity increases, so that $\beta_S > 0$.

Sound speed

Changes in density with respect to pressure at a fixed entropy define the inverse squared sound speed²

$$\frac{1}{c_s^2} = \left[\frac{\partial \rho}{\partial p} \right]_S = -\frac{1}{(\nu_s)^2} \left[\frac{\partial \nu_s}{\partial p} \right]_S = -\frac{1}{[(\partial G/\partial p)_{T, S}]^2} \frac{\partial}{\partial p} \left[\frac{\partial G}{\partial p} \right]_{T, S}. \quad (21.80)$$

The sound speed is a strong function of pressure, generally increasing with higher pressure, as well as temperature, generally decreasing with lower temperature. For the ocean, these two effects compete when moving into the ocean interior. In the upper 500 m to 1000 m, decreasing temperatures cause the sound speed to reduce whereas at deeper regions the higher pressures overcome the temperature

²We discuss sound waves in Section 18.5.

effect thus increasing the sound speed. The result is a sound speed minimum between 500 m and 1000 m. The sound speed minimum and the associated acoustic waveguide play an important role in ocean acoustics, in particular for how certain whales are able to communicate across ocean basins. We consider the sound speed for an ideal gas in Section 21.9.8.

21.6 Maxwell relations for homogeneous fluids

Thermodynamics makes use of basic properties of exact differentials for the purpose of developing identities between partial derivatives. Maxwell relations refer to a suite of partial derivative identities that follow from the equality of mixed second partial derivatives of thermodynamic potentials. In this section we develop the Maxwell relations encountered with homogeneous fluids, with similar relations holding for two-component fluids. We make use of one Maxwell relation in Section 21.7 for expressing the adiabatic lapse rate in terms of readily measurable thermo-mechanical properties.

21.6.1 Maxwell relation from internal energy

As seen from Section 21.4.1, the natural functional dependence for internal energy in a homogeneous fluid is given by its fundamental thermodynamic relation (21.38)

$$d\mathcal{I} = \left[\frac{\partial \mathcal{I}}{\partial S} \right]_{\nu_s} dS + \left[\frac{\partial \mathcal{I}}{\partial \nu_s} \right]_S d\nu_s = T dS - p d\nu_s \implies \mathcal{I} = \mathcal{I}(S, \nu_s). \quad (21.81)$$

The mixed second partial derivatives are equal

$$\left[\frac{\partial}{\partial \nu_s} \right]_{S,C} \left[\frac{\partial}{\partial S} \right]_{\nu_s,C} \mathcal{I} = \left[\frac{\partial}{\partial S} \right]_{\nu_s,C} \left[\frac{\partial}{\partial \nu_s} \right]_{S,C} \mathcal{I}, \quad (21.82)$$

so that, via the fundamental thermodynamic relation (21.81), we have the Maxwell relation

$$\left[\frac{\partial T}{\partial \nu_s} \right]_S = - \left[\frac{\partial p}{\partial S} \right]_{\nu_s}. \quad (21.83)$$

21.6.2 Summary of the Maxwell relations

The other thermodynamic potentials, and their associated fundamental thermodynamical relations, lead to further Maxwell relations as summarized here

$$d\mathcal{J} = T dS - p d\nu_s \implies \left[\frac{\partial T}{\partial \nu_s} \right]_S = - \left[\frac{\partial p}{\partial S} \right]_{\nu_s} \quad (21.84)$$

$$d\mathcal{H} = T dS + \nu_s dp \implies \left[\frac{\partial T}{\partial p} \right]_S = \left[\frac{\partial \nu_s}{\partial S} \right]_p \quad (21.85)$$

$$d\mathcal{G} = -S dT + \nu_s dp \implies \left[\frac{\partial S}{\partial p} \right]_T = - \left[\frac{\partial \nu_s}{\partial T} \right]_p. \quad (21.86)$$

$$d\mathcal{F} = -S dT - p d\nu_s \implies \left[\frac{\partial S}{\partial \nu_s} \right]_T = \left[\frac{\partial p}{\partial T} \right]_{\nu_s}. \quad (21.87)$$

These four Maxwell relations for homogeneous fluids involve permutations on cross derivatives of (T, S) and (p, ν_s) . In statistical mechanics, (T, S) determine the density of accessible microscopic states forming the thermodynamic system, whereas (p, ν_s) involves an external control parameter and its corresponding generalized force.

21.7 Adiabatic lapse rate

The temperature of a fluid can change without the transfer of heat. This *adiabatic* temperature change arises when the fluid pressure changes. We here introduce the *adiabatic lapse rate*, which measures the changes in temperature for a static fluid placed in a gravity field. There are two lapse rates commonly considered: one related to height and one related to pressure. We introduce some manipulations commonly performed with thermodynamic state functions and their partial derivatives, with the goal of expressing the lapse rate in terms of commonly measured response functions.

21.7.1 Isentropic rearrangement

Consider a finite region of a static fluid in a gravitational field. Assume the fluid is initially in a thin horizontal layer with a uniform temperature. Now rearrange the fluid into a vertical column, and do so without changing the entropy; i.e., without the transfer of heat across the fluid boundary (adiabatically) and without mixing. Performing this rearrangement raises the center of mass of the fluid system and thus increases the gravitational potential energy. This process thus requires mechanical work against the gravitational field.

Gravity makes pressure at the bottom of the vertical fluid column greater than at the top. This pressure difference modifies the temperature in the column. We seek a thermodynamic expression for how changes in pressure affects changes in temperature for a static fluid, with the pressure changes imparted reversibly and adiabatically so that entropy does not change. Mathematically, we seek an expression for the partial derivative

$$\hat{\Gamma} \equiv \left[\frac{\partial T}{\partial p} \right]_{C,S}, \quad (21.88)$$

which is known as the *adiabatic lapse rate*. The adiabatic lapse rate can be measured directly, with empirical expressions fit to laboratory measurements. Additionally, it is convenient to express it in terms of other thermodynamic response functions in order to garner further physical insight. The necessary manipulations form the bulk of this section.

21.7.2 Thermodynamic formulation

When the matter concentration is held fixed, the equation of state (21.63) allows us to consider entropy as a function of temperature and pressure so that

$$dS = \left[\frac{\partial S}{\partial T} \right]_p dT + \left[\frac{\partial S}{\partial p} \right]_T dp. \quad (21.89)$$

Substituting the definition of heat capacity from equation (21.75) leads to

$$T dS = c_p dT + T \left[\frac{\partial S}{\partial p} \right]_T dp. \quad (21.90)$$

It is useful to eliminate $(\partial S / \partial p)_T$ in favor of a more easily measurable quantity. For that purpose we make use of the Maxwell relation (21.86) to write

$$\left[\frac{\partial S}{\partial p} \right]_T = - \left[\frac{\partial \nu_s}{\partial T} \right]_p. \quad (21.91)$$

Introducing the thermal expansion coefficient (21.78) yields an expression for changes in entropy in terms of changes in temperature and pressure

$$T dS = c_p dT - T \left[\frac{\partial \nu_s}{\partial T} \right]_p dp = c_p dT - \left[\frac{T \alpha_T}{\rho} \right] dp. \quad (21.92)$$

Since c_p and α_T are readily measurable response functions, the expression (21.92) is a useful means to compute infinitesimal entropy changes when matter concentration is held constant.

21.7.3 Adiabatic lapse rate for pressure changes

Equation (21.92) means that the change in temperature associated with motion through a pressure field, with $dS = 0$ and $dC = 0$, can be written

$$\hat{\Gamma} = \left[\frac{\partial T}{\partial p} \right]_{C,S} = \frac{T \alpha_T}{\rho c_p}. \quad (21.93)$$

Temperature indeed changes when pressure changes, even though there has been no heat exchanged with the environment. With $\hat{\Gamma}$ so defined, we can write the entropy change in equation (21.92) as

$$T dS = c_p (dT - \hat{\Gamma} dp). \quad (21.94)$$

The term $dT - \hat{\Gamma} dp$ subtracts from the *in situ* temperature increment the pressure induced changes in temperature. In Section 21.8 we introduce the *potential temperature*, which is defined just for the purpose of removing changes due to pressure.

21.7.4 Adiabatic lapse rate for height changes

A static fluid in a gravity field is in exact hydrostatic balance, whereby the pressure at a point equals to the weight per area above that point (Section 18.3). Hydrostatic balance in a constant gravity field maintains the following relation between the pressure differential increment and the vertical differential increment

$$dp = -g \rho dz. \quad (21.95)$$

Use of the chain rule within the lapse rate expression (21.93) leads to

$$\Gamma = \left[\frac{\partial T}{\partial z} \right]_{C,S} = \left[\frac{\partial T}{\partial p} \right]_{C,S} \left[\frac{\partial p}{\partial z} \right] = -\rho g \left[\frac{T \alpha_T}{\rho c_p} \right] = -\frac{g T \alpha_T}{c_p}. \quad (21.96)$$

This form for the lapse rates measures the change in temperature (the *lapse*) within a constant composition fluid element as it is isentropically moved vertically through a hydrostatic pressure field.

21.7.5 Further study

[McDougall and Feistel \(2003\)](#) provide a discussion of the lapse rate in terms of molecular dynamics. In particular, they note that the lapse rate, being proportional to the thermal expansion coefficient, can be negative when the thermal expansion is negative. A negative thermal expansion coefficient occurs in cool fresh water, such as the Baltic Sea. Hence, although work is done on the fluid element under increasing pressure, its temperature decreases in these cases.

The addition of water to the atmosphere modifies the lapse rate, as the air is then no longer well approximated by an ideal gas. Chapter 18 of [Vallis \(2017\)](#) offers a pedagogical discussion of the thermodynamics of a moist tropical atmosphere.

21.8 Potential temperature

The *in situ* temperature, T , does not make for a convenient prognostic variable for many theoretical or numerical modeling purposes in either the ocean or atmosphere. Its awkward nature relates to the adiabatic changes in T experienced when pressure changes. Namely, since *in situ* temperature changes when pressure changes, even when there is no heat transport (see Section 21.7), the *in situ* temperature T is modified even when the flow is adiabatic. In contrast, it is generally more convenient to work with properties, like salinity and humidity, that remain materially constant in the absence of mixing.

In a related context, we consider the process of heating and cooling of the ocean, which occurs predominantly near the ocean surface. In contrast, transport in the ocean interior is nearly adiabatic and isohaline (i.e., nearly isentropic). The physical picture is suggested whereby the surface ocean boundary layer experiences irreversible processes that set characteristics of water masses that move quasi-reversibly within the ocean interior. As a means to characterize and thus to label these *water masses*, oceanographers prefer to use properties that maintain constant values when moving within the quasi-isentropic ocean interior. Salinity is a good label for this purpose since it is only altered by mixing between waters of varying concentrations, and in turn it is materially constant in the absence of mixing. This behavior constitutes a basic property of material tracers (tracers that measure the mass per mass of a constituent as discussed in Section 16.1). However, it is *not* a property of the *in situ* temperature, T , which changes even in the absence of mixing due to pressure effects.

We are thus motivated to seek a thermodynamic tracer that evolves analogously to material tracers, so that it can be used as a second material label for fluid elements. A similar motivation stems from the analysis of atmospheric motions. These considerations lead to the introduction of *potential temperature*. Notably, as discussed in Section 22.8, we find that an even better measure of heat transport is provided by Conservative Temperature, Θ , which is the potential enthalpy divided by a constant heat capacity. Nonetheless, potential temperature remains a common field whose use remains ubiquitous in atmospheric and oceanic studies. Furthermore, in some special cases it is directly proportional to the specific entropy. We are thus motivated to provide a discussion of how it is computed.

21.8.1 Adiabatic temperature changes

Vertical motion of a fluid element, without exchange of heat (adiabatic) or matter (constant concentration) changes the pressure of the fluid element. In turn, this motion causes the *in situ* temperature to change in proportion to the adiabatic lapse rate given by (Section 21.7)

$$dT = \hat{\Gamma} dp. \quad (21.97)$$

Consequently, and as already noted, the *in situ* temperature is not a useful thermodynamic variable to label fluid elements since it changes even in the absence of irreversible mixing processes. Instead, it is more useful to remove the adiabatic pressure effects. This is the key reason for introducing potential temperature.

21.8.2 Defining the potential temperature

Removing adiabatic pressure effects from *in situ* temperature leads to the concept of *potential temperature*. Potential temperature is defined as the *in situ* temperature that a fluid element of fixed material composition would have if it were isentropically transported from its *in situ* pressure

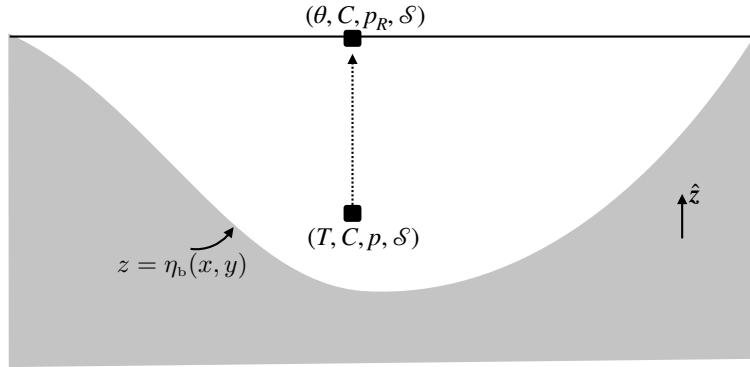


FIGURE 21.1: Potential temperature is the *in situ* temperature that a fluid element of fixed material composition would have if isentropically displaced from its *in situ* pressure to a reference pressure, p_R . The schematic here depicts that displacement for a seawater fluid element with *in situ* temperature, T , salinity, $S = 1000 C$, pressure, p , and specific entropy, \mathcal{S} . The element is moved to the ocean surface with the standard sea level atmospheric pressure providing the reference pressure.

to a reference pressure p_R , with the reference pressure typically taken at the ocean/land surface (see Figure 21.1).

Mathematically, we define the potential temperature, θ , as the reference temperature obtained via integration of $dT = \hat{\Gamma} dp$ for an isentropic *in situ* temperature change with respect to pressure

$$\int_{\theta}^T dT' = \int_{p_R}^p \hat{\Gamma}(T, p', C) dp' \implies T = \theta(T, p_R, C) + \int_{p_R}^p \hat{\Gamma}(T, p', C) dp', \quad (21.98)$$

with $\hat{\Gamma}$ the lapse rate defined in terms of pressure changes (equation (21.93)). By definition, the *in situ* temperature, T , equals the potential temperature, θ , at the reference pressure, $p = p_R$. Elsewhere, they differ by an amount determined by the adiabatic lapse rate. Furthermore, we see that

$$\left[\frac{\partial T}{\partial p} \right]_{C,S} = \left[\frac{\partial \theta}{\partial p} \right]_{C,S} + \hat{\Gamma}. \quad (21.99)$$

However, by definition

$$\left[\frac{\partial T}{\partial p} \right]_{C,S} = \hat{\Gamma} \quad (21.100)$$

so that

$$\left[\frac{\partial \theta}{\partial p} \right]_{C,S} = 0. \quad (21.101)$$

That is, by construction, the potential temperature depends explicitly on the concentration, C , and *in situ* temperature, T , and has a parametric dependence on the reference pressure. It has no explicit dependence on the *in situ* pressure when holding tracer concentration and entropy fixed. Finally, we emphasize that the potential temperature is a function of tracer concentration, C . Hence, the potential temperature generally changes if the tracer concentration changes. For example, potential temperature in the ocean changes if the salinity changes.

21.8.3 Potential temperature and specific entropy

An alternative definition of the potential temperature follows by noting that the entropy of a fluid element remains unchanged as it is reversibly moved to the reference pressure. Consequently,

writing entropy as a function of the *in situ* temperature, pressure, and matter concentration as in equation (21.63)

$$\mathcal{S} = \mathcal{S}(T, p, C) \quad (21.102)$$

leads to the defining identity for potential temperature

$$\mathcal{S} = \mathcal{S}(T, p, C) = \mathcal{S}(\theta, p_R, C). \quad (21.103)$$

This relation directly connects changes in entropy to changes in potential temperature

$$d\mathcal{S} = \left[\frac{\partial \mathcal{S}(\theta, p_R, C)}{\partial \theta} \right]_C d\theta. \quad (21.104)$$

Consequently, the reversible transport of a fluid element with constant matter concentration ($dC = 0$) occurs with both a constant specific entropy and constant potential temperature.

We can go even further than the relation (21.104) by recalling that equation (21.94) relates increments in specific entropy to temperature and pressure

$$T d\mathcal{S} = c_p (dT - \hat{\Gamma} dp), \quad (21.105)$$

where $\hat{\Gamma}$ is the adiabatic lapse rate defined in terms of pressure changes (equation (21.93)), and we set $dC = 0$. To relate $dT - \hat{\Gamma} dp$ to $d\theta$ we write the potential temperature equation (21.98) in the form

$$\theta(T, p_R, C) = T - \int_{p_R}^p \hat{\Gamma}(T, p', C) dp' = T - \Psi(T, p, C, p_R), \quad (21.106)$$

so that the differential increments are given by

$$d\theta = dT - d\Psi. \quad (21.107)$$

We evaluate $d\Psi$ using the chain rule and then specialize to the case of constant composition and with a fixed reference pressure

$$d\Psi = \frac{\partial \Psi}{\partial T} dT + \frac{\partial \Psi}{\partial p} dp + \frac{\partial \Psi}{\partial C} dC + \frac{\partial \Psi}{\partial p_R} dp_R \quad (21.108a)$$

$$= \frac{\partial \Psi}{\partial T} dT + \frac{\partial \Psi}{\partial p} dp \quad (21.108b)$$

$$\equiv \int_{p_R}^p \frac{\partial \hat{\Gamma}(T, p', C)}{\partial T} dp' + \hat{\Gamma}(T, p, C) dp. \quad (21.108c)$$

Evaluating this increment at the reference pressure removes the integral so that

$$d\Psi = \hat{\Gamma}(T, p_R, C) dp, \quad (21.109)$$

in which case the potential temperature increment is

$$d\theta = dT - \hat{\Gamma}(T, p_R, C) dp. \quad (21.110)$$

Making use of this relation in equation (21.105) renders an expression for the entropy increment in terms of the potential temperature increment

$$d\mathcal{S} = c_p \theta^{-1} d\theta \quad p = p_R \text{ and } dC = 0. \quad (21.111)$$

Although evaluated at the reference pressure, as part of exercise 22.4 we see that this relation holds for an ideal gas at all pressures. Furthermore, as part of exercise 22.5 we see that this relation also holds for all pressures in certain liquids.

21.9 A simple ideal gas atmosphere

In an ideal gas, intermolecular forces between molecules are ignored. Also, the molecules in an ideal gas are assumed to occupy zero volume (i.e., they are point particles). As a result, the internal energy of an ideal gas is just due to translation, rotation, and vibration of molecules. We refer to a *simple ideal gas* as an ideal gas where the internal energy is a linear function of temperature. In this section we develop a variety of thermodynamic relations for a simple ideal gas atmosphere in exact hydrostatic balance. Although the real atmosphere is moving, and thus not in exact hydrostatic balance, and the real atmosphere is not a simple ideal gas, it turns out that many of the relations established here are rather accurate approximations to the real atmosphere. Furthermore, by exposing these relations for the ideal gas, we further our understanding of the more general thermodynamic relations established earlier in this chapter.

21.9.1 Equation of state

An ideal gas satisfies the following equation of state (see Section 12.3.2)

$$PV = n R_g T, \quad (21.112)$$

where p is the pressure, V is the volume, n is the number of moles,

$$R_g = 8.314 \text{ J mole}^{-1} \text{ K}^{-1} = 8.314 \text{ kg m}^2 \text{ s}^{-2} \text{ mole}^{-1} \text{ K}^{-1} \quad (21.113)$$

is the *universal gas constant*, and T is the absolute temperature in Kelvin (see Section 12.3.2). The number of moles equals to the mass, M , of the gas divided by the mass per mole, M_{mole}

$$n = M/M_{\text{mole}}. \quad (21.114)$$

The mass density, $\rho = M/V$, is thus given by

$$\rho = \frac{p M_{\text{mole}}}{T R_g} \equiv \frac{p}{T R^M}, \quad (21.115)$$

where

$$R^M = R_g/M_{\text{mole}} \quad (21.116)$$

is the *specific gas constant* as defined by the universal gas constant normalized by the molar mass for the constituent. For air we have (Section 12.3.2)

$$M_{\text{air}} = 28.8 \times 10^{-3} \text{ kg mole}^{-1} \quad (21.117)$$

so that air's specific gas constant is

$$R^{\text{air}} = \frac{R_g}{M_{\text{air}}} = \frac{8.314 \text{ kg m}^2 \text{ s}^{-2} \text{ mole}^{-1} \text{ K}^{-1}}{28.8 \times 10^{-3} \text{ kg mole}^{-1}} = 2.938 \times 10^2 \text{ m}^2 \text{ s}^{-2} \text{ K}^{-1}. \quad (21.118)$$

The relation (21.115) is known as a *thermal equation of state*, or more succinctly just an equation of state (see Section 25.3 for more discussion). It shows that the mass density of an ideal gas is directly proportional to the pressure: increasing pressure increases density. In contrast, mass density is inversely proportional to the temperature: increases in temperature lead to lower mass density. This behavior for the ideal gas density is reflected in certain real gases and liquids.³

³A notable counter-example is water near its freezing point, which becomes more dense as temperature rises. This anomalous behavior is why a body of water freezes from the top down rather than from the bottom up.

21.9.2 Internal energy

An ideal gas is comprised of molecules that interact only through elastic collisions. There are no inter-molecular forces. Furthermore, the volume of the individual molecules is ignored in comparison to the volume of empty space between the molecules, so they are approximated as point masses. Consequently, the internal energy for an ideal gas is independent of density and of the matter concentration. It is hence a function only of the temperature, which measures the kinetic energy of the elastic point molecules

$$\mathcal{I} = \mathcal{I}(T) \quad \text{ideal gas.} \quad (21.119)$$

Consequently, the exact differential of internal energy for an ideal gas is

$$d\mathcal{I} = c_v dT. \quad (21.120)$$

The appearance of c_v , the constant volume specific heat capacity discussed in Section 21.5.1, arises in order for the ideal gas internal energy to satisfy the general equation (21.76). The heat capacity for an ideal gas is generally a function of temperature. However, for many applications it is sufficient to consider a simple ideal gas, in which c_v is a constant so that

$$\mathcal{I} = c_v T + \text{constant} \quad \text{simple ideal gas.} \quad (21.121)$$

The arbitrary constant of integration is generally set to zero so that the internal energy vanishes at absolute zero.

21.9.3 Heat capacity

The heat capacity is a constant for a simple ideal gas (equation 21.121). Results from statistical mechanics show that the thermal/internal energy per molecule equals to $k_B T/2$ per excited molecular degree of freedom, where

$$k_B = 1.3806 \times 10^{-23} \text{ m}^2 \text{ kg s}^{-2} \text{ K}^{-1} \quad (21.122)$$

is the *Boltzmann constant*. Dry air is mostly comprised of the diatomic molecules N_2 and O_2 . Diatomic molecules at temperatures of the lower atmosphere have two rotational and three translational degrees of freedom,⁴ so that $\mathcal{I}_{\text{molecule}} = 5 k_B T/2$.

We convert this energy per molecule to an energy per mole of diatomic molecules by multiplying by Avogadro's number (equation (12.9))

$$\mathcal{I}_{\text{mole diatomic}} = 5 A_v k_B T/2 = 5 R_g T/2, \quad (21.123)$$

where the gas constant is given by

$$R_g = A_v k_B \quad (21.124a)$$

$$= (6.022 \times 10^{23} \text{ mole}^{-1}) (1.3806 \times 10^{-23} \text{ m}^2 \text{ kg s}^{-2} \text{ K}^{-1}) \quad (21.124b)$$

$$= 8.314 \text{ kg m}^2 \text{ s}^{-2} \text{ mole}^{-1} \text{ K}^{-1}. \quad (21.124c)$$

Finally, dividing by the molar mass for dry air (equation (12.10))

$$M_{\text{air}} = 0.028 \text{ kg mole}^{-1} \quad (21.125)$$

⁴ At high temperatures, two vibrational degrees of freedom are also excited so that $\mathcal{I}_{\text{molecule}} = 7 k_B T/2$ at high temperatures.

leads to the simple ideal gas approximation to the dry air heat capacity

$$c_v = \frac{5 R_g}{2 M_{\text{air}}} = 742 \text{ m}^2 \text{ s}^{-2} \text{ K}^{-1}. \quad (21.126)$$

The measured heat capacity for dry air at standard temperature (300 K) is $718 \text{ m}^2 \text{ s}^{-2} \text{ K}^{-1}$, so the simple ideal gas estimate is only $(742 - 718)/718 = 3.3\%$ too large.

21.9.4 Enthalpy

The enthalpy is generally given by equation (21.49), which for a simple ideal gas takes the form

$$\mathcal{H} = \mathcal{I} + p/\rho = c_v T + \frac{T R_g}{M_{\text{mole}}} = T (c_v + R^M) \quad (21.127)$$

where $R^M = R_g/M_{\text{mole}}$ (equation (21.116)) is the specific gas constant for the gas. Recall that the constant pressure heat capacity is given by equation (21.77)

$$c_p = T \left[\frac{\partial \mathcal{S}}{\partial T} \right]_{p,C} = \left[\frac{\partial \mathcal{H}}{\partial T} \right]_{p,C}. \quad (21.128)$$

Consequently, for a simple ideal gas we have

$$c_p = c_v + R^M \quad \text{and} \quad \mathcal{H} = c_p T. \quad (21.129)$$

21.9.5 Thermal expansion coefficient

The thermal expansion coefficient for an ideal gas is given by

$$\alpha_T = -\frac{1}{\rho} \frac{\partial \rho}{\partial T} = \frac{1}{T}, \quad (21.130)$$

so that as temperature increases the thermal expansion decreases.

21.9.6 Fundamental thermodynamic relations

The fundamental thermodynamic relation, written in terms of internal energy (equation (21.33b)) and enthalpy (equation (21.49)), are given by

$$d\mathcal{I} = T d\mathcal{S} - p d\nu_s + \mu dC \quad (21.131)$$

$$d\mathcal{H} = T d\mathcal{S} + \nu_s dp + \mu dC. \quad (21.132)$$

For a simple ideal gas these relations take the form

$$c_v dT = T d\mathcal{S} - p d\nu_s + \mu dC \quad (21.133)$$

$$c_p dT = T d\mathcal{S} + \nu_s dp + \mu dC. \quad (21.134)$$

21.9.7 Isothermal compressibility

The isothermal compressibility measures the change in volume when holding the temperature and matter concentration fixed and it is determined by the partial derivatives

$$-\frac{1}{V} \left[\frac{\partial V}{\partial p} \right]_{T,C} = \frac{1}{\rho} \left[\frac{\partial \rho}{\partial p} \right]_{T,C}. \quad (21.135)$$

For an ideal gas the compressibility is given by

$$-\frac{1}{V} \left[\frac{\partial V}{\partial p} \right]_{T,C} = \frac{1}{p}, \quad (21.136)$$

so that the compressibility decreases when pressure increases.

21.9.8 Sound speed

Sound travels through a fluid through compression and expansion (Section 18.5), so we expect the sound speed to be related to the compressibility. But rather than the isothermal compressibility considered above, sound waves are largely adiabatic waves so that the entropy is constant. We are thus in need of the isentropic compressibility to compute the sound speed. That is, as defined by equation (21.80), the sound speed is the pressure derivative of density computed with entropy and matter concentration held fixed. We make use of the fundamental relations (21.133) and (21.134), with $dS = 0$ and $dC = 0$ to have

$$\frac{c_v}{c_p} = \frac{p}{\rho} \left[\frac{\partial \rho}{\partial p} \right]_{S,C} = (p/\rho) c_s^{-2} \implies c_s^2 = (p/\rho)(c_p/c_v) = T R^M (c_p/c_v). \quad (21.137)$$

For an ideal diatomic gas, such as nitrogen and oxygen, the ratio $c_p/c_v = 7/5$. Taking $R^M = 2.938 \times 10^2 \text{ m}^2 \text{ s}^{-2} \text{ K}^{-1}$ for air from equation (21.118) then leads to

$$c_s \approx 350 \text{ m s}^{-1} \quad \text{for } T = 300 \text{ K}. \quad (21.138)$$

21.9.9 Adiabatic lapse rate

For an ideal gas, the thermal expansion coefficient is given by (equation (21.130)) $\alpha_T = T^{-1}$ so that the lapse rates are

$$\hat{\Gamma} = \frac{1}{\rho c_p} \quad \text{and} \quad \Gamma = -\frac{g}{c_p}. \quad (21.139)$$

The measured specific heat capacity for a dry atmosphere at standard temperature (300 K) is

$$c_p = 1005 \text{ m}^2 \text{ s}^{-2} \text{ K}^{-1} \quad (21.140)$$

so that the adiabatic lapse rate for a dry atmosphere is roughly

$$\Gamma_d = -9.8 \text{ K}/(1000 \text{ m}). \quad (21.141)$$

Hence, temperature decreases by nearly 10 K when rising 1000 m in a dry and ideal gas atmosphere.

21.9.10 Geopotential thickness

We now establish basic relations for a static atmosphere satisfying the hydrostatic balance. These relations also hold to a very good approximation for the large-scale atmosphere given the dominance of approximate hydrostatic balance for these scales (see Section 23.3).

From the hydrostatic equation (18.38) we know that the pressure on a geopotential, Φ_1 , equals to

$$p(\Phi_1) = \int_{\Phi_1}^{\infty} \rho(\Phi) d\Phi \quad (21.142)$$

where we assumed that $p(\Phi_2 = \infty) = 0$. Equation (18.37) allows us to write the integrand as

$$d\Phi = -\rho^{-1} dp = -\frac{T R^{\text{air}} dp}{p}, \quad (21.143)$$

where the second equality assumed an ideal gas atmosphere. Vertical integration of equation (21.143) leads to the *hypsometric equation*, which provides the geopotential thickness between two pressure isosurfaces

$$\Phi(z_2) - \Phi(z_1) = -R^{\text{air}} \int_{p_1}^{p_2} T d(\ln p). \quad (21.144)$$

Recall that $dp < 0$ if $dz > 0$ since the hydrostatic pressure decreases when moving up in the atmosphere. We define the *geopotential height* according to

$$Z = \Phi/g, \quad (21.145)$$

where g is the gravitational acceleration at sea level. The geopotential height is close to the geometric height in the troposphere and lower stratosphere. The hypsometric equation (21.144) says that the geopotential thickness between two isobars is

$$Z_2 - Z_1 = \frac{R^{\text{air}}}{g} \int_{p_2}^{p_1} T d(\ln p). \quad (21.146)$$

Defining the layer mean temperature

$$\langle T \rangle = \frac{\int_{p_2}^{p_1} T d(\ln p)}{\int_{p_2}^{p_1} d(\ln p)} \quad (21.147)$$

and the layer mean *scale height*

$$H = \frac{R^{\text{air}} \langle T \rangle}{g} \quad (21.148)$$

leads to the geopotential thickness

$$Z_2 - Z_1 = -H \ln(p_2/p_1) \quad (21.149)$$

The geopotential thickness is thus directly proportional to the mean temperature within the pressure layer, with thicker layers, for example, with higher mean temperatures.

We can invert the geopotential thickness relation (21.149) for the pressures to render

$$p_1 = p_2 e^{-(Z_1 - Z_2)/H}. \quad (21.150)$$

The scale height is a function of pressure through its dependence on the layer averaged temperature in equation (21.148). For the special case of an atmosphere with a constant temperature T (unrealistic but useful special limit), then the scale height is a constant

$$H_{\text{const}} = \frac{R^{\text{air}} T_{\text{const}}}{g}. \quad (21.151)$$

Setting $T = 300$ K and using the specific gas constant for air from equation (21.118) leads to the scale height

$$H_{\text{const}} = \frac{2.938 \times 10^2 \text{ m}^2 \text{ s}^{-2} \text{ K}^{-1} \times 300 \text{ K}}{9.8 \text{ m s}^{-2}} \approx 9 \times 10^3 \text{ m}. \quad (21.152)$$

It is furthermore convenient to set $Z_2 = 0$ with $p_2 = p_{\text{slp}}$ the sea level pressure, whose global average is

$$\langle p_{\text{slp}} \rangle = 101.325 \times 10^3 \text{ N m}^{-2}. \quad (21.153)$$

The pressure in an isothermal atmosphere thus decreases exponentially with geopotential height according to the scale height

$$p(Z) = \langle p_{\text{slp}} \rangle \exp(-Z/H_{\text{const}}). \quad (21.154)$$

21.9.11 Potential temperature

The fundamental thermodynamic relation for a simple ideal gas (21.134) takes on the following form for an isentropic change

$$c_p dT = \nu_s dp. \quad (21.155)$$

Dividing both sides by temperature and using the ideal gas relation

$$\frac{\nu_s}{T} = \frac{R^M}{p} \quad (21.156)$$

leads to

$$c_p d(\ln T) = R^M d(\ln p). \quad (21.157)$$

Since c_p and R^M are constants, we can integrate this relation from the reference pressure to an arbitrary pressure

$$c_p \int_{\theta}^T d(\ln T) = R^M \int_{p_R}^p d(\ln p), \quad (21.158)$$

which renders the explicit expression for the potential temperature of a simple ideal gas

$$\theta = T \left[\frac{p_R}{p} \right]^{R^M/c_p} \quad \text{where} \quad c_p = \frac{7 R^M}{2}, \quad (21.159)$$

with c_p the constant pressure heat capacity of a simple ideal gas of diatomic molecules (Section 21.9.3). In Exercise 21.5 we show that $\partial\theta/\partial p = 0$ for the ideal gas, thus exemplifying the removal of explicit pressure effects from the potential temperature. Furthermore, it follows from equation (21.159) that increments in potential temperature are related to increments in temperature and pressure via

$$\frac{\delta\theta}{\theta} = \frac{\delta T}{T} - \frac{\delta p}{p}. \quad (21.160)$$

In particular, if the increment is between points in space within a fluid, then we are led to the relationship between gradients

$$\frac{\nabla\theta}{\theta} = \frac{\nabla T}{T} - \frac{\nabla p}{p}. \quad (21.161)$$

21.9.12 Further study

Atmospheric sciences and dynamic meteorology books have thorough discussions of ideal gas thermodynamics. Some of the material in section 1.6 of [Holton \(1992\)](#) was used in the present section.

21.10 Exercises

EXERCISE 21.1: DERIVATION OF THE GIBBS-DUHEM RELATION

Show all of the steps leading to the Gibbs-Duhem relation (21.24).

EXERCISE 21.2: CHEMICAL POTENTIAL IDENTITY

As seen in Section 22.4.7, we have need to consider the partial derivative

$$\left[\frac{\partial \mathcal{H}}{\partial C} \right]_{T,p} \quad (21.162)$$

when determining the chemical work done by mixing within a fluid. Prove the identity

$$\left[\frac{\partial \mathcal{H}}{\partial C} \right]_{T,p} = \left[\frac{\partial \mathcal{H}}{\partial C} \right]_{S,p} - T \left[\frac{\partial \mu}{\partial T} \right]_{C,p} = \mu - T \left[\frac{\partial \mu}{\partial T} \right]_{C,p}. \quad (21.163)$$

Hint: as seen in Section 21.4.3, the natural functional dependence for enthalpy is $\mathcal{H}(S, p, C)$, whereas in Section 21.4.5 we found the natural function dependence of the Gibbs potential to be $\mathcal{G}(T, p, C)$. Equate the exact differential expressions for enthalpy using the two functional dependencies $\mathcal{H}(S, p, C)$ and $\mathcal{G}(T, p, C)$, and then derive a Maxwell relation based on the the fundamental thermodynamic relation written in terms of the Gibbs potential.

EXERCISE 21.3: CONSTANT OF MOTION FOR ADIABATIC FLOW

Show that for a simple ideal gas, isentropic flow (i.e., both adiabatic and of constant matter concentration) maintains

$$p \nu_s^{c_p/c_v} = \text{constant}, \quad (21.164)$$

where $\nu_s = \rho^{-1}$ is the specific volume.

EXERCISE 21.4: GEOPOTENTIAL HEIGHT

The *geopotential height* is the height above the earth of a chosen pressure surface.

- (a) Show that an ideal gas atmosphere in exact hydrostatic balance with a uniform lapse rate

$$\frac{dT}{dz} = -|\Gamma| = \text{constant} \quad (21.165)$$

has a geopotential height at a pressure p given by

$$z = \frac{T_0}{|\Gamma|} \left[1 - \left(\frac{p_0}{p} \right)^{-R^M |\Gamma| / g} \right], \quad (21.166)$$

where T_0 is the temperature at $z = 0$.

- (b) For an isothermal atmosphere, obtain an expression for the geopotential height as a function of pressure, and show that this result is consistent with the expression (21.166) in the appropriate limit.

EXERCISE 21.5: POTENTIAL TEMPERATURE FOR AN IDEAL GAS

Show that $\partial\theta/\partial p = 0$ for the potential temperature of an ideal gas given by equation (21.159)

$$\theta = T \left[\frac{p_R}{p} \right]^{R^M/c_p}. \quad (21.167)$$

Hint: remember that $\partial T/\partial p \neq 0$.

EXERCISE 21.6: THERMODYNAMIC RELATIONS FOR AN ATMOSPHERE

In this exercise, we establish some relations for an ideal gas atmosphere, and one relation holding for an arbitrary equation of state. We assume that the gravitational acceleration is constant throughout the full depth of the atmosphere. This assumption becomes questionable when integrating to the top of the atmosphere. We furthermore ignore differences in the horizontal cross-sectional area of a fluid column at the bottom and top of the atmosphere arising from the spherical nature of the planet. These two assumption are sufficient for our purposes.

- (a) **PRESSURE-HEIGHT IDENTITY:** Prove the following identity and state your assumptions

$$\int_0^{p_s} z \, dp = \int_{z=0}^{z_{top}} p \, dz. \quad (21.168)$$

This identity will be of use for some of the following questions.

- (b) **IDEAL GAS $\mathcal{I} + \Phi$ INTEGRATED OVER DEPTH OF A HYDROSTATIC ATMOSPHERE:** For an ideal gas atmosphere in exact hydrostatic balance, show that the integral of the gravitational potential energy plus internal energy from the surface to the top of the atmosphere is equal to the integral of the enthalpy of the atmosphere

$$\int_0^{z_{top}} (\Phi + \mathcal{I}) \rho \, dz = \int_0^{z_{top}} \mathcal{H} \rho \, dz, \quad (21.169)$$

where

$$\mathcal{H} = p \alpha + \mathcal{I} \quad (21.170)$$

is the enthalpy per mass,

$$\Phi = g z \quad (21.171)$$

is the simple form of the geopotential, which is also the gravitational potential energy per mass (Section 10.12.4), and \mathcal{I} is the internal energy per mass. The height integral extends from the surface where $z = 0$, to the top of the atmosphere where $z = z_{top}$.

- (c) **VERTICAL DERIVATIVE OF DRY STATIC ENERGY:** For an ideal gas atmosphere in hydrostatic balance, show that

$$\frac{d\sigma}{dz} = \Pi \frac{d\theta}{dz}, \quad (21.172)$$

where

$$\sigma = \mathcal{H} + \Phi \quad (21.173)$$

is the dry static energy and

$$\Pi = \frac{c_p T}{\theta} \quad (21.174)$$

is the *Exner function*.

- (d) FIRST IDENTITY FOR HORIZONTAL PRESSURE GRADIENT: For an ideal gas atmosphere (either hydrostatic or non-hydrostatic), derive the following expression for the pressure gradient force

$$-\frac{1}{\rho} \nabla p = -\theta \nabla \Pi. \quad (21.175)$$

- (e) SECOND IDENTITY FOR HORIZONTAL PRESSURE GRADIENT: For an ideal gas atmosphere (either hydrostatic or non-hydrostatic), derive the following expression for the pressure gradient force

$$-\frac{1}{\rho} \nabla p = -\frac{c_s^2}{\rho \theta} \nabla(\rho \theta), \quad (21.176)$$

where c_s is the sound speed.

- (f) $\mathcal{I} + \Phi$ INTEGRATED OVER DEPTH OF A HYDROSTATIC ATMOSPHERE: Show that for a hydrostatic atmosphere with an arbitrary equation of state

$$\int_0^{p_s} (\Phi + \mathcal{I}) dp = \int_0^{p_s} \mathcal{H} dp. \quad (21.177)$$

That is, show that the relation in the first part of this problem holds even without making the ideal gas assumption.

EXERCISE 21.7: UNIT KNUDSEN NUMBER

Recall from Section 12.2.2 that the Knudsen number is the ratio $\text{Kn} = L_{\text{mfp}}/L_{\text{macro}}$, where $L_{\text{macro}} \approx 10^{-4}$ m is the macroscopic length scale used in the discussion of the continuum approximation, and L_{mfp} is the molecular mean free path (Section 12.3.3). Throughout this exercise make use of $p_{\text{stand}} = 101.325 \times 10^3$ Pa for standard atmospheric pressure.

- (a) Consider a mole of an isothermal and ideal gas atmosphere of $T = 300\text{K}$ with a constant gravitational acceleration. At what pressure is the Knudsen number unity? Write your answer as a fraction of standard sea atmospheric pressure, p_{stand} .
- (b) Compute the altitude corresponding to the above pressure, assuming the sea level pressure is p_{slp} and the geopotential is $\Phi = g z$. Hint: make use of results from Section 21.9.10.
- (c) Assuming $p = p_{\text{stand}}$, at what temperature is $\text{Kn} = 1$? Hint: assume the ideal gas law holds regardless the temperature.
- (d) Comment on what is the least atmospherically relevant assumption made during this exercise.



Energy dynamics

The energy of a static fluid consists of the potential energy of the fluid within a gravitational field plus the internal energy arising from molecular degrees of freedom. In this chapter, we are concerned with moving fluids, in which the macroscopic motion possesses kinetic energy as well as potential energy. We develop the dynamical equations for mechanical energy (kinetic plus gravitational potential) of this macroscopic motion and examine how it can be exchanged with internal energy.

As part of this analysis we extend the equilibrium thermodynamics of Chapter 21 to include time dependent moving fluid phenomena. Making this transition requires us to assume that moving fluid elements are in *local thermodynamic equilibrium*. This assumption is based on the quasi-static nature of macroscopic motions (Section 21.1.3) whereby microscopic motions have a much shorter equilibration time relative to the slower macroscopic processes of interest for fluid mechanics such as advection, waves, turbulence, and turbulent mixing. The extremely wide time scale separation between the microscopic and macroscopic allows us to use phenomenological thermodynamic laws to develop evolution equations for continuous thermodynamic properties of moving continuum fluid elements. Consistent with the time scale separation, we can assume that the macroscopic motion of a fluid element does not alter its entropy. That is, advective transport in the absence of mixing is a reversible process. In contrast, mixing of properties between fluid elements, whether through molecular or turbulent processes, is irreversible and thus increases entropy.

The application of thermodynamics to a moving fluid is known as *quasi-equilibrium thermodynamics* or *linear irreversible thermodynamics*, with fundamentals developed in [DeGroot and Mazur \(1984\)](#) and [Landau and Lifshitz \(1987\)](#). The term “linear” refers to an assumption that the system is close to thermodynamic equilibrium throughout its motion. This assumption is very accurate for the atmospheric and oceanic motions studied in this book, whereby local thermodynamic equilibrium is maintained for fluid elements even as those elements move as a result of imbalances in macroscopic forces. Thermodynamic fluxes are thus linear functions of the gradients of the thermodynamic state variables. The assumption of local thermodynamic equilibrium breaks down in the rarefied gas dynamics of the upper atmosphere and is outside the scope of this book.

READER’S GUIDE FOR THIS CHAPTER

This chapter builds from the momentum dynamics of Chapter 18 and the thermodynamics of Chapter 21. We develop a variety of energy budgets and consider their physical content. We also study the heat equation as represented by the equation for potential enthalpy, with variations on the heat equation and the transport of enthalpy of primary importance for large-scale ocean and atmospheric phenomena and climate dynamics.

22.1	Loose threads	459
22.2	Gravitational potential energy	459
22.2.1	Material evolution	459
22.2.2	Reference geopotential	460
22.2.3	Regionally integrated gravitational potential energy	460
22.2.4	Potential energy and vertical stratification	461
22.2.5	Gravitational potential energy and mixing	462
22.3	Kinetic energy and mechanical energy	463
22.3.1	Kinetic energy	463
22.3.2	Frictional dissipation of kinetic energy	464
22.3.3	Mechanical energy budget	465
22.3.4	Further study	466
22.4	Thermodynamics of a moving fluid	466
22.4.1	Concerning the transition to a moving fluid	466
22.4.2	Thermodynamic properties as space-time fields	467
22.4.3	Thermodynamic relations for steady state flow	467
22.4.4	Materially constant specific entropy for a perfect fluid	468
22.4.5	First law for a moving fluid element	468
22.4.6	Enthalpy budget	469
22.4.7	Thermal and chemical processes and fluxes	469
22.4.8	First law for homogeneous fluids in terms of potential temperature .	470
22.4.9	Further study	470
22.5	Budget for total energy	470
22.5.1	Postulating the budget for total energy	470
22.5.2	Budgets for internal energy and enthalpy	471
22.5.3	Joule heating from friction	471
22.5.4	Comments on gauge symmetry	472
22.5.5	Further study	472
22.6	Bernoulli potential	472
22.6.1	Bernoulli potential and mechanical injection work	472
22.6.2	Bernoulli's theorem for a steady perfect fluid	473
22.6.3	Traditional derivation of Bernoulli's theorem	474
22.6.4	Steady flow of constant density perfect fluid in a pipe	475
22.6.5	Further study	476
22.7	Entropy budget for the ocean	476
22.7.1	Non-advective entropy flux and entropy source	477
22.7.2	Constraints from the second law of thermodynamics	478
22.7.3	A summary presentation	479
22.7.4	Comments and further study	480
22.8	Temperature evolution	480
22.8.1	Enthalpy equation again	480
22.8.2	Evolution of <i>in situ</i> temperature	480
22.8.3	Evolution of potential temperature	481
22.8.4	Conservative Temperature for the ocean	481
22.8.5	Further study	483
22.9	Atmospheric moist static energy	483
22.9.1	Further study	483
22.10	Summary of the ocean-atmosphere equations	483
22.11	Concerning conservation laws	485
22.11.1	Flux-form conservation laws	485
22.11.2	Conservation laws that are not flux-form	486
22.11.3	Non-material or wave-like transport of properties	486
22.11.4	Local versus global conservation	486
22.11.5	Concerning potential scalar properties	487

22.1 Loose threads

- Figures needed
- Moist static energy as per David Romps' notes in Section 22.9.
- Available internal energy
- Potential energy as per *Landau and Lifshitz* (1987) and possible extention to APE.
- More discussion of the entropy budget in Section 22.7.
- Discuss mixing of T versus Θ in Section 22.8.

22.2 Gravitational potential energy

Geophysical fluids move within a gravitational field created by the mass of the planet, including the mass of the fluid itself. We typically focus on a rather simple form for the effective gravitational acceleration exemplified by the geopotential $\Phi = g z$, where g is the gravitational acceleration that includes effects from both central gravity plus planetary centrifugal (Section 10.12.4). However, we offer some discussion of astronomical tide producing forces in Chapter 29, whereby the geopotential is a more complicated function of space and time, $\Phi(\mathbf{x}, t)$. In this section we study the potential energy of a fluid element due to its presence in a gravitational field, with the geopotential providing the effective gravitational potential energy per mass of fluid elements.¹

22.2.1 Material evolution

The gravitational potential energy per mass of a fluid element is given by the geopotential, Φ , so that the gravitational potential energy is $\Phi \rho \delta V$. Hence, the evolution of potential energy for a constant mass fluid element is given by

$$\frac{D(\Phi \rho \delta V)}{Dt} = \rho \delta V \frac{D\Phi}{Dt}, \quad (22.1)$$

where $D(\rho \delta V)/Dt = 0$ since the fluid element has a constant mass. The material time derivative for the geopotential

$$\frac{D\Phi}{Dt} = \frac{\partial \Phi}{\partial t} + \mathbf{v} \cdot \nabla \Phi, \quad (22.2)$$

contains a local time dependence that arises from astronomical tide forcing or movement of mass on the planet (Chapter 29). As further explored in Section 22.3, the advective term represents an exchange of mechanical energy between the kinetic energy contained in fluid motion and the gravitational potential energy due to the fluid being within a gravitational field. This energy exchange arises from fluid motion across constant geopotential surfaces. For example, motion up

¹In the following we generally refer to the more concise “gravitational potential energy” or even briefer “potential energy” rather than “potential energy from the effective gravitational field.”

the geopotential gradient, $\mathbf{v} \cdot \nabla \Phi > 0$, increases gravitational potential energy and motion down the geopotential gradient decreases potential energy. With the geopotential $\Phi = g z$, we have

$$\mathbf{v} \cdot \nabla \Phi = g w, \quad (22.3)$$

so that vertically upward motion ($w > 0$) increases potential energy.

22.2.2 Reference geopotential

There is no change to the energetics if we modify the gravitational reference state by modifying the geopotential

$$\Phi \rightarrow \Phi + \Phi_r \quad (22.4)$$

with Φ_r an arbitrary constant. In particular, this offset has no effect on the evolution of gravitational potential energy of the constant mass fluid element since

$$\frac{D(\Phi_r \rho \delta V)}{Dt} = \Phi_r \frac{D(\rho \delta V)}{Dt} = 0. \quad (22.5)$$

Hence, as is well known from classical mechanics, it is not the value of the gravitational potential energy that is important, but instead it is the space and time changes that affect energetics.

22.2.3 Regionally integrated gravitational potential energy

Now consider the evolution of the gravitational potential energy integrated over a finite region \mathcal{R} . If the fluid region is closed to mass transport, as per a material boundary, then we can make use of the Leibniz-Reynolds transport theorem in the form of equation (16.41d) to write

$$\frac{d}{dt} \int_{\mathcal{R}} \Phi \rho dV = \int_{\mathcal{R}} \frac{D\Phi}{Dt} \rho dV, \quad (22.6)$$

which is an extension of the material evolution equation (22.1). If the region is open to material mass transport, we make use Leibniz-Reynolds transport theorem in the form of equation (16.34) to find

$$\frac{d}{dt} \int_{\mathcal{R}} \rho \Phi dV = \int_{\mathcal{R}} \left[\frac{\partial(\rho \Phi)}{\partial t} + \nabla \cdot (\rho \Phi \mathbf{v}^{(b)}) \right] dV \quad (22.7a)$$

$$= \int_{\mathcal{R}} \left[\rho \frac{D\Phi}{Dt} - \nabla \cdot (\rho \Phi \mathbf{v}) + \nabla \cdot (\rho \Phi \mathbf{v}^{(b)}) \right] dV \quad (22.7b)$$

$$= \int_{\mathcal{R}} \left[\rho \frac{D\Phi}{Dt} + \nabla \cdot [\rho \Phi (\mathbf{v}^{(b)} - \mathbf{v})] \right] dV \quad (22.7c)$$

$$= \int_{\mathcal{R}} \rho \frac{D\Phi}{Dt} dV + \oint_{\partial\mathcal{R}} \rho \Phi (\mathbf{v}^{(b)} - \mathbf{v}) \cdot \hat{\mathbf{n}} d\mathcal{S}, \quad (22.7d)$$

where $\mathbf{v}^{(b)}$ is the velocity of a point on the boundary of the domain. The evolution thus consists of the mass integrated material time evolution of the geopotential, plus a surface term that contributes to the transport of the geopotential across the regional boundaries.

If the region is a vertical column of ocean fluid with fixed horizontal cross-section, extending from the ocean surface to the ocean bottom, then there is horizontal transport across the vertical

column bounds, plus vertical transport of mass across the ocean free surface. For the free surface we make use of the surface kinematic boundary condition (15.73c) to write

$$\int_{z=\eta} \rho \Phi (\mathbf{v}^{(b)} - \mathbf{v}) \cdot \hat{\mathbf{n}} d\mathcal{S} = \int_{z=\eta} Q_m \Phi dA. \quad (22.8)$$

In this equation, Q_m is the mass per time per horizontal area of matter crossing the ocean free surface at $z = \eta$ where $Q_m > 0$ for matter entering the ocean domain, and $d\mathcal{S}$ is the area element on the free surface with dA its horizontal projection. As noted in Section 22.2.2, we can add a constant to the geopotential without affecting the energetics, which is here seen by noting that mass conservation means that

$$\frac{d}{dt} \int \rho dV = - \int_{\mathcal{R}} \nabla \cdot [\rho (\mathbf{v} - \mathbf{v}^{(b)})] dV, \quad (22.9)$$

which results from setting $\Phi = 1$ in equation (22.7d). In general, we expect the transfer of mass across the surface boundary to affect the gravitational potential energy both because it adds or removes mass to the ocean domain, and because it affects the geopotential. To help interpret the sign from the boundary term, it is useful to define the reference state geopotential so that $\Phi > 0$ at the ocean surface, no matter what the value of η . We can do so by defining the reference geopotential at or below the ocean bottom. In this case, adding mass increases the gravitational potential energy and removing mass reduces it.

Consider the special case of a geopotential $\Phi = g z$, so returning to a $z = 0$ reference state, in which case the global ocean potential energy equation is written

$$\frac{d}{dt} \int \rho z dV = \int \rho w dV + \int Q_m \eta dA, \quad (22.10)$$

where we cancelled the constant gravitational acceleration. Now decompose Q_m and η into their global area means and deviations

$$Q_m = \overline{Q_m} + Q'_m \quad \text{and} \quad \eta = \bar{\eta} + \eta', \quad (22.11)$$

so that

$$g \int Q_m \eta dA = g \overline{Q_m} \bar{\eta} A + g \int Q'_m \eta' dA. \quad (22.12)$$

As before, the $\overline{Q_m} \bar{\eta}$ term alters potential energy relative to the arbitrary reference state, here taken as $z = 0$. The area correlation term increases potential energy in regions where $Q'_m \eta' > 0$, which acts to increase the relative deviation of the free surface from its mean value. That is, $Q'_m \eta' > 0$ in regions where $Q'_m > 0$ and $\eta' > 0$ as well as in regions where $Q'_m < 0$ and $\eta' < 0$. Conversely, the correlation term reduces potential energy where Q'_m and η' are anti-correlated, which acts to decrease the relative deviation of the free surface height.

22.2.4 Potential energy and vertical stratification

Consider the potential energy of a region of horizontally homogeneous fluid centered at a vertical position, $z = z_c$, and with constant horizontal cross-sectional area, A . Assuming we do not move vertically far away from the central position, we can write the density in the linear form

$$\rho(z) = \rho(z_c) + \frac{d\rho(z_c)}{dz} (z - z_c) \equiv \rho_c - K(z - z_c), \quad (22.13)$$

where $K = -d\rho(z_c)/dz > 0$ is a shorthand for the vertical density gradient at the central point. The potential energy per volume ($\Delta V = A \Delta z$) for fluid in the vertical region $z \in [z_c - \Delta z/2, z_c + \Delta z/2]$ is given by

$$(g/\Delta z) \int z \rho dz = (g/\Delta z) \int_{z_c - \Delta z/2}^{z_c + \Delta z/2} [\rho_c - K(z - z_c)] z dz \quad (22.14a)$$

$$= g \rho_c z_c - g K (\Delta z)^2 / 12. \quad (22.14b)$$

We thus see that the gravitational potential energy decreases as the vertical stratification, $K > 0$, increases, with the maximum potential energy when the stratification vanishes, $K = 0$. As seen in Section 22.2.5, potential energy is maximized when $K = 0$ since the center of mass moves vertically upward as the stratification reduces to zero.

22.2.5 Gravitational potential energy and mixing

Consider a fluid in exact hydrostatic balance (Section 18.3) with a gravitationally stable vertical stratification where light fluid is above heavy fluid.² Now introduce a physical process, such as vertical mixing associated with a kinetic energy source, that reduces the vertical stratification. Reducing vertical stratification requires mixing to move heavy fluid up and light fluid down. In so doing, the kinetic energy supporting the mixing is converted into gravitational potential energy since the center of mass for the fluid column rises.

We can formulate this thought experiment by considering a column of seawater that is vertically stratified in salinity, S , and Conservative Temperature, Θ , and another column that is vertically unstratified with constant values S_m and Θ_m that result from vertically mixing (to infinite time) the original column.³ We assume the mass of the two columns is the same so that the bottom pressure is the same. However, the volumes will generally differ since the density differs, so that the two free surfaces, η and η_m , are different. Assuming a geopotential, $\Phi = g z$, leads to the difference between the gravitational potential energies per horizontal area in the two columns

$$g \int_{\eta_b}^{\eta_m} \rho(S_m, \Theta_m, p) z dz - g \int_{\eta_b}^{\eta} \rho(S, \Theta, p) z dz = \int_{p_a}^{p_b} (z_m - z) dp \quad (22.15a)$$

$$= (p_b - p_a) (\bar{z}_m - \bar{z}). \quad (22.15b)$$

To reach this result we used the same pressure increment, dp , when integrating over both columns, which we can do since mass is conserved. Furthermore, we used the hydrostatic balance to convert from a depth integral into a pressure integral

$$dp = -g \rho(S, \Theta, p) dz = -g \rho(S_m, \Theta_m, p) dz_m \implies z \rho(S, \Theta, p) dz = z_m \rho(S_m, \Theta_m, p) dz_m. \quad (22.16)$$

We also introduced the center of mass positions for the vertically homogeneous column, \bar{z}_m , and the stratified column, \bar{z}

$$\bar{z}_m = \frac{1}{p_b - p_a} \int_{p_a}^{p_b} z_m dp \quad \text{and} \quad \bar{z} = \frac{1}{p_b - p_a} \int_{p_a}^{p_b} z dp. \quad (22.17)$$

²See Section 25.5 for a more precise discussion of gravitational stability.

³See Section 25.3 for discussion of the seawater equation of state. For present purposes it is sufficient to know that seawater density is a function of the material tracer S , the thermodynamic tracer, Θ , and pressure, p . When a column is vertically homogenized that means S and Θ are constant throughout the column. However, pressure remains hydrostatic and thus is not vertically constant. Since density is a function of pressure, it too retains a vertical gradient.

There are two terms in the potential energy difference in equation (22.15b). The first is the mass per horizontal area, as measured by the difference in bottom pressure and applied surface pressure, $p_b - p_a$. The second is the difference between the center of mass for the two columns, $\bar{z}_m - \bar{z}$, which is a positive number since mixing moves heavier water up and lighter water down so that $\bar{z}_m > \bar{z}$. Hence, we see that the potential energy of the mixed column is larger than the stratified column.

22.3 Kinetic energy and mechanical energy

Mechanical energy is a dynamical property formed by adding the energy due to motion of fluid elements (kinetic energy) to the energy arising from the position of a fluid element within the gravitational field (gravitational potential energy). Energy is a scalar field, which often means that its budget is simpler to understand than that for the vector momentum field. In this section we develop the evolution equations for kinetic energy and then mechanical energy of a fluid on a rotating sphere, making use of the momentum equation written in the general form of equation (18.16), repeated here for convenience

$$\rho \frac{D\mathbf{v}}{Dt} + 2\rho \boldsymbol{\Omega} \wedge \mathbf{v} = -\rho \nabla \Phi + \nabla \cdot (-p \mathbb{I} + \tau) \quad (22.18a)$$

$$= -\rho \nabla \Phi - \nabla p + \rho \mathbf{F}, \quad (22.18b)$$

where we wrote

$$\mathbf{F} = \rho^{-1} \nabla \cdot \tau \quad (22.19)$$

for acceleration due to friction.

22.3.1 Kinetic energy

Following our discussion of mechanical energy of a point particle in Section 11.9, consider the kinetic energy per mass of a fluid element

$$\mathcal{K} = \mathbf{v} \cdot \mathbf{v}/2. \quad (22.20)$$

Use of the momentum equation in the form of equations (22.18a) and (22.18b) reveals that the material evolution of the kinetic energy per mass is given by

$$\rho \frac{D\mathcal{K}}{Dt} = -\rho \mathbf{v} \cdot \nabla \Phi - \mathbf{v} \cdot \nabla p + \rho \mathbf{v} \cdot \mathbf{F}. \quad (22.21)$$

We thus see that the kinetic energy of a fluid element is affected by work done by the geopotential, $-\rho \mathbf{v} \cdot \nabla \Phi$, along with work done by stresses through both pressure and friction. We detail these processes in the following.

Contribution from the geopotential

Kinetic energy increases for motion directed down the geopotential gradient

$$\mathbf{v} \cdot \nabla \Phi < 0 \implies \text{increases kinetic energy.} \quad (22.22)$$

For example, with a simple geopotential, $\Phi = g z$, kinetic energy increases where the vertical velocity is downward,

$$w < 0 \implies -w g \rho > 0 \longleftrightarrow \text{downward motion increases } \mathcal{K}. \quad (22.23)$$

As seen in Section 22.2.1, this increase in kinetic energy due to motion down the geopotential gradient is exactly balanced by a decrease in gravitational potential energy, with this kinetic to potential energy conversion also seen in Section 11.9 for the point particle.

Contribution from the pressure force

Kinetic energy increases in regions where the velocity projects down the pressure gradient,

$$\mathbf{v} \cdot \nabla p < 0 \implies \text{increase kinetic energy}, \quad (22.24)$$

thus resulting in an increase in fluid speed imparted by the pressure gradient force. Conversely, kinetic energy is reduced in regions where the flow is directed up the pressure gradient. It is notable that horizontal geostrophic flows (Section 26.4) with

$$\mathbf{v}_g = \frac{\hat{\mathbf{z}} \wedge \nabla p}{f\rho} \quad (22.25)$$

have the velocity oriented perpendicular to the horizontal pressure gradient ($\mathbf{v}_g \cdot \nabla p = 0$). Consequently, pressure has no impact on the horizontal kinetic energy of a geostrophic fluid.

Contribution from friction

Kinetic energy is reduced in regions where the velocity has a negative projection onto the direction of the friction vector, $\rho \mathbf{v} \cdot \mathbf{F} < 0$. As detailed in Section 22.3.2, the friction arising from a viscous stress tensor appropriate for a Newtonian fluid gives rise to two contributions to kinetic energy: the divergence of a viscous flux plus a sign-definite sink.

22.3.2 Frictional dissipation of kinetic energy

We here detail the role of friction on kinetic energy

$$\text{friction power per volume} = \rho \mathbf{v} \cdot \mathbf{F}, \quad (22.26)$$

where friction is determined by the stress-rate of strain relation (19.78). We interpret $\rho \mathbf{v} \cdot \mathbf{F}$ as the frictional power (energy per time) per volume acting to alter the kinetic energy per volume of a fluid element. To proceed, we expose Cartesian tensor labels to have

$$\rho \mathbf{v} \cdot \mathbf{F} = v_m \rho F_m \quad (22.27a)$$

$$= v_m \partial_n \tau_{nm} \quad (22.27b)$$

$$= 2 v_m \partial_n (\rho \nu \mathbb{S}_{mn}^{\text{dev}}) \quad (22.27c)$$

$$= 2 \partial_n (\rho \nu v_m \mathbb{S}_{mn}^{\text{dev}}) - 2 \rho \nu \partial_n v_m \mathbb{S}_{mn}^{\text{dev}} \quad (22.27d)$$

$$= 2 \nabla \cdot (\rho \nu \mathbf{v} \cdot \mathbb{S}^{\text{dev}}) - 2 \rho \nu \mathbb{S}_{mn} \mathbb{S}_{mn}^{\text{dev}}, \quad (22.27e)$$

where we recall from Section 19.7.5 that the *deviatoric rate of strain tensor* has elements given by

$$\mathbb{S}_{mn}^{\text{dev}} = \mathbb{S}_{mn} - \delta_{mn} \mathbb{S}_{qq}/3 \quad \text{with} \quad \mathbb{S}_{qq} = \nabla \cdot \mathbf{v}. \quad (22.28)$$

To reach equation (22.27e) required the identity

$$2 \partial_n v_m \mathbb{S}_{mn}^{\text{dev}} = (\partial_n v_m + \partial_m v_n) \mathbb{S}_{mn}^{\text{dev}} + (\partial_n v_m - \partial_m v_n) \mathbb{S}_{mn}^{\text{dev}} = 2 \mathbb{S}_{mn} \mathbb{S}_{mn}^{\text{dev}}, \quad (22.29)$$

where

$$(\partial_n v_m - \partial_m v_n) \mathbb{S}_{mn}^{\text{dev}} = 2 \mathbb{A}_{mn} \mathbb{S}_{mn}^{\text{dev}} = 0 \quad (22.30)$$

due to symmetry of the deviatoric rate of strain tensor, $\mathbb{S}_{mn}^{\text{dev}} = \mathbb{S}_{nm}^{\text{dev}}$, and anti-symmetry of the rotation tensor, \mathbb{A}_{mn} (see Section 14.2.5 as well as Exercise 1.2). We can show that the second term in equation (22.27e) is sign-definite by noting that

$$\mathbb{S}_{mn}^{\text{dev}} \mathbb{S}_{mn}^{\text{dev}} = (\mathbb{S}_{mn} - \delta_{mn} \mathbb{S}_{qq}/3)^2 \quad (22.31a)$$

$$= \mathbb{S}_{mn} \mathbb{S}_{mn} + \delta_{mn} \delta_{mn} \mathbb{S}_{qq} \mathbb{S}_{qq}/9 - 2 \mathbb{S}_{mn} \delta_{mn} \mathbb{S}_{qq}/3 \quad (22.31b)$$

$$= \mathbb{S}_{mn} \mathbb{S}_{mn} + (\mathbb{S}_{qq})^2/3 - 2 (\mathbb{S}_{qq})^2/3 \quad (22.31c)$$

$$= \mathbb{S}_{mn} (\mathbb{S}_{mn} - \delta_{mn} \mathbb{S}_{qq}/3) \quad (22.31d)$$

$$= \mathbb{S}_{mn} \mathbb{S}_{mn}^{\text{dev}}. \quad (22.31e)$$

We are thus left

$$\rho \mathbf{v} \cdot \mathbf{F} = 2 \nabla \cdot (\rho \nu \mathbf{v} \cdot \mathbb{S}^{\text{dev}}) - 2 \rho \nu \mathbb{S}^{\text{dev}} \cdot \mathbb{S}^{\text{dev}}, \quad (22.32)$$

with

$$\mathbb{S}^{\text{dev}} \cdot \mathbb{S}^{\text{dev}} = \mathbb{S}_{mn}^{\text{dev}} \mathbb{S}_{mn}^{\text{dev}}. \quad (22.33)$$

We interpret the two contributions to the frictional power in equation (22.32) as

$$\rho \mathbf{v} \cdot \mathbf{F} = \text{divergence of viscous flux} - \text{kinetic energy dissipation}. \quad (22.34)$$

The divergence theorem means that when integrated over the full domain, the flux divergence becomes a contribution from boundary stresses, and boundary stresses can either increase or decrease kinetic energy according to the boundary processes. In contrast, the sign-definite dissipation term provides a sink to the kinetic energy at each point in the fluid interior. This frictional dissipation is commonly written

$$\epsilon \equiv [\mathbf{v} \cdot \mathbf{F}]_{\text{dissipate}} = 2 \nu \mathbb{S}^{\text{dev}} \cdot \mathbb{S}^{\text{dev}} \geq 0. \quad (22.35)$$

The dimensions of ϵ are $\text{L}^2 \text{T}^{-3}$, which in SI units are $\text{m}^2 \text{s}^{-3} = \text{W kg}^{-1}$. We thus refer to ϵ as the kinetic energy dissipation per mass due to viscous effects.

22.3.3 Mechanical energy budget

Adding the material time evolution equations for kinetic energy per mass and gravitational potential energy per mass leads to the material form of the mechanical energy per mass

$$\rho \frac{Dm}{Dt} = -\mathbf{v} \cdot \nabla p + 2 \nabla \cdot (\rho \nu \mathbf{v} \cdot \mathbb{S}^{\text{dev}}) + \rho (-\epsilon + \partial_t \Phi) \quad (22.36a)$$

$$= -\nabla \cdot (p \mathbf{v} - 2 \rho \nu \mathbf{v} \cdot \mathbb{S}^{\text{dev}}) + p \nabla \cdot \mathbf{v} + \rho (-\epsilon + \partial_t \Phi) \quad (22.36b)$$

$$= -\nabla \cdot \mathbf{J}_{\text{mech}} + p \nabla \cdot \mathbf{v} + \rho (-\epsilon + \partial_t \Phi), \quad (22.36c)$$

where

$$m = \mathcal{K} + \Phi \quad (22.37)$$

is the mechanical energy per mass for a fluid element. As already anticipated, there is a cancellation of the mechanical energy exchange due to motion through the gravitational field, though the time-dependent geopotential provides a source of mechanical energy. The final equality introduced the mechanical energy flux

$$\mathbf{J}_{\text{mech}} = p \mathbf{v} - 2 \rho \nu \mathbf{v} \cdot \mathbb{S}^{\text{dev}} = -\mathbf{v} \cdot (-p \mathbb{I} + \tau) = -\mathbf{v} \cdot \mathbb{T}. \quad (22.38)$$

We see that the material evolution of mechanical energy arises from the convergence of the mechanical energy flux, plus the work done by pressure, effects from frictional dissipation (more discussed in Section 22.5.3), and the time changes to the geopotential.

22.3.4 Further study

The physics of mechanical energy dissipation forms a central role in the study of ocean mixing and its measurement. [MacKinnon et al. \(2013\)](#) is a useful place to begin studying this active field of physical oceanography.

As shown in Section 17.8 of [Griffies \(2004\)](#), we can relate the global integral of the kinetic energy dissipation to the friction vector by taking the functional derivative of the dissipation with respect to the velocity field. This connection follows from the self-adjoint nature of the friction operator and it has been found to be a useful framework for numerical discretization of viscous friction (e.g., [Griffies and Hallberg \(2000\)](#)).

22.4 Thermodynamics of a moving fluid

Recall the fundamental thermodynamic relation (21.33b) for a two-component fluid such as seawater or the atmosphere

$$dJ = T dS - p d\nu_s + \mu dC. \quad (22.39)$$

This relation is an expression of the first law of thermodynamics for a quasi-static process. It provides an expression for the exact differential of internal energy for a thermodynamic system that is infinitesimally close to equilibrium.

Now consider a finite region of fluid comprised of fluid elements. The fluid region generally experiences mechanical and thermal forces that support macroscopic motion. However, we assume that each fluid element is in local thermodynamic equilibrium. This assumption is supported by noting that the equilibrium time scale for the microscopic degrees of freedom within an individual fluid element is tiny compared to the equilibrium time scales of the macroscopic motion. We discussed this extremely large time scale separation in Section 21.1.3 when presenting the notion of a quasi-static process. Within this regime, we are justified in making use of quasi-equilibrium thermodynamics in which we use equilibrium thermodynamics locally yet allow for macroscopic gradients in fluid properties. That is, each fluid element is locally in thermodynamic equilibrium while it is out of equilibrium with its neighbors.

22.4.1 Concerning the transition to a moving fluid

For a moving continuous fluid, each of the thermodynamic properties in the equilibrium first law expression (22.39) are continuous functions of space and time. Furthermore, equation (22.39) provides a relation between exact differentials as detailed in Section 2.8. As exact differentials of continuous space-time fields, we can make use of the space-time increments detailed in Section 13.5.1 to write

$$d\Psi = \Psi(\mathbf{x} + d\mathbf{x}, t + dt) - \Psi(\mathbf{x}, t) = dt \partial_t \Psi + d\mathbf{x} \cdot \nabla \Psi, \quad (22.40)$$

where Ψ is one of the thermodynamic properties, dt is the infinitesimal time increment, and $d\mathbf{x}$ is the vector of infinitesimal space increments. Following the discussion in Section 13.5.2, we are thus led to the total time derivative for a property following a trajectory $\mathbf{x} = \mathbf{X}(t)$

$$\frac{d\Psi}{dt} = \frac{\partial\Psi}{\partial t} + \frac{d\mathbf{X}}{dt} \cdot \nabla \Psi. \quad (22.41)$$

Restricting the trajectory to that defined by a fluid particle, so that $\mathbf{v} = d\mathbf{X}/dt$, renders the material time derivative as in Section 13.5.4

$$\frac{D\Psi}{Dt} = \frac{\partial\Psi}{\partial t} + \mathbf{v} \cdot \nabla \Psi. \quad (22.42)$$

We make use of this result in Section 22.4.5 to transition the quasi-static relation (22.39) to a moving fluid.

22.4.2 Thermodynamic properties as space-time fields

There is room for confusion when allowing thermodynamic properties to extend to space-time functions. The confusion arises when not being clear on whether a mathematical expression refers to an equilibrium thermodynamic relation between thermodynamic variables, as in the fundamental thermodynamic relation (22.39), or whether it expresses a relation involving space-time field representations of thermodynamic properties. The distinction is important especially when considering derivatives and integrals as it is necessary to know what are the other variables to be held fixed in the process of performing the operations.

For example, consider the middle relation in equation (21.62), which says that the partial derivative of the Gibbs potential with respect to pressure, holding temperature and concentration fixed, equals to the specific volume

$$\left[\frac{\partial \mathcal{G}}{\partial p} \right]_{T,C} = \nu_s = \rho^{-1}. \quad (22.43)$$

However, if we encounter the Gibbs potential as a space-time function, and we use pressure as a generalized vertical coordinate, then we might find need to compute the distinct partial derivative

$$\left[\frac{\partial \mathcal{G}}{\partial p} \right]_{x,y,t} \neq \left[\frac{\partial \mathcal{G}}{\partial p} \right]_{T,C}. \quad (22.44)$$

When appropriate, we offer reminders to help avoid a cascade of misunderstandings. One point where this reminder is particularly useful is when discussing energetics for a Boussinesq fluid in Section 24.6.

22.4.3 Thermodynamic relations for steady state flow

Before considering time evolution, let us establish some steady (stationary) state relations defined for flows that have no Eulerian time dependence. In this case, the exact differential for a thermodynamic potential arises just from spatial variations

$$d\Psi = \Psi(\mathbf{x} + d\mathbf{x}) - \Psi(\mathbf{x}) = d\mathbf{x} \cdot \nabla \Psi \quad \Longleftarrow \text{steady state.} \quad (22.45)$$

Applying this result to the fundamental thermodynamic expression (22.39) renders the steady state relation

$$d\mathbf{x} \cdot (\nabla \mathcal{I} - T \nabla \mathcal{S} + p \nabla \nu_s - \mu \nabla C) = 0. \quad (22.46)$$

Since the spatial increment $d\mathbf{x}$ is arbitrary, the bracketed term generally vanishes

$$\nabla \mathcal{I} = T \nabla \mathcal{S} - p \nabla \nu_s + \mu \nabla C \quad \Longleftarrow \text{steady state.} \quad (22.47)$$

Analogous identities hold for steady state thermodynamic potentials following from the fundamental thermodynamic relations derived in Section 21.4

STEADY STATE FLOW

$$\nabla \mathcal{I} = T \nabla \mathcal{S} - p \nabla \nu_s + \mu \nabla C \quad (22.48)$$

$$T \nabla \mathcal{S} = \nabla \mathcal{I} + p \nabla \nu_s - \mu \nabla C \quad (22.49)$$

$$\nabla \mathcal{H} = T \nabla \mathcal{S} + \nu_s \nabla p + \mu \nabla C \quad (22.50)$$

$$\nabla \mathcal{G} = -\mathcal{S} \nabla T + \nu_s \nabla p + \mu \nabla C. \quad (22.51)$$

22.4.4 Materially constant specific entropy for a perfect fluid

Each material fluid parcel within a perfect fluid maintains a constant specific entropy given that it experiences no dissipation (friction is absent), maintains a constant composition (mixing is absent), and encounters no heating (adiabatic). Consequently, specific entropy for each fluid parcel is reversibly stirred through advection

$$\frac{D\mathcal{S}}{Dt} = \frac{\partial \mathcal{S}}{\partial t} + \mathbf{v} \cdot \nabla \mathcal{S} = 0. \quad (22.52)$$

Notably, a perfect fluid generally admits nonzero gradients of specific entropy even as each fluid element moves without altering its specific entropy.

22.4.5 First law for a moving fluid element

Section 22.4.1 provides the key operational means for developing the equations of quasi-equilibrium thermodynamics, in which we apply the equilibrium thermodynamic relations to moving and evolving fluid elements. Consequently, the equilibrium relation (22.39), which is the first law for a quasi-static process, takes the form for a moving fluid element

$$\frac{D\mathcal{J}}{Dt} = T \frac{D\mathcal{S}}{Dt} - p \frac{D\nu_s}{Dt} + \mu \frac{DC}{Dt}. \quad (22.53)$$

Exposing the Eulerian space and time derivatives renders

$$\partial_t \mathcal{J} - T \partial_t \mathcal{S} + p \partial_t \nu_s - \mu \partial_t C + \mathbf{v} \cdot (\nabla \mathcal{J} - T \nabla \mathcal{S} + p \nabla \nu_s - \mu \nabla C) = 0. \quad (22.54)$$

As discussed in Section 22.4.3, for a steady state flow all Eulerian time derivatives vanish and the spatial gradients reduce to the steady state relation (22.48). For the general case of a time dependent flow, the thermodynamic potentials are changing in both space and time so that the first law takes on the material time evolution expression (22.53).

We can massage the result (22.53) to further reveal its connection to the first law written in the form (22.39). For this purpose, recall that mass conservation as discussed in Section 15.2 means that changes in the volume of a fluid element are related to density changes via

$$\frac{1}{\delta V} \frac{D\delta V}{Dt} = \frac{1}{\nu_s} \frac{D\nu_s}{Dt}, \quad (22.55)$$

where $\nu_s = \rho^{-1}$ is the specific volume. Hence, equation (22.53) can be written

$$\delta M \frac{D\mathcal{J}}{Dt} = T \delta M \frac{D\mathcal{S}}{Dt} - p \frac{D\delta V}{Dt} + \mu \delta M \frac{DC}{Dt}, \quad (22.56)$$

where $\delta M = \rho \delta V$ is the mass of the fluid element. Since the mass of the fluid element is constant, equation (22.56) is the fluid element extension of the first law given by equation (22.39). Alternatively, we can use the further result from mass conservation (equation (15.26))

$$\frac{1}{\delta V} \frac{D\delta V}{Dt} = \nabla \cdot \mathbf{v} \quad (22.57)$$

to write

$$\frac{D\mathcal{J}}{Dt} = -p \nu_s \nabla \cdot \mathbf{v} + T \frac{D\mathcal{S}}{Dt} + \mu \frac{DC}{Dt}. \quad (22.58)$$

Processes affecting internal energy that appear on the right hand side are (i) mechanical work from pressure modifying the volume of the fluid element, (ii) entropy production, whose form is developed in Sections 22.5 and 22.7, and (iii) mixing (chemical work) through the exchange of matter constituents between fluid elements. We provide further details for the first law in Section 22.5 when considering the budget for the total energy of a fluid element.

22.4.6 Enthalpy budget

Rather than work with internal energy, it is often more convenient to consider the specific enthalpy (Section 21.4.3),

$$\mathcal{H} = \mathcal{I} + p/\rho = \mathcal{I} + p\nu_s. \quad (22.59)$$

The mass continuity equation (15.20) and the internal energy equation (22.58) yield

$$\frac{D\mathcal{H}}{Dt} = \frac{D\mathcal{I}}{Dt} + \frac{1}{\rho} \frac{Dp}{Dt} - \frac{p}{\rho^2} \frac{D\rho}{Dt} \Rightarrow \rho \frac{D\mathcal{H}}{Dt} = \frac{Dp}{Dt} + \rho \left[T \frac{Ds}{Dt} + \mu \frac{DC}{Dt} \right]. \quad (22.60)$$

The specific enthalpy equation (22.60) says that for constant pressure processes, changes to specific enthalpy of a moving fluid element arise just from those processes that give rise to changes in specific entropy and changes in matter concentration

$$\frac{Dp}{Dt} = 0 \Rightarrow \frac{D\mathcal{H}}{Dt} = T \frac{Ds}{Dt} + \mu \frac{DC}{Dt}. \quad (22.61)$$

Since many boundary processes occur approximately at near constant pressure (e.g., air-sea fluxes), this result motivates formulating boundary fluxes of matter and thermal energy in terms of enthalpy fluxes rather than internal energy fluxes. Also, the mixing of fluid elements occurs locally in space so that pressure of fluid elements is the same when they mix, again making enthalpy a useful thermodynamic potential when studying mixing.

22.4.7 Thermal and chemical processes and fluxes

The internal energy of a fluid element is modified by the diabatic transfer of thermal energy. We are generally concerned with two thermal fluxes, one due to conduction and one due to radiation

$$\mathbf{J}_{\text{therm}} = \mathbf{J}_{\text{cond}} + \mathbf{J}_{\text{rad}}. \quad (22.62)$$

Details of the radiant flux require topics outside our scope so we leave it unspecified. However, we consider forms for heat conduction in Section 22.7, with the simplest form being *Fourier's law* whereby \mathbf{J}_{cond} is directed down the temperature gradient.

The internal energy of a fluid element also changes through changes in the matter concentration. This change occurs when fluid elements mix at constant pressure. As mentioned in Section 22.4.6, enthalpy is the proper thermodynamic potential to consider for examining the mixing of matter concentrations. We thus assume that the transfer of chemical energy associated with a matter flux, \mathbf{J}_C , leads to a corresponding chemical energy flux given by

$$\mathbf{J}_{\text{chem}} = \left[\frac{\partial \mathcal{H}}{\partial C} \right]_{T,p} \mathbf{J}_C. \quad (22.63)$$

The enthalpy partial derivative is computed at constant pressure and temperature so to isolate the energy change associated just with mixing of matter. It is important to distinguish this partial derivative with the distinct derivative that leads to the chemical potential as in equation (21.52)

$$\left[\frac{\partial \mathcal{H}}{\partial C} \right]_{s,p} = \mu. \quad (22.64)$$

As shown by Exercise 21.2, the two partial derivatives are related by

$$\left[\frac{\partial \mathcal{H}}{\partial C} \right]_{T,p} = \mu - T \left[\frac{\partial \mu}{\partial T} \right]_{C,p}, \quad (22.65)$$

and we make use of this relation when studying the entropy budget in Section 22.7.

Details of the thermal and chemical fluxes are undetermined at this point. We garner further insights through studying the budget for total energy in Section 22.5 as well as the entropy budget in Section 22.7.

22.4.8 First law for homogeneous fluids in terms of potential temperature

Equation (21.111) says that the change in entropy for a fluid element moving with constant matter concentration and at the reference pressure is given in terms of the potential temperature

$$\frac{D\mathcal{S}}{Dt} = \frac{c_p}{\theta} \frac{D\theta}{Dt} \quad (22.66)$$

Since the potential temperature equals to the temperature when $p = p_R$, we have

$$c_p \frac{D\theta}{Dt} = \theta \frac{D\mathcal{S}}{Dt} = \dot{Q} \quad \text{at } p = p_R \text{ and } dC = 0. \quad (22.67)$$

In Exercise 22.4 we consider these relations for an ideal gas.

22.4.9 Further study

DeGroot and Mazur (1984) provide an authoritative accounting of quasi-equilibrium thermodynamics as applied to continuum matter such as a fluid. *Gregg* (1984) and *Davis* (1994) apply these methods to small-scale mixing in the ocean. Slightly different formulations can be found in *Landau and Lifshitz* (1987) and *Batchelor* (1967).

22.5 Budget for total energy

Recall from Section 11.9 that a point particle conserves its mechanical energy in the absence of friction. In contrast, the mechanical energy for a fluid element is not materially constant even when only conservative forces act on the element. The reason is that for the continuum fluid, there is (i) a conversion between mechanical energy and internal energy as pressure does work to alter the volume of fluid elements, and (ii) frictional dissipation of kinetic energy irreversibly converts some kinetic energy to internal energy through Joule heating (Section 22.5.3). In this section we combine the mechanical energy budget from Section 22.3 to the internal energy budget from Section 22.4, thus rendering the budget for total energy of a fluid element. We furthermore postulate that the domain integrated total energy changes only due to boundary effects, as well as possible changes in the geopotential. This assumption then leads to further specifications of the processes contributing to the internal energy and enthalpy budgets.

22.5.1 Postulating the budget for total energy

The total energy per mass of a fluid element is the sum of the internal energy plus the mechanical energy,

$$\mathcal{E} = \mathcal{I} + \mathcal{M} = \mathcal{I} + \mathcal{K} + \Phi. \quad (22.68)$$

We postulate that this total energy satisfies a conservation law whereby it is affected only by the convergence of a total energy flux, plus a source due to temporal changes in the geopotential

$$\rho \frac{D\mathcal{E}}{Dt} = -\nabla \cdot \mathbf{J}_{\mathcal{E}} + \rho \partial_t \Phi. \quad (22.69)$$

Correspondingly, the Eulerian flux form equation is

$$\frac{\partial(\rho \mathcal{E})}{\partial t} + \nabla \cdot (\rho \mathbf{v} \mathcal{E} + \mathbf{J}_{\mathcal{E}}) = \rho \partial_t \Phi. \quad (22.70)$$

The flux of total energy is given by

$$\rho \mathbf{v} \mathcal{E} + \mathbf{J}_{\mathcal{E}} = \rho \mathbf{v} \mathcal{E} + \mathbf{J}_{\text{therm}} + \mathbf{J}_{\text{chem}} + \mathbf{J}_{\text{mech}} \quad (22.71a)$$

$$= \rho \mathbf{v} \mathcal{E} + \mathbf{J}_{\text{therm}} + \mathbf{J}_{\text{chem}} - \mathbf{v} \cdot (-p \mathbb{I} + \tau) \quad (22.71b)$$

$$= \rho \mathbf{v} (\mathcal{E} + p/\rho) + \mathbf{J}_{\text{therm}} + \mathbf{J}_{\text{chem}} - \mathbf{v} \cdot \tau, \quad (22.71c)$$

where $\mathbf{J}_{\text{therm}}$ and \mathbf{J}_{chem} were discussed in Section 22.4.7, and \mathbf{J}_{mech} was derived in Section 22.3.3. We have more to say concerning the flux of total energy in Section 22.6 when studying the Bernoulli potential.

We are led to postulate the total energy equation (22.69) through assuming that $\int_{\mathcal{R}} \rho \mathcal{E} dV$ remains constant in time for a region, \mathcal{R} , that is closed to thermal, material, and mechanical interactions, and one where the geopotential is constant in time. This assumption is based on our understanding of molecular and atomic mechanics.

22.5.2 Budgets for internal energy and enthalpy

We now have the total energy budget as postulated in the form of equation (22.69), along with the mechanical energy budget derived in equation (22.36c). Subtracting the two yields the internal energy budget

$$\rho \frac{D(\mathcal{E} - m)}{Dt} = \rho \frac{D\mathcal{J}}{Dt} = -\nabla \cdot (\mathbf{J}_{\text{therm}} + \mathbf{J}_{\text{chem}}) - p \nabla \cdot \mathbf{v} + \rho \epsilon. \quad (22.72)$$

This equation provides yet another expression for the first law of thermodynamics for a moving fluid element. It says that the internal energy of a fluid element is modified through the convergence of thermal and chemical fluxes, pressure work that alters the volume of a fluid element, and frictional dissipation through viscosity. Both the pressure work and frictional dissipation are exchanged with mechanical energy, and so they appear with opposite signs in the two budget equations. We further specify the thermal and chemical fluxes in Section 22.7 when studying the entropy budget.

Finally, we make use of the enthalpy equation (22.60) to render the enthalpy budget

$$\rho \frac{D\mathcal{H}}{Dt} - \frac{Dp}{Dt} = -\nabla \cdot (\mathbf{J}_{\text{therm}} + \mathbf{J}_{\text{chem}}) + \rho \epsilon. \quad (22.73)$$

We return to this equation in Section 22.8 when studying the evolution of temperature, where we introduce the potential enthalpy and Conservative Temperature.

22.5.3 Joule heating from friction

Frictional dissipation, $\epsilon > 0$, measures the conversion of kinetic energy into heat, and it is thus a conversion from mechanical energy to internal energy

$$\dot{Q}_{\text{Joule}} \equiv \epsilon. \quad (22.74)$$

This term is referred to as *Joule heating* in analog to the process that occurs in electrical circuits. The Joule heating of a fluid by molecular viscosity is larger in regions where the fluid strains are larger, signalling a more efficient transfer of power to the microscales where molecular viscosity can act on the flow.

In the ocean interior, measurements indicate that $\epsilon \approx 10^{-9} \text{ W kg}^{-1}$. Dividing by $c_p = 3900 \text{ J kg}^{-1} \text{ K}^{-1}$ leads to a heating rate of less than $10^{-3} \text{ K century}^{-1}$, which is a very small rate of ocean heating. Consequently, ocean Joule heating has a negligible role in the ocean heat budget and as such is generally ignored. Atmospheric flows are roughly two orders faster so that the kinetic energy per mass is four orders larger. The larger flow speeds lead to larger shears thus creating larger viscous dissipation that reaches roughly $\epsilon \approx 2 \text{ W m}^{-2}$ globally averaged. Hence, Joule heating is an important part of the global atmosphere enthalpy budget (Becker, 2003).

22.5.4 Comments on gauge symmetry

Consider again the Eulerian flux equation for total energy (22.70). It is notable that the time tendency for the total energy remains unchanged if we shift the flux of total energy by a curl,

$$\rho \mathcal{E} \mathbf{v} + \mathbf{J}_{\mathcal{E}} \rightarrow \rho \mathcal{E} \mathbf{v} + \mathbf{J}_{\mathcal{E}} + \nabla \wedge \mathbf{G}, \quad (22.75)$$

with \mathbf{G} referred to as a *gauge function*. This arbitrariness in the definition of total energy flux is ubiquitous in physics; e.g., see the discussion of the electromagnetic field energy flux in Section 27.4 in Volume II of Feynman et al. (1963).⁴ It means that the energy flux itself has no unique local physical meaning; only its convergence has a meaning as it gives rise to time changes in the energy at a point in space.

We also encounter such *gauge symmetry* in the potential vorticity flux discussed in Chapter 39, as well as the vector streamfunction for an incompressible fluid in Section 17.5.1. In some cases we can exploit the symmetry to our subjective desires, such as discussed in Section 39.4.6 for potential vorticity. However, we know of no strategic use of gauge symmetry for the study of energy budgets.

22.5.5 Further study

The postulate of globally integrated total energy conservation in Section 22.5.1, and the associated discussion of energy budgets, follow that from Chapter 14 of Callen (1985), Sections 49 and 58 of Landau and Lifshitz (1987), Section II.4 of DeGroot and Mazur (1984), Chapters 3 and 4 of Müller (2006), Appendix B of IOC et al. (2010), Section 2.4 of Olbers et al. (2012), Chapter 1 of Vallis (2017), and Section 13.5.5 of Thorne and Blandford (2017).

22.6 Bernoulli potential

The total energy equation (22.69) reveals that the material time change for the total energy of a fluid element is affected by the convergence of pressure times velocity. Hence, even in the absence of irreversible processes and with a time-independent geopotential, the total energy of a fluid element is not materially invariant. The energy source term $p \mathbf{v}$ is fundamental to energy within the continuum. As shown in this section, it is the pressure work required for the fluid element to mechanically exist within the continuum. We refer to it as the mechanical *injection work*.

22.6.1 Bernoulli potential and mechanical injection work

Consider the Eulerian flux form equation for the total energy, (22.70), written here in the form

$$\frac{\partial(\rho \mathcal{E})}{\partial t} + \nabla \cdot [\rho \mathbf{v} (\mathcal{E} + p/\rho) + \mathbf{J}_{\text{therm}} + \mathbf{J}_{\text{chem}} - \mathbf{v} \cdot \boldsymbol{\tau}] = \rho \partial_t \Phi. \quad (22.76)$$

⁴The Feynman lectures are available online through the California Institute of Technology.

The left hand side indicates that total energy of a fluid element is locally modified by the advective transport of the quantity

$$\mathcal{E} + p/\rho = (\mathcal{K} + \Phi) + (\mathcal{J} + p/\rho) = m + \mathcal{H} \equiv \mathcal{B}, \quad (22.77)$$

where we introduced the *Bernoulli potential*, which is the sum of the mechanical energy per mass plus the enthalpy per mass

$$\mathcal{B} = \mathcal{M} + \mathcal{H} = \mathcal{K} + \Phi + \mathcal{J} + p/\rho. \quad (22.78)$$

For a perfect fluid there is no irreversible transfer of heat, matter, or momentum so that

$$\mathbf{J}_{\text{therm}} = 0 \quad \text{and} \quad \mathbf{J}_{\text{chem}} = 0 \quad \text{and} \quad \tau = 0. \quad (22.79)$$

Hence, we see that integration over a region with zero boundary transfer of $\mathbf{v}\mathcal{B}$ leads to the conservation of total energy for a perfect fluid with a time independent geopotential.

Why is $\rho\mathcal{E}$ affected by the convergence of $\rho\mathbf{v}\mathcal{B}$ rather than the convergence of $\rho\mathbf{v}\mathcal{E}$? To answer this question,⁵ again note that the Bernoulli potential is the sum of the total energy per mass of a fluid element, \mathcal{E} , plus the term $p/\rho = p\nu_s$. So what is $p\nu_s$? Imagine carving out a tiny region from within a continuous fluid with pressure p and specific volume ν_s , leaving behind a “hole”. The mechanical work required to carve out this hole is precisely equal to $p\nu_s$. Correspondingly, we interpret $p\nu_s$ as the mechanical work required to inject a unit mass of fluid with specific volume ν_s into a region with pressure p . We thus refer to $p\nu_s$ as the *injection work*, and we in turn see that specific enthalpy, $\mathcal{H} = \mathcal{J} + p/\rho$ (equation (22.59)), measures the internal energy plus the mechanical work required for a fluid element to exist within a continuum.

We can support the above interpretation by considering the flux, $\rho\mathbf{v}\mathcal{B}$, in a perfect fluid that penetrates a static closed fluid region

$$\oint_{\partial\mathcal{R}} \rho\mathbf{v}\mathcal{B} \cdot \hat{\mathbf{n}} dS = \oint_{\partial\mathcal{R}} \rho\mathbf{v}\mathcal{E} \cdot \hat{\mathbf{n}} dS + \oint_{\partial\mathcal{R}} p\mathbf{v} \cdot \hat{\mathbf{n}} dS. \quad (22.80)$$

The first term is the flux of total energy (mechanical plus internal) that penetrates the boundary, $\partial\mathcal{R}$. The second term is the mechanical work done by pressure acting on the surface. The pipe flow example in Section 22.6.2 further supports this perspective.

22.6.2 Bernoulli's theorem for a steady perfect fluid

Consider a perfect fluid flow in steady state (vanishing Eulerian time derivatives). Steady state mass conservation means that

$$\frac{\partial \rho}{\partial t} = -\nabla \cdot (\rho\mathbf{v}) = 0. \quad (22.81)$$

This relation, along with a steady state energy in equation (22.76) (absent friction, heating, mixing, and with a time-independent geopotential), means that the steady state velocity field is parallel to contours of constant Bernoulli potential

$$\mathbf{v} \cdot \nabla \mathcal{B} = 0. \quad (22.82)$$

We thus see that for the perfect fluid to be in a steady state, the Bernoulli potential is constant along streamlines, which is a result known as *Bernoulli's theorem*. Hence, as the fluid moves along

⁵This argument follows Section 13.5.4 of [Thorne and Blandford \(2017\)](#) as well as Section 6 of [Landau and Lifshitz \(1987\)](#).

a streamline, there is an exchange between the total energy per mass, \mathcal{E} , and the injection work, $p\nu_s$, such that their sum remains constant.

A constant Bernoulli potential for steady flow is used frequently in engineering fluid dynamics to interpret flow around objects, such as for flow around a wing, in which case the sum $p + \rho v^2/2$ is sometimes referred to as the *total pressure* or *stagnation pressure*. It leads to a realization of *Bernoulli's Principle*, whereby in regions of low pressure the energy per mass is relatively large, whereas the converse holds in regions of high pressure. The change in energy is largely due to a change in the kinetic energy, so that flow is fast in regions of low pressure (e.g., top of the wing, flow around a train moving through a tunnel) and slow in regions of high pressure (e.g., bottom of the wing). We revisit Bernoulli's Principle in Section 30.1.2 when studying inviscid and irrotational flow.

22.6.3 Traditional derivation of Bernoulli's theorem

For completeness we offer a second derivation of Bernoulli's theorem that follows a more traditional route and reveals some useful manipulations. For this purpose, convert the advective-form momentum equation (18.16) into its vector-invariant form by making use of the vector identity (see Section 2.3.4)

$$\boldsymbol{\omega} \wedge \mathbf{v} = -\mathcal{K} + (\mathbf{v} \cdot \nabla) \mathbf{v}. \quad (22.83)$$

This identity allows us to eliminate velocity self-advection in favor of the vorticity and kinetic energy per mass

$$\frac{\partial \mathbf{v}}{\partial t} + \boldsymbol{\omega}_a \wedge \mathbf{v} = -\frac{1}{\rho} \nabla p - \nabla m, \quad (22.84)$$

where

$$\boldsymbol{\omega}_a = \boldsymbol{\omega} + 2\boldsymbol{\Omega} \quad (22.85)$$

is the absolute vorticity (see Chapter 36) and we set the irreversible terms to zero since we are assuming a perfect fluid. The Eulerian time evolution for the kinetic energy per mass is therefore given by

$$\frac{\partial \mathcal{K}}{\partial t} = -\frac{1}{\rho} \mathbf{v} \cdot \nabla p - \mathbf{v} \cdot \nabla m, \quad (22.86)$$

where we set $\mathbf{v} \cdot (\boldsymbol{\omega}_a \wedge \mathbf{v}) = 0$.

For Bernoulli's theorem we are interested in the steady state, with a steady kinetic energy per mass realized by the balance

$$\rho^{-1} \mathbf{v} \cdot \nabla p = -\mathbf{v} \cdot \nabla m. \quad (22.87)$$

We can connect this steady state balance to the Bernoulli potential by noting that for a steady perfect fluid, equation (22.50) relates the pressure gradient force to

$$-\rho^{-1} \nabla p = -\nabla \mathcal{H} + T \nabla S. \quad (22.88)$$

Combining with equation (22.87) renders

$$\mathbf{v} \cdot (\nabla \mathcal{H} + \nabla m - T \nabla S) = \mathbf{v} \cdot (\nabla \mathcal{B} - T \nabla S) = 0. \quad (22.89)$$

A perfect fluid maintains materially constant specific entropy (Section 22.4.4), which in a steady state means that

$$\mathbf{v} \cdot \nabla S = 0 \quad \text{and} \quad \mathbf{v} \cdot \nabla \mathcal{B} = 0 \quad \Leftarrow \text{steady state perfect fluid.} \quad (22.90)$$

That is, for a steady perfect fluid the velocity is aligned with isosurfaces of specific entropy and Bernoulli potential.

22.6.4 Steady flow of constant density perfect fluid in a pipe

To help further understand Bernoulli's theorem and the contribution from the mechanical work provided by pressure forces, consider the steady flow of a constant density perfect fluid in a frictionless pipe as depicted in Figure 22.1. For this system, Bernoulli's theorem says that the following quantity is constant for flow along a streamline

$$v^2/2 + p/\rho + g z = \text{constant}. \quad (22.91)$$

Note that internal energy dropped out since for a constant density fluid the internal energy is a constant and so plays no role in the energetics. We now show that the statement (22.91) of Bernoulli's theorem can be derived through traditional energetic arguments, whereby the work done on the fluid system equals to the system's change in kinetic energy (see Section 10.1.5 for the particle mechanics version of this *work-energy theorem*).

For this purpose, let the system under examination be a control volume of fluid as described in the caption to Figure 22.1, and examine the work done on the control volume over an arbitrary time increment, Δt . During this time, a mass of fluid given by

$$M = \rho A_1 u_1 \Delta t = \rho A_2 u_2 \Delta t \quad (22.92)$$

moves through the pipe, with $A_1 u_1 = A_2 u_2$ following from volume conservation, and we assumed that the $u_{1,2} = \Delta x_{1,2}/\Delta t$ measures the average velocity across the pipe cross-section. Mechanical work is applied to the fluid in the control volume by pressure acting on the end caps (contact force) and by gravity acting throughout the fluid (body force).

- **PRESSURE WORK:** At the left end cap, pressure from fluid to the left of the control volume does work on the control volume by the amount $p_1 A_1 \Delta x_1 = p_1 M/\rho$. On the right end, the control volume does work on the fluid to its right, which means that a negative work is applied to the control volume in the amount $-p_2 A_2 \Delta x_2 = -p_2 M/\rho$.
- **GRAVITATIONAL WORK:** Fluid downstream at the right end is higher than fluid upstream on the left end. The control volume must do work against gravity to achieve this altitude increase and this work is given by $-g M (z_2 - z_1)$.

As the fluid moves from left to right, the control volume changes its kinetic energy by the amount $(M/2)(u_2^2 - u_1^2)$. Equating this kinetic energy change to the work applied to the control volume renders

$$(1/2)(u_2^2 - u_1^2) = (1/\rho)(p_1 - p_2) - g(z_2 - z_1), \quad (22.93)$$

where the mass, M , dropped out. Rearrangement then leads to

$$u^2/2 + p/\rho + g z = \text{constant}, \quad (22.94)$$

which is a statement of Bernoulli's theorem (22.91). Furthermore, the pressure difference between the left and right end of the pipe is given by

$$p_1 - p_2 = \rho g (z_2 - z_1) + (\rho/2) u_1^2 [1 - (A_1/A_2)^2] > 0. \quad (22.95)$$

We expect $p_1 > p_2$ given the assumed fluid motion from left to right, whereby pressure on the left end must be larger than the right end to support this flow. This example thus supports our understanding of how pressure provides a mechanical work on fluid control volume boundaries. Indeed, taking the control volume to be a tiny fluid element furthers our interpretation of the p/ρ contribution to the Bernoulli potential (22.77).

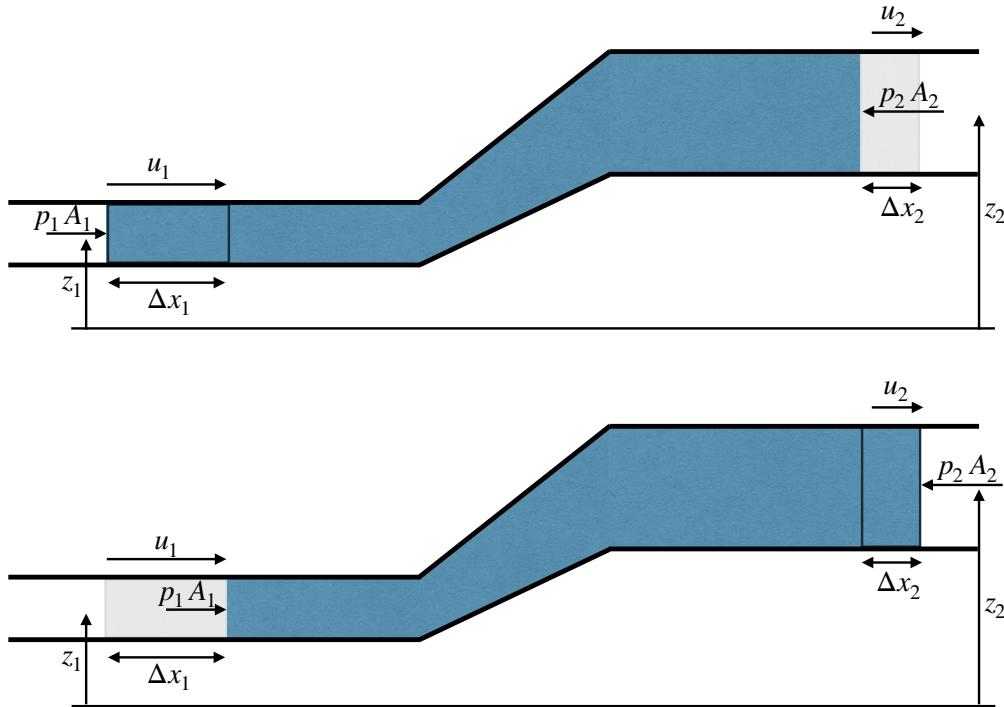


FIGURE 22.1: An example to illustrate the basic physics of Bernoulli’s theorem and pressure work, whereby we depict the flow of a perfect and constant density fluid from left to right in a pipe of variable cross-section and variable height. We study the energetics of a control volume (dark blue region) moving with the fluid. The top panel shows the control volume at one time and the lower panel shows the control volume at a time Δt later, after which a mass of fluid, M , has moved through the system. Volume conservation means that $(u_1 \Delta t) A_1 = (u_2 \Delta t) A_2$, where $\Delta x_{1,2}$ is the horizontal displacement of the fluid plug over time Δt , $u_{1,2}$ is the cross-sectional area average velocity, $A_{1,2}$ is the pipe cross-sectional area, and $M = \rho \Delta x_1 A_1 = \rho \Delta x_2 A_2$ is the mass of fluid moving over the Δt time increment. Pressure forces, $p_{1,2}$, at the end caps point inward (compressive), with pressure on the left larger than that on the right to support the fluid moving to the right. As the fluid moves upward it increases its gravitational potential energy and in so doing the fluid does work against gravity.

22.6.5 Further study

For an examination of Bernoulli’s theorem for non-rotating flows, such as flow in laminar boundary layers, see [this video](#) from Prof. Shapiro.

22.7 Entropy budget for the ocean

In this section we consider the entropy budget for the ocean and make use of the second law of thermodynamics to infer some specific forms for the thermal and chemical fluxes that were introduced in Section 22.4.7. To start, we note that the entropy budget follows simply by rearranging the enthalpy equation (22.60)

$$\rho T \frac{D\mathcal{S}}{Dt} = \rho \frac{D\mathcal{H}}{Dt} - \frac{Dp}{Dt} - \rho \mu \frac{DC}{Dt}, \quad (22.96)$$

with the enthalpy budget in the form (22.73) yielding

$$T \rho \frac{D\mathcal{S}}{Dt} = -\nabla \cdot (\mathbf{J}_{\text{therm}} + \mathbf{J}_{\text{chem}}) + \rho \epsilon + \mu \nabla \cdot \mathbf{J}_C, \quad (22.97)$$

where we wrote the tracer equation as

$$\rho \frac{DC}{Dt} = -\nabla \cdot \mathbf{J}_C \quad (22.98)$$

as in Section 16.1.3. Through some manipulations pursued in Section 22.7.1, we identify a non-advection entropy flux and a corresponding entropy source. This form of the entropy budget is useful since we can then invoke the second law to insist that the entropy source is non-negative, and that statement renders some constraints on the thermal and chemical fluxes.

22.7.1 Non-advection entropy flux and entropy source

To start the manipulations, divide equation (22.97) by T and make use of the chemical energy flux in the form

$$\mathbf{J}_{\text{chem}} = \left[\frac{\partial \mathcal{H}}{\partial C} \right]_{T,p} \mathbf{J}_C = \left(\mu - T \left[\frac{\partial \mu}{\partial T} \right]_{C,p} \right) \mathbf{J}_C = (\mu - T \mu_T) \mathbf{J}_C \quad (22.99)$$

where we used the identity (22.65) for the second equality, and introduced a shorthand for the third equality. In this manner we have

$$\rho \frac{DS}{Dt} = -\frac{\nabla \cdot (\mathbf{J}_{\text{therm}} + (\mu - T \mu_T) \mathbf{J}_C)}{T} + \frac{\rho \epsilon}{T} + \frac{\mu \nabla \cdot \mathbf{J}_C}{T} \quad (22.100a)$$

$$= -\nabla \cdot (\mathbf{J}_{\text{therm}}/T - \mu_T \mathbf{J}_C) + \frac{\rho \epsilon}{T} + \frac{\nabla T}{T} \cdot (\mu_T \mathbf{J}_C - \mathbf{J}_{\text{therm}}/T) - (\mathbf{J}_C \cdot \nabla \mu)/T \quad (22.100b)$$

$$= -\nabla \cdot (\mathbf{J}_{\text{therm}}/T - \mu_T \mathbf{J}_C) + \frac{\rho \epsilon}{T} - \frac{\nabla T \cdot \mathbf{J}_{\text{therm}}}{T^2} + \frac{\mathbf{J}_C \cdot (\mu_T \nabla T - \nabla \mu)}{T} \quad (22.100c)$$

$$= -\nabla \cdot (\mathbf{J}_{\text{therm}}/T - \mu_T \mathbf{J}_C) + \frac{\rho \epsilon}{T} - \frac{\nabla T \cdot \mathbf{J}_{\text{therm}}}{T^2} - \frac{\mathbf{J}_C \cdot (\mu_C \nabla C + \mu_p \nabla p)}{T}, \quad (22.100d)$$

where we wrote the gradient of the chemical potential, $\mu(T, p, C)$, as

$$\nabla \mu = \left[\frac{\partial \mu}{\partial T} \right]_{p,C} \nabla T + \left[\frac{\partial \mu}{\partial p} \right]_{T,C} \nabla p + \left[\frac{\partial \mu}{\partial C} \right]_{T,p} \nabla C = \mu_T \nabla T + \mu_p \nabla p + \mu_C \nabla C. \quad (22.101)$$

We can thus identify the non-advection entropy flux

$$\mathbf{J}_S = \frac{\mathbf{J}_{\text{therm}}}{T} - \left[\frac{\partial \mu}{\partial T} \right]_{p,C} \mathbf{J}_C, \quad (22.102)$$

and the entropy source

$$\Sigma_S = \frac{\rho \epsilon}{T} - \frac{\nabla T \cdot \mathbf{J}_{\text{therm}}}{T^2} - \frac{\mathbf{J}_C}{T} \cdot \left(\left[\frac{\partial \mu}{\partial C} \right]_{T,p} \nabla C + \left[\frac{\partial \mu}{\partial p} \right]_{T,C} \nabla p \right). \quad (22.103)$$

The non-advection entropy flux (22.102) arises from the thermal and chemical fluxes, whereas the entropy source includes contributions from those fluxes as well as the frictional dissipation of mechanical energy.

22.7.2 Constraints from the second law of thermodynamics

The second law of thermodynamics states that the entropy source is non-negative

$$\Sigma_S \geq 0 \iff \text{second law of thermodynamics.} \quad (22.104)$$

This condition imposes constraints on the frictional dissipation, thermal flux, and tracer flux. Since frictional dissipation in Newtonian fluids (equation (22.35)) is non-negative, $\epsilon \geq 0$, we make use of the second law to constrain just the thermal flux and tracer flux. Furthermore, we split off the radiant flux from the thermal flux, since the radiant flux is generally determined by processes external to the fluid. We thus use the second law to constrain the conductive portion of the thermal flux along with the tracer flux so that

$$-\nabla T \cdot \mathbf{J}_{\text{cond}} - T \mu_C \mathbf{J}_C \cdot \left[\nabla C + \frac{\mu_p}{\mu_C} \nabla p \right] \geq 0. \quad (22.105)$$

This constraint can be satisfied by assuming the conductive and tracer fluxes are of the form

$$\rho^{-1} \mathbf{J}_{\text{cond}} = -c_p \kappa_T \nabla T - \kappa_{TC} \left[\nabla C + \frac{\mu_p}{\mu_C} \nabla p \right] \quad (22.106a)$$

$$\rho^{-1} \mathbf{J}_C = -\kappa_C \left[\nabla C + \frac{\mu_p}{\mu_C} \nabla p \right] - \kappa_{CT} \nabla T. \quad (22.106b)$$

The first term in the conductive thermal flux (22.106a) is known as *Fourier's law of conduction*

$$\mathbf{J}_{\text{cond}}^{\text{Fourier}} = -\rho c_p \kappa_T \nabla T, \quad (22.107)$$

in which case the conductive thermal flux is directed down the gradient of the *in situ* temperature. The second term leads to a conductive thermal flux in the presence of matter concentration gradients and thermal gradients, and this process is known as the *Dufour effect*

$$\mathbf{J}_{\text{cond}}^{\text{Dufour}} = -\rho \kappa_{TC} \left[\nabla C + \frac{\mu_p}{\mu_C} \nabla p \right]. \quad (22.108)$$

The first term in the matter flux (22.106b) is known as *Fick's law of diffusion*

$$\mathbf{J}_C^{\text{Fick}} = -\rho \kappa_C \nabla C. \quad (22.109)$$

We generally also encounter a matter diffusive flux in the presence of pressure gradients, which accounts for the second contribution to the matter flux in equation (22.106b). Finally, the matter flux has a contribution from temperature gradients, which is known as the *Soret effect*

$$\mathbf{J}_C^{\text{Soret}} = -\rho \kappa_{CT} \nabla T. \quad (22.110)$$

The Onsager reciprocity conditions⁶ are now invoked to relate the two off-diagonal coefficients according to

$$T \mu_C \kappa_{CT} = \kappa_{TC}, \quad (22.111)$$

which brings the entropy condition to the form

$$c_p \kappa_T |\nabla T|^2 + \kappa_C \mu_C T \left| \nabla C + \frac{\mu_p}{\mu_C} \nabla p \right|^2 + 2 \kappa_{TC} \nabla T \cdot \left[\nabla C + \frac{\mu_p}{\mu_C} \nabla p \right] \geq 0. \quad (22.112)$$

This condition then constrains the phenomenological constants κ_C , κ_T and κ_{TC} so that

$$|\kappa_{TC}|^2 \leq \kappa_T \kappa_C c_p T \mu_S. \quad (22.113)$$

⁶See Chapter 14 of [Callen \(1985\)](#) for these conditions and their underlying dynamical connections.

22.7.3 A summary presentation

We here summarize the previous material by skipping details for the entropy flux, thermal flux, and matter flux. For this purpose, write the budget for the total energy in the form

$$\rho \frac{D\mathcal{E}}{Dt} = -\nabla \cdot (p\mathbf{v} - \mathbf{v} \cdot \tau) + \rho \left[-\epsilon + T \frac{DS}{Dt} + \mu \frac{DC}{Dt} \right] + \rho \partial_t \Phi. \quad (22.114)$$

Now assume that the specific entropy and matter concentration satisfy the evolution equations

$$\rho \frac{DS}{Dt} = -\nabla \cdot \mathbf{J}_S + \Sigma_S \quad \text{and} \quad \rho \frac{DC}{Dt} = -\nabla \cdot \mathbf{J}_C, \quad (22.115)$$

thus rendering

$$\rho \left[T \frac{DS}{Dt} + \mu \frac{DC}{Dt} \right] = -T \nabla \cdot \mathbf{J}_S + T \Sigma_S - \mu \nabla \cdot \mathbf{J}_C \quad (22.116a)$$

$$= -\nabla \cdot (T \mathbf{J}_S + \mu \mathbf{J}_C) + \nabla T \cdot \mathbf{J}_S + \nabla \mu \cdot \mathbf{J}_C + T \Sigma_S, \quad (22.116b)$$

which then brings the total energy equation (22.114) into the form

$$\rho \frac{D\mathcal{E}}{Dt} = -\nabla \cdot (p\mathbf{v} - \mathbf{v} \cdot \tau + T \mathbf{J}_S + \mu \mathbf{J}_C) + [-\rho \epsilon + \nabla T \cdot \mathbf{J}_S + \nabla \mu \cdot \mathbf{J}_C + T \Sigma_S] + \rho \partial_t \Phi. \quad (22.117)$$

We now postulate that the globally integrated total energy is constant in the absence of boundary processes and with a time independent geopotential. In the presence of mechanical dissipation and matter constituent mixing, a necessary condition for such global energy conservation is for the specific entropy source to take the form

$$T \Sigma_S = \rho \epsilon - \nabla T \cdot \mathbf{J}_S - \nabla \mu \cdot \mathbf{J}_C. \quad (22.118)$$

That is, the entropy source arises from frictional dissipation, entropy mixing, and matter mixing. With this form for the entropy source, the total energy budget (22.117) is given by the material form and the equivalent Eulerian flux-form expressions

$$\rho \frac{D\mathcal{E}}{Dt} = -\nabla \cdot (p\mathbf{v} - \mathbf{v} \cdot \tau + T \mathbf{J}_S + \mu \mathbf{J}_C) + \rho \partial_t \Phi \quad (22.119a)$$

$$\frac{\partial(\rho \mathcal{E})}{\partial t} = -\nabla \cdot (\mathcal{E} \mathbf{v} + p \mathbf{v} - \mathbf{v} \cdot \tau + T \mathbf{J}_S + \mu \mathbf{J}_C) + \rho \partial_t \Phi. \quad (22.119b)$$

The modified form of the internal energy budget (22.58) is found by subtracting the mechanical energy budget (22.36c) from the total energy budget (22.119a)

$$\rho \frac{DJ}{Dt} = \rho \frac{D(\mathcal{E} - \mathcal{M})}{Dt} = -p \nabla \cdot \mathbf{v} - \nabla \cdot (T \mathbf{J}_S + \mu \mathbf{J}_C) + \rho \epsilon. \quad (22.120)$$

The corresponding enthalpy budget (22.60) is given by

$$\rho \frac{D\mathcal{H}}{Dt} = \frac{Dp}{Dt} - \nabla \cdot (T \mathbf{J}_S + \mu \mathbf{J}_C) + \rho \epsilon. \quad (22.121)$$

22.7.4 Comments and further study

It is remarkable how the second law of thermodynamics is able to predict new physical processes through considering the various forms that the thermal and matter fluxes can take to ensure a positive entropy source. Even so, for most purposes of oceanography, the fluxes that are considered in this section are far smaller than those induced by turbulent flow processes, in which case the most accurate flux relations reduce to the Fickian and Fourier expressions, thus making the off-diagonal processes (Dufour, Soret, pressure induced diffusion) less important in practice.

Much of our presentation in this section followed that from Sections 2.4 and 2.5 from [Olbers et al. \(2012\)](#). Similar presentations are also found in Appendix B of [IOC et al. \(2010\)](#). The physical ideas underlying the Onsager reciprocity relations are lucidly discussed in Chapter 14 of [Callen \(1985\)](#).

22.8 Temperature evolution

In specifying the state of a fluid element it is sensible to make use of the temperature, pressure, and tracer concentration given that these state properties are readily measured in the laboratory and environment. Furthermore, these properties are the natural variables for the Gibbs potential (Section 21.4.5). Hence, given values for (T, p, C) we can determine the Gibbs potential and then determine all other thermodynamic properties by taking partial derivatives.

How do we specify the evolution of (T, p, C) for a fluid element? Evolution of the matter concentration follows from the tracer equation (an advection-diffusion equation) as developed in Section 16.1. Pressure measures the compressive stress acting on each fluid element (Section 19.7), with its specification depending on the dominant dynamical balances (see Section 22.10). Temperature reflects the energy of the internal microscopic degrees of freedom within a fluid element, with its evolution the subject of this section. We show how of Conservative Temperature, Θ , rather than *in situ* temperature, T , or potential temperature, θ , offers the simplest prognostic equation of the three temperature variables. The key reason is that Θ evolves almost precisely like a material tracer, driven by the convergence of fluxes, whereas the equations for T and θ contain extra source terms in addition to flux convergences.

22.8.1 Enthalpy equation again

In developing the temperature equation it is useful to start from the prognostic equation for enthalpy as developed in Sections 22.4.6. For that purpose we write the enthalpy equation (22.121) as

$$\rho \frac{D\mathcal{H}}{Dt} = \frac{Dp}{Dt} - \nabla \cdot \mathbf{J}_{\mathcal{H}} + \rho \epsilon, \quad (22.122)$$

with the enthalpy flux written in terms of the entropy and tracer fluxes

$$\mathbf{J}_{\mathcal{H}} = T \mathbf{J}_S + \mu \mathbf{J}_C. \quad (22.123)$$

22.8.2 Evolution of *in situ* temperature

To reveal a prognostic equation for temperature, we write enthalpy as a function of (T, p, C) so that

$$\frac{D\mathcal{H}}{Dt} = \left[\frac{\partial \mathcal{H}}{\partial T} \right]_{p,C} \frac{DT}{Dt} + \left[\frac{\partial \mathcal{H}}{\partial p} \right]_{T,C} \frac{Dp}{Dt} + \left[\frac{\partial \mathcal{H}}{\partial C} \right]_{T,p} \frac{DC}{Dt}. \quad (22.124)$$

The partial derivatives can be related to response functions via the following. First, the specific heat capacity at constant pressure is given by equation (21.77)

$$\left[\frac{\partial \mathcal{H}}{\partial T} \right]_{p,C} = c_p. \quad (22.125)$$

Next, we make use of the Gibbs potential identities in Section 21.4.5 to write

$$\left[\frac{\partial \mathcal{H}}{\partial p} \right]_{T,C} = \left[\frac{\partial \mathcal{G}}{\partial p} \right]_{T,C} - T \left[\frac{\partial}{\partial p} \right]_{T,C} \left[\frac{\partial \mathcal{G}}{\partial T} \right]_{p,C} \quad (22.126a)$$

$$= \nu_s - T \left[\frac{\partial}{\partial T} \right]_{p,C} \left[\frac{\partial \mathcal{G}}{\partial p} \right]_{T,C} \quad (22.126b)$$

$$= \nu_s - T \left[\frac{\partial \nu_s}{\partial T} \right]_{p,C} \quad (22.126c)$$

$$= \nu_s (1 - T \alpha_T), \quad (22.126d)$$

where α_T is the thermal expansion coefficient given by equation (21.78). Use of these identities in the enthalpy equation (22.124) and rearrangement leads to the *in situ* temperature equation

$$c_p \rho \frac{DT}{Dt} = -\nabla \cdot \mathbf{J}_{\mathcal{H}} + \left[\frac{\partial \mathcal{H}}{\partial C} \right]_{T,p} \nabla \cdot \mathbf{J}_C + \alpha_T T \frac{Dp}{Dt} + \rho \epsilon. \quad (22.127)$$

The *in situ* temperature of a fluid element thus evolves according to convergence of the enthalpy fluxes, matter concentration fluxes, pressure changes, and frictional dissipation.

22.8.3 Evolution of potential temperature

We can convert the *in situ* temperature equation (22.127) into a version of the potential temperature equation by recalling the expression (21.93) for the lapse rate

$$\hat{\Gamma} = \left[\frac{\partial T}{\partial p} \right]_{C,S} = \frac{T \alpha_T}{\rho c_p} \quad (22.128)$$

so that equation (22.127) takes the form

$$c_p \rho \left[\frac{DT}{Dt} - \hat{\Gamma} \frac{Dp}{Dt} \right] = -\nabla \cdot \mathbf{J}_{\mathcal{H}} + \left[\frac{\partial \mathcal{H}}{\partial C} \right]_{T,p} \nabla \cdot \mathbf{J}_C + \rho \epsilon. \quad (22.129)$$

Making use of the definition (21.98) for potential temperature renders

$$c_p \rho \frac{D\theta}{Dt} = -\nabla \cdot \mathbf{J}_{\mathcal{H}} + \left[\frac{\partial \mathcal{H}}{\partial C} \right]_{T,p} \nabla \cdot \mathbf{J}_C + \rho \epsilon. \quad (22.130)$$

As expected, pressure changes are removed from the evolution equation for potential temperature.

22.8.4 Conservative Temperature for the ocean

Rather than expressing enthalpy as a function $\mathcal{H}(T, p, C)$, we make use of its natural coordinate dependence $\mathcal{H}(S, p, C)$ from Section 21.4.3, which leads to the enthalpy equation in the form (22.121)

$$\rho \frac{D\mathcal{H}}{Dt} = \frac{Dp}{Dt} - \nabla \cdot \mathbf{J}_H + \rho \epsilon. \quad (22.131)$$

The pressure term arises just like for *in situ* temperature. Its presence suggests we introduce the *potential enthalpy*.

Potential enthalpy and Conservative Temperature

The potential enthalpy is defined to be the enthalpy of a fluid element moved to a reference pressure, p_R , while maintaining fixed specific entropy and fixed tracer concentration

$$\mathcal{H}^{\text{pot}}(\mathcal{S}, C) = \mathcal{H}(\mathcal{S}, p_R, C). \quad (22.132)$$

As for potential temperature (Section 21.8.2), it is most convenient to take p_R as the standard atmospheric pressure, thus corresponding to the standard pressure at the air-sea interface. This definition parallels that for potential temperature given by equation (21.103). By construction, the material time derivative of potential enthalpy is given by

$$\rho \frac{D\mathcal{H}^{\text{pot}}}{Dt} = \rho \left[\frac{\partial \mathcal{H}^{\text{pot}}}{\partial \mathcal{S}} \right]_C \frac{D\mathcal{S}}{Dt} + \rho \left[\frac{\partial \mathcal{H}^{\text{pot}}}{\partial C} \right]_{\mathcal{S}} \frac{DC}{Dt} \quad (22.133a)$$

$$= \theta (-\nabla \cdot \mathbf{J}_{\mathcal{S}} + \Sigma_{\mathcal{S}}) - \mu_R \nabla \cdot \mathbf{J}_C \quad (22.133b)$$

$$= (\theta/T) [\rho \epsilon - \nabla \cdot \mathbf{J}_H] - [\mu_R - (\theta/T) \mu] \nabla \cdot \mathbf{J}_C, \quad (22.133c)$$

where we set

$$\theta = \left[\frac{\partial \mathcal{H}^{\text{pot}}}{\partial \mathcal{S}} \right]_C \quad \text{and} \quad \mu_R = \left[\frac{\partial \mathcal{H}^{\text{pot}}}{\partial C} \right]_{\mathcal{S}} \quad (22.134)$$

and used equation (22.118) for the entropy source, $\Sigma_{\mathcal{S}}$. Now define the Conservative Temperature, Θ , via

$$c_p^{\text{ref}} \Theta \equiv \mathcal{H}^{\text{pot}}(\mathcal{S}, C) = \mathcal{H}(\mathcal{S}, p_R, C), \quad (22.135)$$

where c_p^{ref} is an arbitrary reference specific heat capacity. For the ocean, [McDougall \(2003\)](#) suggested that c_p^{ref} be chosen so that $\Theta = \theta$ at a salinity of 35 parts per thousand. [McDougall \(2003\)](#) furthermore argued that the terms appearing in the potential enthalpy equation (22.133c) are well approximated for the ocean by just the convergence of the enthalpy flux. Hence, the Conservative Temperature satisfies, to a very good approximation, the source-free tracer equation

$$\rho c_p^{\text{ref}} \frac{D\Theta}{Dt} = -\nabla \cdot \mathbf{J}_H. \quad (22.136)$$

Key points regarding the Conservative Temperature equation

The Conservative Temperature equation (22.136) is mathematically identical to the material tracer equation, and as such it offers an elegant means to prognose thermodynamic properties of the fluid and to perform budget analyses. We further emphasize two points in regards to this equation relative to the potential temperature equation (22.130).

- The source terms (those not associated with flux convergences) on the right hand side of the potential temperature equation are much larger than those in the Conservative Temperature equation. In particular, [McDougall \(2003\)](#) argued that they are roughly 100 times larger in certain regions of the ocean.
- The heat capacity appearing in the Conservative Temperature equation is a fixed constant, by construction. This feature contrasts to the space-time variable heat capacity, c_p , appearing in both the *in situ* temperature equation (22.127) and potential temperature equation (22.130). The space-time variations of c_p are not negligible (e.g., order 5% for the global ocean), thus making the non-constant heat capacity required for the T and θ equations very inconvenient for purposes of budget analyses.

We close by noting that the enthalpy flux, \mathbf{J}_H , is related to the entropy flux and concentration flux as per equation (22.123). As discussed in Section 2.6 of [Olbers et al. \(2012\)](#), the dominant terms appearing in this flux arise from entropy, which itself is largely due to fluxes of temperature. Consequently, the flux \mathbf{J}_H is well approximated as a flux just of Θ .

22.8.5 Further study

The discussion in this section largely followed the more complete ocean discussion given in Section 2.6 of [Olbers et al. \(2012\)](#), which is itself based on [McDougall \(2003\)](#). Development for a realistic atmosphere must consider phase changes (liquid-vapor and liquid-solid), since the associated latent heat exchanges are leading order contributions to the enthalpy budget. Additionally, the role of frictional dissipation is not negligible in the atmosphere whereas it is negligible in the ocean (see Section 22.5.3).

22.9 Atmospheric moist static energy

Follow the discussion of Section 4.8 in [Gill \(1982\)](#). Then derive the lapse rate as per David Romps' discussion at the convection workshop 8-10 Feb 2018 in Princeton.

Let us consider the enthalpy equation (21.134) for the special case in which entropy changes occur only through heating in a homogeneous fluid

$$T dS = dQ. \quad (22.137)$$

Now assume heating occurs through radiation plus another term whose

$$T dS = dQ \quad (22.138a)$$

$$= L dT + dQ_{rad}, \quad (22.138b)$$

where L is the latent heat of vaporization.

22.9.1 Further study

This section is incomplete. The intent is to merge material from Section 1.10 of [Vallis \(2017\)](#) to lecture notes from David Romps.

22.10 Summary of the ocean-atmosphere equations

We have now completed the derivation of the full suite of partial differential equations describing rotating and stratified fluids. We thus find it useful to again list these *thermo-hydrodynamical* equations and summarize their content:

$$\rho \frac{D\mathbf{v}}{Dt} + 2\rho \boldsymbol{\Omega} \wedge \mathbf{v} = -\rho \nabla \Phi - \nabla p + \nabla \cdot \tau \quad \text{momentum} \quad (22.139)$$

$$\frac{D\rho}{Dt} = -\rho \nabla \cdot \mathbf{v} \quad \text{mass continuity} \quad (22.140)$$

$$\rho \frac{DC}{Dt} = -\nabla \cdot \mathbf{J}(C) \quad \text{matter conservation} \quad (22.141)$$

$$\rho \frac{D\Theta}{Dt} = -\nabla \cdot \mathbf{J}(\Theta) \quad \text{potential enthalpy conservation} \quad (22.142)$$

$$\rho = \rho(C, \Theta, p) \quad \text{equation of state.} \quad (22.143)$$

It is a testament to the success of classical continuum mechanics that these equations are of use for describing fluid phenomena from the millimetre scale to the astrophysical scale. We summarize the following terms in these equations.

- **LINEAR MOMENTUM AND VELOCITY:** Newton's second law of motion, as developed for a fluid in Chapter 18, provides the prognostic equation for the velocity field, \mathbf{v} . Each of the three velocity components evolves according to its respective dynamical equation (22.139). As noted at the end of Section 18.1.3, we write the momentum equation in the form (22.139) by separating the time dependence of the basis vectors into a term arising from solid-body rotation (which leads to planetary Coriolis and planetary centrifugal accelerations) and a term arising from the motion of the fluid relative to the rotating sphere (which leads to the metric acceleration when using non-Cartesian coordinates).
- **MASS CONSERVATION:** Kinematics provides a constraint on the velocity field according to the needs of mass conservation for a fluid element (Section 15.2). This constraint leads to the continuity equation (22.140).
- **MATERIAL TRACER CONSERVATION:** Kinematic constraints from the conservation of matter (Section 16.1) leads to the material tracer equation (22.141). Evolution is determined by the convergence of tracer fluxes, \mathbf{J} , with this flux specified by molecular diffusion as discussed in Section 48.3, or through other parameterized processes when sampling flow on scales larger than millimetres (see Chapter 50).
- **THERMODYNAMIC TRACER:** The Conservative Temperature, Θ , (Section 22.8.4), evolves according to the convergence of fluxes, just like a material tracer.
- **DENSITY:** The *in situ* density can be updated in time via mass continuity (equation (22.140)) or via knowledge of (C, Θ, p) . We discussed the many forms of density for the ocean and atmosphere in Section 25.3.
- **PRESSURE:** There is no prognostic equation for pressure. Rather, pressure is diagnosed based on knowledge of other fields. Here are sketches of how that diagnostic calculation is performed.
 - For an ideal gas, pressure is diagnosed from the ideal gas relation (21.115) using the density and temperature.
 - For fluid flow maintaining an approximate hydrostatic balance (Section 23.3), pressure is diagnosed at a point through knowledge of the weight per area above the point.
 - For an incompressible flow as per the oceanic Boussinesq fluid (Chapter 24), pressure is no longer connected thermodynamically to partial derivatives of the thermodynamic potentials (Section 21.4). Instead, pressure is determined kinematically by the incompressibility constraint. In particular, for a non-hydrostatic Boussinesq fluid, pressure is diagnosed by solving a Poisson equation derived from taking the divergence of the momentum equation (see Section 24.3).
- **GEOPOTENTIAL:** The geopotential, Φ , is specified once the height above an arbitrary reference level is known, as well as the effective gravitational acceleration (Section 10.12.4). For geophysical fluid studies, the reference level is generally taken at the level of a resting sea surface. We thus often write the radial coordinate as

$$r = R_e + z \tag{22.144}$$

where $R_e = 6.371 \times 10^6$ m is the earth radius (equation (10.134)), and z is the geopotential coordinate measuring the height above sea level.

- EARTH'S SPIN: The earth's angular velocity, Ω , is constant for geophysical fluid studies of concern here. Its value is discussed in Section 10.2.

- FRICTION: The friction vector,

$$\rho \mathbf{F} = \nabla \cdot \boldsymbol{\tau}, \quad (22.145)$$

is the divergence of a symmetric and trace-free deviatoric stress tensor, $\boldsymbol{\tau}$ (Section 19.7). The stress tensor is determined through a constitutive relation as a function of the strain and viscous properties.

- BOUNDARY CONDITIONS: Boundary conditions consist of the exchange of matter, momentum, and enthalpy with the surrounding media, such as the solid earth or another fluid component (e.g., atmosphere-ocean exchange). We discuss the boundary conditions for matter in Section 15.4; for momentum in Section 19.9; and for ocean buoyancy in Section 51.4.

22.11 Concerning conservation laws

A central facet of theoretical physics concerns the development of concepts and tools to expose conservation laws and their underlying symmetries.⁷ We routinely make use of such laws in geophysical fluid mechanics to provide constraints on the fluid motion and to study budgets of corresponding properties to help understand fundamental processes. As such, conservation laws offer great physical insight and predictive utility. We close this chapter by summarizing some conceptual points concerning conservation laws. In particular, we identify the need to distinguish laws that involve just the convergence of a flux from those that also include non-conservative “source” terms. We also distinguish local or material conservation for a property, as reflected by the property remaining invariant following a fluid element, versus global conservation, whereby a property is invariant only when integrated over a finite region whose boundaries are closed to fluxes that affect the property.

22.11.1 Flux-form conservation laws

Certain scalar properties studied in fluid mechanics satisfy Eulerian flux-form conservation laws that are written as

$$\frac{\partial(\rho\psi)}{\partial t} = -\nabla \cdot (\rho \mathbf{v} \psi + \mathbf{J}). \quad (22.146)$$

The right hand side involves a flux that is comprised of an advective term, $\rho \mathbf{v} \psi$, plus a non-advective term, \mathbf{J} . Examples of flux-form conservation laws of this form include the material tracer concentration, $\psi = C$, as in equation (22.141); the Conservative Temperature, $\psi = \Theta$, as in equation (22.142); the total energy, $\psi = \mathcal{E}$, in the absence of astronomical forces, as in equation (22.119b); and the potential vorticity, $\psi = Q$, as in equation (38.40). In Chapter 16, we saw how this differential equation leads to finite volume conservation properties for the integral of ψ -stuff within a region, $\int_{\mathcal{R}} \psi \rho dV$, with the evolution of this integral only affected by area integrated fluxes, $\rho \mathbf{v} \psi + \mathbf{J}$, penetrating the region boundary (mathematically seen by applying the divergence theorem).

Such conservation laws are a direct consequence of the local conservation of ψ -stuff within the fluid. That is, the amount of ψ -stuff changes at a point only through the local convergence of

⁷See Chapter 11 for the connection between conservation laws and symmetries, as embodied by Noether's theorem.

fluxes onto that point, and likewise for a finite region. Conservation laws of the form (22.146) are consistent with basic notions of causality and locality that appear throughout physics, with a discussion of such conservation laws offered in Section 27-1 of [Feynman et al. \(1963\)](#).

22.11.2 Conservation laws that are not flux-form

The presence of source/sinks are relevant for chemical or biogeochemical reactions, whereby matter is converted from one form to another. Such processes are not mathematically represented as the convergence of a flux. As such, they are not contained in the conservation law (22.146) and they are correspondingly referred to as *non-conservative processes*. Even without chemical reactions, not all fluid properties satisfy flux-form conservation laws of the form (22.146). For example, linear momentum of a fluid element is affected by pressure, Coriolis, and effective gravity, and these processes are not represented as the convergence of a flux.

As discussed in Chapter 11, conservation laws are associated with symmetries of the physical system. Correspondingly, non-conservative terms appearing in an evolution equation often reflect the breaking of a symmetry. For example, motion around a sphere does not conserve linear momentum even in the absence of forces, whereas linear momentum is conserved for free motion in a planar geometry.

22.11.3 Non-material or wave-like transport of properties

Pressure is of particular note since pressure perturbations travel through a compressible fluid via acoustic waves (Section 18.5), or, in the case of an incompressible flow, a pressure perturbation is felt globally and instantaneously as reflected in the elliptic Poisson equation satisfied by pressure (Section 24.3). More generally, the wave mediated transfer of forces, or other fluid properties such as momentum or mechanical energy, is an example of a non-material transfer; i.e., a transfer of information not arising from the transfer of matter. Non-material transfer generally occurs much faster than material transfer, with matter transport only mediated through advection and diffusion. Correspondingly, material substances (and potential enthalpy) are not directly affected by wave transport. Rather, waves affect material substances only so far as they affect advection and diffusion.

22.11.4 Local versus global conservation

In fluid mechanics we often encounter properties that remain invariant when following a fluid element, such as the mass of a fluid element. Mass invariance for a fluid element reflects matter conservation in classical physics, which in turn motivates a kinematic perspective that follows fluid elements whose mass remains constant. A related example concerns the matter content of a fluid element. In the absence of irreversible mixing, the matter content (i.e., tracer concentration) of the fluid element remains invariant so that its tracer concentration is materially constant

$$\rho \frac{D\psi}{Dt} = 0. \tag{22.147}$$

For example, in the absence of mixing the salt content of seawater and the water content of moist air are invariant, so that the salt concentration and water concentration in a fluid element remains invariant. Such material invariances are termed *local* conservation laws since they hold locally to every fluid element. Correspondingly, in the presence of mixing between two fluid elements, the net material tracer in the combined fluid element equals to the sum of the tracer content in the contributing elements. That is, we can perform local budgets on fields that possess a local

conservation law, and it is for this reason we sometimes refer to fluid properties satisfying local conservation laws as *conservative properties*.

But what of fluid properties that satisfy a flux-form conservation law of the form (22.146), and yet do not satisfy the material invariance law (22.147) even in the absence of irreversible processes? For example, consider the total energy, \mathcal{E} , from Section 22.5. Even in the absence of entropy sources (i.e., no mixing) and astronomical forces (i.e., constant gravity), mechanical work from pressure modifies the internal energy of the fluid element via the energy equation (22.119a)

$$\rho \frac{D\mathcal{E}}{Dt} = -\nabla \cdot (p\mathbf{v}). \quad (22.148)$$

Pressure work means that when two fluid elements are combined, the total energy of the combined fluid, \mathcal{E}_{12} , is not generally equal to the sum of their separate total energies,

$$\mathcal{E}_{12} \neq \mathcal{E}_1 + \mathcal{E}_2. \quad (22.149)$$

Hence, total energy is not a locally conservative property, in contrast to material tracers. However, when integrated globally over a domain that is closed to the fluxes of total energy (i.e., there is no pressure work applied to the domain boundaries), and when there are no astronomical forces, then the domain integrated total energy, $\int \rho \mathcal{E} dV$, remains constant. This *global* conservation law means that the total energy is conserved globally, which is familiar from classical physics. Global conservation is a less powerful statement than local conservation. Even so, global conservation can be of great use when studying energy transformations within a closed domain.⁸

We can summarize the above discussion by noting that for a fluid property to be conservative (i.e., to satisfy a local conservation law), it is not sufficient that the material derivative of that property to be given by the divergence of a flux. Rather, what is needed is that the material derivative must equal to the divergence of a flux that is zero in the absence of mixing at that point in space and time. A diffusive flux satisfies this property (see Chapter 48 for more on diffusion). In contrast, the pressure flux convergence acting to modify total energy, $-\nabla \cdot (p\mathbf{v})$, can be nonzero even in regions where there is no mixing since pressure is transported by waves (Section 22.11.3).

22.11.5 Concerning potential scalar properties

As introduced in Sections 21.8 and 22.8.4, to study fluid mixing it is useful to work with scalar fields that are not affected by adiabatic and isentropic pressure work. For this reason, rather than *in situ* temperature, we prefer to work with potential temperature, θ , or Conservative Temperature, Θ . Some potential properties are also endowed with the local conservation property discussed above, which makes local budgets available just like for a material tracer. For example, Conservative Temperature is very well approximated as a conservative property, with its sources far smaller than potential temperature (Section 22.8.4). In contrast, neither *in situ* temperature nor total energy are potential properties since an adiabatic and isentropic change in pressure alters the total energy and *in situ* temperature of a fluid element. Correspondingly, as noted in Section 22.11.4, total energy and *in situ* temperature are not conservative properties.

22.12 Exercises

EXERCISE 22.1: CROCCO'S THEOREM

Prove that the spatial gradient of the Bernoulli potential for a single-component steady state perfect

⁸We further encounter these distinctions between local and global conservation laws when studying potential vorticity in Chapter 39.

fluid can be written

$$\nabla \mathcal{B} = T \nabla \mathcal{S} + \mathbf{v} \wedge \boldsymbol{\omega}_a. \quad (22.150)$$

This result is known as Crocco's Theorem ([Crocco, 1937](#)). It means that the velocity for a single-component perfect fluid in steady state is aligned parallel to isosurfaces of both the Bernoulli potential and the specific entropy. We encounter another form of this theorem for the steady state shallow water equations in Exercise [35.6](#).

Hint: to help formulate the proof, study the discussion in Section [22.6.2](#) where we showed that the Bernoulli potential is constant along a steady flow streamline in a perfect fluid. Also recall equation [\(22.50\)](#) valid for a steady state and applied here to a single-component fluid.

EXERCISE 22.2: FRICTIONAL DISSIPATION FROM VISCOSITY

This exercise works through a simple case of the more general considerations from Section [22.3.2](#). Here, assume the friction in the momentum equation takes the form

$$\rho \mathbf{F} = \nabla \cdot (\rho \nu \nabla \mathbf{v}) = \partial_n (\rho \nu \partial_n \mathbf{v}), \quad (22.151)$$

with $\nu > 0$ a scalar kinematic viscosity (generally non-constant). Show that when integrated over the full domain

$$\int \mathbf{F} \cdot \mathbf{v} \rho dV < 0, \quad (22.152)$$

where boundary terms are dropped. Hence, the global integrated kinetic energy is dissipated (reduced) through the impacts of viscosity. This dissipation of mechanical energy is converted to an increase in internal energy through Joule heating. Hint: for this exercise, it is sufficient to assume Cartesian tensors so that

$$\rho \mathbf{F} \cdot \mathbf{v} = \rho F_m v_m = \partial_n (\rho \nu \partial_n v_m) v_m. \quad (22.153)$$

EXERCISE 22.3: CENTER OF MASS TRANSPORT THEOREM

Consider a field, ψ , that satisfies the standard conservation law

$$\rho \frac{D\psi}{Dt} = -\nabla \cdot \mathbf{J}, \quad (22.154)$$

within a region, \mathcal{R} , that moves with the fluid flow.

- (a) Prove the transport theorem

$$\frac{d}{dt} \int_{\mathcal{R}} \psi \mathbf{x} \rho dV = \int_{\mathcal{R}} (\mathbf{J} + \rho \psi \mathbf{v}) dV - \oint_{\partial \mathcal{R}} \mathbf{x} (\hat{\mathbf{n}} \cdot \mathbf{J}) dV. \quad (22.155)$$

Hint: multiply both sides of equation [\(22.154\)](#) by x_i .

- (b) Offer an interpretation of equation [\(22.155\)](#). Hint: first consider the special case that the total ψ -stuff, defined by $\Psi \equiv \int_{\mathcal{R}} \psi \rho dV$, is constant when following the flow. Further hint: consider the more specialized case of $\psi = 1$ and $\mathbf{J} = 0$, and then make use of Exercise [15.1](#).

EXERCISE 22.4: THERMODYNAMIC MANIPULATIONS FOR IDEAL GASES

This question develops some manipulations with the potential temperature.

- (a) Beginning with the expression [\(21.159\)](#) for potential temperature of an ideal gas, show that

$$d\theta = \frac{\theta}{T} \left[dT - \frac{\nu_s}{c_p} dp \right]. \quad (22.156)$$

(b) Given the result (22.156), show that an ideal gas satisfies the following relation

$$T dS = \frac{c_p T}{\theta} d\theta. \quad (22.157)$$

Whereas the relation (21.111) holds for a general fluid only at the reference pressure, this exercise shows that it holds for an ideal gas at all pressures. As a result, a moving fluid of ideal gas satisfies the material time relation

$$T \frac{D\dot{S}}{Dt} = \frac{c_p T}{\theta} \frac{D\theta}{Dt} \Rightarrow \frac{c_p T}{\theta} \frac{D\theta}{Dt} = \dot{Q}. \quad (22.158)$$

EXERCISE 22.5: THERMODYNAMIC MANIPULATIONS FOR A LIQUID

Consider seawater with specific entropy given by (see Section 1.7.2 of [Vallis \(2017\)](#))

$$\dot{S}(S, T, p) = S_0 + c_{p0} \ln(T/T_o) [1 + \beta_s^*(S - S_o)] - \alpha_o p \left[\beta_T + \beta_T \gamma^* \frac{p}{2} + \beta_T^* (T - T_o) \right], \quad (22.159)$$

and corresponding specific heat capacity at constant pressure

$$c_p(S, T, p) = c_{p0} [1 + \beta_s^*(S - S_o)] - \alpha_o p \beta_T^* T. \quad (22.160)$$

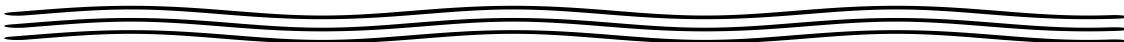
In these equations, T is the *in situ* temperature, S is the salinity, and p is the *in situ* pressure. All other terms on the right hand side to these expressions are empirical constants. Verify that an infinitesimal change in entropy for a fluid element with constant composition is given by

$$\theta dS = c_p(S, \theta, p_R) d\theta, \quad (22.161)$$

where θ is the potential temperature and p_R is the corresponding reference pressure. Consequently, we can write for a moving fluid element

$$\dot{Q} = \frac{c_p T}{\theta} \frac{D\theta}{Dt}, \quad (22.162)$$

where we evaluate the non-constant heat capacity at $c_p(S, \theta, p_R)$. We see that certain liquids have an expression for heating that is analogous to that for an ideal gas, with the ideal gas case discussed in Exercise 22.4. Hint: Make use of the identity (21.103).



Hydrostatic and tangent plane approximations

The ocean and atmosphere thermo-hydrodynamical equations (22.139)-(22.143) explain a huge range of phenomena. Yet by encapsulating so many physical scales of motion and associated dynamical processes, the equations are difficult to manage when studying a focused dynamical regime. Therefore, it is common to approximate or filter the equations to remove scales of little interest, thus enabling a more telescopic view of the dynamics. In this chapter, we develop certain approximations commonly used in geophysical fluid mechanics. In particular we develop the hydrostatic primitive equations as well as the tangent plane approximations.

READER'S GUIDE TO THIS CHAPTER

The equations developed in this chapter, and their associated approximations, will be used extensively throughout the remainder of this book.

23.1	The hydrostatic primitive equations	492
23.1.1	Hydrostatic approximation	492
23.1.2	Shallow fluid approximation	492
23.1.3	Traditional approximation	493
23.1.4	Summary of the hydrostatic primitive equations	493
23.1.5	Comments and further study	494
23.2	The f -plane and β -plane approximations	494
23.2.1	Tangent plane approximation	494
23.2.2	Traditional approximation and the f -plane	495
23.2.3	β -plane approximation	495
23.3	Approximate hydrostatic balance	496
23.3.1	Expressions for the hydrostatic pressure	496
23.3.2	Evolution of hydrostatic pressure	496
23.3.3	Heuristic scaling	498
23.3.4	Removal of a dynamically irrelevant background state	499
23.3.5	Decomposing the horizontal pressure gradient	500
23.3.6	Horizontal hydrostatic pressure gradient in a mass conserving fluid	501
23.3.7	Ocean dynamic topography	503
23.3.8	Surfaces of atmospheric geopotential height and pressure	504
23.3.9	Concerning vertical motion	504
23.3.10	Further study	505
23.4	Homogeneous fluid in a rotating tank	505
23.4.1	What about the planet's rotation?	505
23.4.2	Formulating the equations of motion	505
23.4.3	Solid-body rotation and free surface shape	506

23.4.4 Further study	508
23.5 Exercises	508

23.1 The hydrostatic primitive equations

The hydrostatic *primitive equations* provide a set of approximate equations for use in studying large-scale atmospheric and oceanic phenomena. Indeed, nearly all numerical models of the large-scale atmospheric and oceanic circulation are based on the primitive equations. They make use of the following three approximations.

23.1.1 Hydrostatic approximation

As discussed in Section 18.3, a static fluid in a gravity field maintains an exact hydrostatic balance, whereby the pressure at a point equals to the weight per area of fluid above that point. As shown in Section 23.3, the hydrostatic balance is very closely maintained column-by-column for the large scales in a moving geophysical fluid. Hence, it is appropriate for many purposes to take the *hydrostatic approximation* for the vertical momentum equation, with this approximation central to the study of geophysical fluid dynamics.

The hydrostatic approximation results in a balance in the vertical momentum equation (18.21) between the vertical pressure gradient and the effective gravitational force

$$\frac{\partial p}{\partial r} = -\rho g, \quad (23.1)$$

with this balance holding separately for each vertical column. Vertical integration of this equation, while assuming g is constant, renders a diagnostic expression for the hydrostatic pressure at a point as a function of the weight per horizontal area above the point

$$p(r, \lambda, \phi, t) = p(r_0, \lambda, \phi, t) + g \int_r^{r_0} \rho(r', \lambda, \phi, t) dr'. \quad (23.2)$$

Note that we exposed the horizontal space dependence along with the time dependence for the density and hence the hydrostatic pressure. That is, an approximate hydrostatic fluid has horizontal pressure gradients as well as time dependence.

We emphasize that in making the hydrostatic approximation that we are *not* assuming that vertical motion vanishes. In fact, there is vertical motion. But with the hydrostatic approximation, the vertical motion is not prognosed by the vertical momentum equation. Instead, it must be diagnosed via the constraints imposed on the motion. We have more comment on this point in Section 23.3.9.

23.1.2 Shallow fluid approximation

The ocean and atmosphere each form a fluid shell that envelopes the outer portion of the planet. The thickness of the these fluids is small relative to the earth's radius. The shallow fluid approximation¹ builds this scale separation into the equations of motion by setting the radial coordinate equal to the earth's radius

$$r = R_e + z \approx R_e. \quad (23.3)$$

¹The shallow fluid approximation is distinct from the *shallow water approximation* treated in Part V of this book.

This approximation is made where r appears as a multiplier, but not as a derivative operator. For example, the spherical coordinate gradient operator (18.23) takes the approximate form

$$\nabla = \frac{\hat{\lambda}}{r \cos \phi} \frac{\partial}{\partial \lambda} + \frac{\hat{\phi}}{r} \frac{\partial}{\partial \phi} + \hat{r} \frac{\partial}{\partial r} \approx \frac{\hat{\lambda}}{R_e \cos \phi} \frac{\partial}{\partial \lambda} + \frac{\hat{\phi}}{R_e} \frac{\partial}{\partial \phi} + \hat{r} \frac{\partial}{\partial r}. \quad (23.4)$$

23.1.3 Traditional approximation

The *traditional approximation* comprises three approximations that come as a package.

Coriolis acceleration

The traditional approximation sets to zero the Coriolis terms in the horizontal momentum equations involving the vertical velocity. We are thus concerned only with the local vertical component of the earth's angular rotation vector (see discussion in Section 10.11.7)

$$\boldsymbol{\Omega} = \Omega \hat{\mathbf{Z}} = \Omega (\hat{\phi} \cos \phi + \hat{r} \sin \phi) \rightarrow \Omega \sin \phi \hat{r} = \mathbf{f}/2, \quad (23.5)$$

where

$$\mathbf{f} = (2 \Omega \sin \phi) \hat{r} \quad (23.6)$$

is the Coriolis parameter and $\hat{\mathbf{Z}}$ is the spherical earth unit vector pointing out of the north pole (Figure 8.1).²

Metric terms

The traditional approximation also drops the metric terms, uw/r and vw/r , associated with the vertical velocity as they appear in the horizontal momentum equations (18.19) and (18.20). These terms are generally smaller than the other terms since w is much smaller than the horizontal velocity for large-scale geophysical fluid flow.

Self consistency

The shallow fluid approximation and both parts of the traditional approximation must be taken together in order to maintain a consistent energy and angular momentum conservation principle for the resulting equations. As shown in Exercise 23.1, taking one but not the other leads to an inconsistent set of equations

23.1.4 Summary of the hydrostatic primitive equations

The above approximations lead to the primitive equations written in spherical coordinates

$$\frac{Du}{Dt} - \frac{uv \tan \phi}{R_e} - fv = -\frac{1}{\rho R_e \cos \phi} \frac{\partial p}{\partial \lambda} + F^\lambda \quad (23.7)$$

$$\frac{Dv}{Dt} + \frac{u^2 \tan \phi}{R_e} + fu = -\frac{1}{\rho R_e} \frac{\partial p}{\partial \phi} + F^\phi \quad (23.8)$$

$$\frac{\partial p}{\partial z} = -g \rho, \quad (23.9)$$

²We use the capital $\hat{\mathbf{Z}}$ to distinguish this north pole unit vector from the local $\hat{\mathbf{z}} = \hat{r}$ unit vector pointing vertical relative to a tangent plane discussed in Section 23.2.

where the gradient operator is given by equation (23.4). We can write these equations in the succinct form

$$\left[\frac{D}{Dt} + (f + u \tan \phi / R_e) \hat{z} \wedge \right] \mathbf{u} = -\rho \nabla \Phi - \nabla p + \mathbf{F}, \quad (23.10)$$

where

$$\mathbf{F} = \hat{\lambda} F^\lambda + \hat{\phi} F^\phi \quad (23.11)$$

is the horizontal friction vector. Furthermore, the material time derivative in this equation signifies the relative acceleration

$$\frac{Du}{Dt} = \hat{\lambda} \frac{Du}{Dt} + \hat{\phi} \frac{Dv}{Dt}. \quad (23.12)$$

23.1.5 Comments and further study

The primitive equations make use of the momentum equations, which contrasts to *non-primitive* equation methods that develop evolution equations for the vorticity and divergence. [Smagorinsky \(1963\)](#) was an early proponent of the hydrostatic primitive equations for use in studying the large-scale ocean and atmospheric circulation. These equations form the basis for many general circulation models of the atmosphere and ocean. However, it is notable that even finer resolution simulations are making use of the non-hydrostatic equations. Non-hydrostatic simulations are particularly relevant when studying clouds in the atmosphere and fine-scale mixing in the ocean, with both of these processes involving nontrivial vertical accelerations that break the hydrostatic approximation. These models sometimes also time step the momentum equations, and as such are referred to as *non-hydrostatic primitive equation* models.

23.2 The f -plane and β -plane approximations

Spherical coordinates are suited for the study of planetary fluid dynamics for cases where the fluid samples the earth's sphericity. However, spherical coordinates remain more complicated to work with than Cartesian coordinates. We are thus led to consider the utility of an idealized tangent plane configuration as part of a hierarchy of theoretical models to help understand geophysical fluid motion. This motivation leads to the f -plane and β -plane approximations, which are the two cases considered as part of the tangent plane approximations.

23.2.1 Tangent plane approximation

Consider a tangent plane located at latitude $\phi = \phi_0$ and introduce a Cartesian set of coordinates according to

$$(x, y, z) = (R_e \lambda \cos \phi_0, R_e (\phi - \phi_0), z) \quad (23.13)$$

$$(\hat{x}, \hat{y}, \hat{z}) = (\hat{\lambda}, \hat{\phi}, \hat{r}). \quad (23.14)$$

Use of these Cartesian coordinates leads to the following inviscid (i.e., no friction) equations of motion

$$\frac{Du}{Dt} + 2(\Omega^y w - \Omega^z v) = -\frac{1}{\rho} \frac{\partial p}{\partial x} \quad (23.15)$$

$$\frac{Dv}{Dt} + 2(\Omega^z u - \Omega^x w) = -\frac{1}{\rho} \frac{\partial p}{\partial y} \quad (23.16)$$

$$\frac{Dw}{Dt} + 2(\Omega^x v - \Omega^y u) = -\frac{1}{\rho} \frac{\partial p}{\partial z} - g, \quad (23.17)$$

with rotational vector components

$$\boldsymbol{\Omega} = \Omega (\cos \phi_0 \hat{\mathbf{y}} + \sin \phi_0 \hat{\mathbf{z}}). \quad (23.18)$$

Note the absence of metric terms due to the use of Cartesian coordinates on a flat planar geometry.

The tangent plane approximation originates from the geopotential vertical coordinate system (Section 10.13.3). In that coordinate system, the effective gravitational acceleration (gravity plus planetary centrifugal) is aligned with the local vertical direction. Correspondingly, the resulting tangent plane equations have the effective gravitational force aligned in the $\hat{\mathbf{z}}$ direction. As seen in Section 23.4, these equations are slightly different from those describing a fluid in a rotating tank, in which the effective gravity is not aligned with the vertical.

23.2.2 Traditional approximation and the f -plane

The traditional approximation retains only the local vertical component of the rotation vector, resulting in

$$\frac{Du}{Dt} - fv = -\frac{1}{\rho} \frac{\partial p}{\partial x} \quad (23.19)$$

$$\frac{Dv}{Dt} + fu = -\frac{1}{\rho} \frac{\partial p}{\partial y} \quad (23.20)$$

$$\frac{Dw}{Dt} = -\frac{1}{\rho} \frac{\partial p}{\partial z} - g. \quad (23.21)$$

The f -plane makes use of a constant Coriolis parameter

$$f = 2\Omega \sin \phi_0 \equiv f_0. \quad (23.22)$$

The f -plane approximation is the simplest model for a rotating fluid, and as such it provides an end member in the hierarchy of theoretical models for geophysical fluid flows.

23.2.3 β -plane approximation

As seen in Section 34.7, *Rossby waves* are planetary scale waves that sample the earth's spherical geometry. The essential ingredient for their existence is the latitudinal dependence of the Coriolis parameter. To capture Rossby waves on a tangent plane requires the meridional gradient of the Coriolis parameter, with a linear dependence sufficient

$$f = f_0 + R_e^{-1} (2\Omega \cos \phi_0) (y - y_0). \quad (23.23)$$

For many purposes the β -plane approximation only depends on the meridional gradient of the Coriolis parameter, in which case we generally set

$$f = f_0 + \beta y \quad (23.24)$$

$$\beta = \partial f / \partial y = (2\Omega / R_e) \cos \phi_0, \quad (23.25)$$

thus ignoring the constant $-(2\Omega \cos \phi_0) y_0 / R_e$.

23.3 Approximate hydrostatic balance

For a static fluid with identically zero net acceleration, the vertical pressure gradient precisely balances the weight of fluid in thus realizing exact hydrostatic balance. We discussed this static solution to the equations of motion in Sections 18.3 and 19.4. For a moving fluid with scales of motion that maintain a small vertical to horizontal aspect ratio, the presentation in this section reveals that the vertical pressure gradient and gravitational acceleration individually remain far larger than other accelerations acting on a fluid element. In this case, the vertical momentum equation, even for the moving fluid, remains approximately in hydrostatic balance column-by-column. We thus have a fluid state whereby each vertical fluid column is in hydrostatic balance, and yet there are horizontal pressure gradients that drive motion. In this section we study aspects of such approximately hydrostatic fluids.

23.3.1 Expressions for the hydrostatic pressure

Making the hydrostatic approximation in the vertical momentum equation leads to the local balance

$$\frac{\partial p}{\partial z} = -\rho g. \quad (23.26)$$

Vertically integrating upward from a point within the ocean to the ocean surface leads to the hydrostatic pressure

$$p(x, y, z, t) = p_a(x, y, t) + g \int_z^\eta \rho(x, y, z', t) dz'. \quad (23.27)$$

In this equation we wrote $p(\eta) = p_a$ for the pressure at the ocean free surface, $z = \eta(x, y, t)$, arising from the weight of the overlying atmosphere or sea ice; i.e., this is the applied pressure acting on the top of the ocean fluid. A similar integration applies to the atmosphere

$$p(x, y, z, t) = g \int_z^{z_{\text{top}}} \rho(x, y, z', t) dz', \quad (23.28)$$

where $z = z_{\text{top}}$ is the top of the atmosphere, sometimes approximated by $z_{\text{top}} = \infty$. For both the ocean and the atmosphere, we assume g remains a constant over the vertical extent of the fluid, which is a sensible approximation even for the top of the atmosphere.

In both the ocean and atmosphere, the hydrostatic pressure at a vertical position, z , equals to the weight per horizontal area of matter above that position, with equations (23.27) and (23.28) providing explicit expressions in terms of *in situ* density and boundary contributions. These expressions offer a huge simplification for how we determine pressure, with the remainder of this section providing example implications.

23.3.2 Evolution of hydrostatic pressure

We expect that hydrostatic pressure evolves according to the convergence of mass onto the column of fluid above that point. The ocean hydrostatic pressure also changes due to time changes in the applied upper boundary pressure. Here we derive mathematical expressions that support these expectations, with Figure 23.1 providing a schematic.

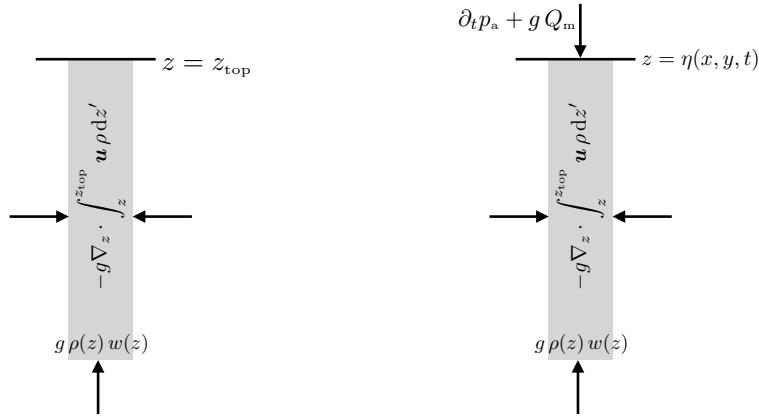


FIGURE 23.1: Evolution of hydrostatic pressure for a vertical position in the atmosphere (left panel) and ocean (right panel) according to equations (23.30b) and (23.33a). Hydrostatic pressure at a vertical position, z , (here the bottom of each column) arises from the convergence of mass onto the column over the region above z . The ocean column also has a contribution from the time tendency of applied surface pressure plus the mass of coming across the top boundary. For the atmosphere as assume the top boundary is at $z_{\text{top}} = \infty$ and so there is no mass coming across that boundary.

Hydrostatic pressure in the atmosphere

A time derivative of the atmospheric hydrostatic pressure expression (23.28) renders

$$\partial_t p = g \int_z^{z_{\text{top}}} \partial_t \rho(x, y, z', t) dz'. \quad (23.29)$$

Note the absence of a time derivative on z_{top} . We ensure this time derivative is not relevant by setting z_{top} to a constant value well above anything of physical relevance; e.g., $z_{\text{top}} \approx \infty$. Now insert the mass continuity equation (15.6) and make further use of Leibniz's rule to write

$$\partial_t p = -g \int_z^{z_{\text{top}}} \nabla \cdot (\mathbf{v} \rho) dz' \quad (23.30a)$$

$$= g \rho(z) w(z) - g \nabla_z \cdot \int_z^{z_{\text{top}}} \mathbf{u} \rho dz', \quad (23.30b)$$

where we set $\rho(z_{\text{top}}) = 0$ and assumed z_{top} is independent of horizontal position. The first term on the right hand side arises from the vertical mass flux into the fluid column from below. The second term arises from the horizontal convergence of mass over the column above the position, z .

Hydrostatic pressure in the ocean

The derivation for the ocean requires some more work since the ocean free surface is a permeable space-time dependent function. A time derivative of the ocean pressure expression (23.27) renders

$$\partial_t p = \partial_t p_a + g \rho(\eta) \partial_t \eta + g \int_z^{\eta} \partial_t \rho(x, y, z', t) dz', \quad (23.31)$$

where we made use of Leibniz's rule to take the time derivative of the upper limit at $z = \eta(x, y, t)$, and with the shorthand $\rho(\eta) = \rho(x, y, z = \eta, t)$. Now insert the mass continuity equation (15.6)

and make further use of Leibniz's rule to write

$$\partial_t(p - p_a) - g \rho(\eta) \partial_t \eta = g \int_z^\eta \partial_t \rho(x, y, z', t) dz' \quad (23.32a)$$

$$= -g \int_z^\eta \nabla \cdot (\mathbf{v} \rho) dz' \quad (23.32b)$$

$$= -g [\rho(\eta) w(\eta) - \rho(z) w(z)] - g \int_z^\eta \nabla_z \cdot (\mathbf{u} \rho) dz' \quad (23.32c)$$

$$= g \rho(z) w(z) + g [(-w + \mathbf{u} \cdot \nabla \eta) \rho]_{z=\eta} - g \nabla_z \cdot \int_z^\eta \mathbf{u} \rho dz' \quad (23.32d)$$

$$= g \rho(z) w(z) + g (Q_m - \rho(\eta) \partial_t \eta) - g \nabla_z \cdot \int_z^\eta \mathbf{u} \rho dz', \quad (23.32e)$$

where the last step made use of the kinematic boundary condition (15.79) for the ocean free surface, with Q_m the mass flux entering the ocean across the free surface. Rearrangement, and cancellation of the $\rho(\eta) \partial_t \eta$ term appearing on both sides, leads to

$$\partial_t p = \partial_t p_a + g \rho(z) w(z) + g Q_m - g \nabla \cdot \int_z^\eta \mathbf{u} \rho dz'. \quad (23.33a)$$

The first term on the right hand side arises from time fluctuations of the applied pressure at $z = \eta$. The second and third terms measure the vertical convergence of mass onto the column of fluid sitting above the vertical position, with $\rho(z) w(z)$ the mass flux entering the column from below and Q_m the mass flux entering from across the free surface. The final term arises from the vertically integrated horizontal mass transport converging onto the column above the position of interest.

23.3.3 Heuristic scaling

We here present a scale analysis to justify the hydrostatic approximation. This analysis serves to introduce a common method used in fluid mechanics to identify those processes that may be dominant for a particular flow regime. In particular, the flow regime of interest here occurs with a small vertical to horizontal aspect ratio

$$\alpha_{\text{aspect}} \equiv \frac{H}{L} \ll 1, \quad (23.34)$$

with H a typical length scale for vertical motion and L the horizontal length scale. This regime is fundamental to the large-scale circulation of the ocean and atmosphere. As the hydrostatic approximation is concerned with the force balances in a fluid column, it is sufficient to ignore rotation when performing a scale analysis.

Consider the vertical momentum equation (23.17) from the tangent plane and traditional approximations (Section 23.2), along with the associated scales for the various terms

$$\frac{Dw}{Dt} = -\frac{1}{\rho} \frac{\partial p}{\partial z} - g \quad (23.35a)$$

$$\frac{W}{T} + \frac{U W}{L} + \frac{W W}{H} = -\frac{1}{\rho} \frac{\partial p}{\partial z} - g. \quad (23.35b)$$

In the second equation we introduced the following scales for the terms appearing on the left hand side of the first equation.

- L is the horizontal scale of the motion.

- H is the vertical scale of the motion.
- W is the vertical velocity scale.
- U is the horizontal velocity scale. For this analysis we do not distinguish between the zonal and meridional velocity scales, writing U for both. This assumption is not always valid, such as when scaling for jet stream or equatorial flows, both of which have larger zonal speeds than meridional.
- T is the time scale of the motion. We assume that the time scale is determined by horizontal advection so that $T \sim L/U$. For studies of waves we may wish to instead consider time scales according to a wave speed and wave length, $T \sim \lambda/c$.

To get a sense for the numbers, consider the atmospheric case with $W = 10^{-2} \text{ m s}^{-1}$, $L = 10^5 \text{ m}$, $H = 10^3 \text{ m}$, $U = 10 \text{ m s}^{-1}$. These numbers lead to $T = L/U = 10^4 \text{ s}$ and to the values for the vertical momentum equation

$$10^{-6} \text{ m s}^{-2} \sim -\frac{1}{\rho} \frac{\partial p}{\partial z} - g. \quad (23.36)$$

With $g \sim 10 \text{ m s}^{-2}$, the only term that can balance the gravitational acceleration is the vertical pressure gradient. A similar analysis holds for the ocean where we set $W = 10^{-3} \text{ m s}^{-1}$, $L = 10^3 \text{ m}$, $H = 10^1 \text{ m}$, $U = 10^{-1} \text{ m s}^{-1}$. These numbers lead to $T = L/U = 10^4 \text{ s}$ and to

$$10^{-7} \text{ m s}^{-2} \sim -\frac{1}{\rho} \frac{\partial p}{\partial z} - g. \quad (23.37)$$

In either case, large scale motion maintains an approximate hydrostatic balance whereby $\partial p/\partial z = -\rho g$.

We offer a more formal scale analysis in Section 24.2, making use of the oceanic Boussinesq equations derived in Chapter 24. For the remainder of this section we explore certain properties of a fluid maintaining an approximate hydrostatic balance.

23.3.4 Removal of a dynamically irrelevant background state

The previous analysis pointed to the dominance of the hydrostatic balance in the vertical momentum equation for large scale motions. However, is that analysis sufficient to understand what causes motion? To help answer that question, consider a density field that is decomposed into a constant, ρ_0 , plus a deviation

$$\rho(\mathbf{x}, t) = \rho_0 + \delta\rho(\mathbf{x}, t), \quad (23.38)$$

with a corresponding decomposition of the pressure field

$$p(\mathbf{x}, t) = p_0(z) + p'(\mathbf{x}, t) \quad \text{with} \quad \frac{dp_0}{dz} = -\rho_0 g. \quad (23.39)$$

That is, the pressure is decomposed into a background static hydrostatic pressure field that is just a function of z , plus a deviation from the background pressure. In this case, the non-rotating vertical momentum equation takes the form

$$\rho \frac{Dw}{Dt} = -\frac{\partial p'}{\partial z} - \left[\frac{dp_0}{dz} + \rho_0 g \right] = -\frac{\partial p'}{\partial z}. \quad (23.40)$$

We thus see that the exact hydrostatically balanced background pressure, $p_0(z)$, has no dynamical implications. Correspondingly, to garner a more relevant scaling for the hydrostatic balance it is appropriate to ask whether the dynamically active pressure, p' , is approximately hydrostatic.

For flows with small aspect ratios, the vertical momentum equation remains approximately hydrostatic even when removing the dynamically inactive background pressure field. So our intuition about hydrostatic dominance holds unchanged even for the dynamical pressure. The formal justification of this approximation is nicely framed within the Boussinesq equations of Chapter 24 since the pressure force in these equations exposes just the dynamically active pressure. We thus postpone further discussion of hydrostatic scaling until Section 24.2.

23.3.5 Decomposing the horizontal pressure gradient

In contrast to an exact hydrostatic fluid, where there is no motion, there are generally horizontal pressure gradients in an approximate hydrostatic fluid, and these horizontal gradients drive horizontal motion. Such horizontal gradients in the hydrostatic pressure can arise from horizontal differences in the fluid mass density. We refer to such pressure gradients as *internal pressure gradients* since they arise from density gradients internal to the fluid. Horizontal pressure gradients can also arise from horizontal gradients in the total mass of a fluid column, with such pressure gradients referred to as *external pressure gradients*. In developing an understanding of the horizontal pressure gradients in an approximate hydrostatic fluid, it can be useful to examine the mathematical expressions for the pressure gradient. We do so in this section, considering the specific case of a column of ocean fluid.³

Recall equation (23.27), which expresses the hydrostatic pressure at a point within the ocean

$$p(z, y, z, t) = p_a(x, y, t) + g \int_z^\eta \rho(x, y, z', t) dz'. \quad (23.41)$$

In this equation, $p_a = p[x, y, z = \eta(x, y, t), t]$ is the pressure applied to the ocean free surface at $z = \eta(x, y, t)$ from any mass above the ocean, such as the atmosphere or cryosphere. In many idealized cases we assume the media above the ocean is massless, in which case $p_a = 0$. Nonetheless, we maintain $p_a \neq 0$ for generality. Now introduce the globally referenced buoyancy (see Chapter 25) as defined by

$$b = -g(\rho - \rho_0)/\rho_0, \quad (23.42)$$

in which case the pressure is

$$p = -g \rho_0 z + g \rho_0 [\eta + p_a/(g \rho_0)] - \rho_0 \int_z^\eta b dz'. \quad (23.43)$$

The first term is a background pressure that increases moving downward. However, this background pressure has no horizontal dependence and so it does not contribute to the horizontal pressure gradients. In contrast, the second and third terms have horizontal gradients and are thus sometimes referred to as the *dynamical pressure*. The second term arises from the free surface height plus the applied surface pressure. This term is uniformly felt throughout the fluid column since it has no vertical dependence. The free surface term is the product of the large number, $g \rho_0$, times a small free surface undulation, η . The third term arises from buoyancy within the fluid computed relative to the constant background density, ρ_0 , and this term is a function of vertical position.

³Note that some authors refer to *baroclinic* and *barotropic* pressure gradients rather than the *internal* and *external* used here.

Furthermore, it is the vertical integral over a generally large depth range of the relatively small buoyancy. In this manner, the second and third terms can generally be of comparable magnitude.

The horizontal gradient of the hydrostatic pressure (23.43) is given by

$$\nabla_z p = \underbrace{\nabla_z p_a + g \rho(\eta) \nabla_z \eta}_{\text{external contribution}} - \underbrace{\rho_0 \int_z^\eta \nabla_z b dz'}_{\text{internal contribution}}. \quad (23.44)$$

Again, the external contribution acts throughout the vertical fluid column since it is only a function of horizontal position and time. Every point within the fluid column instantly feels this term whenever there is a gradient in the applied surface pressure, the surface height, or the surface buoyancy, with $\rho_0[g - b(\eta)] = g \rho(\eta)$. The internal term arises from horizontal gradients in the *in situ* buoyancy that are integrated vertically over the region above the point of interest.

We can invert the above formulation by introducing the bottom pressure

$$p_b = p_a + g \int_{\eta_b}^z \rho dz, \quad (23.45)$$

in which case

$$p = p_b - g \rho_0 (z - \eta_b) + \rho_0 \int_{\eta_b}^z b dz', \quad (23.46)$$

so that the corresponding expression for the horizontal hydrostatic pressure gradient is

$$\nabla_z p = \underbrace{\nabla_z p_b + g \rho(\eta_b) \nabla_z \eta_b}_{\text{external contribution}} + \underbrace{\rho_0 \int_{\eta_b}^z \nabla_z b dz'}_{\text{internal contribution}}. \quad (23.47)$$

The bottom pressure contribution is dominated by gradients in the bottom topography so that it can be useful to write

$$\nabla_z p_b = -\rho_0 g \nabla_z \eta_b + \nabla_z p'_b, \quad (23.48)$$

in which case the external contribution takes the form

$$\nabla_z p = \underbrace{\nabla_z p'_b - \rho_0 b(\eta_b) \nabla_z \eta_b}_{\text{barotropic contribution}} + \underbrace{\rho_0 \int_{\eta_b}^z \nabla_z b dz'}_{\text{baroclinic contribution}}. \quad (23.49)$$

As for the expression (23.44), we have decomposed the horizontal pressure gradient into an external and internal contribution. Here, the external contributions arise from gradients in the bottom pressure, which measures the mass per area of fluid within the column, plus gradients in the bottom topography as weighted by the difference between the gravitational acceleration and bottom buoyancy, with $\rho_0[g - b(\eta_b)] = g \rho(\eta_b)$. The internal term arises from gradients in the buoyancy as integrated below the depth of interest.

23.3.6 Horizontal hydrostatic pressure gradient in a mass conserving fluid

We here consider a specific example that is emblematic of how one determines the sign for horizontal gradients in a fluid maintaining an approximate hydrostatic balance, and thus to determine the direction for the pressure force. In so doing, we further our understanding of the pressure field driving motion in a hydrostatic fluid. The example is posed for a bounded column of fluid, as in the ocean, but analogous considerations hold for the atmosphere where the upper boundary is the top of the atmosphere (i.e., effectively unbounded). Furthermore, the fluid is assumed to be compressible so that fluid elements conserved their mass. In Section 24.2.5 we reconsider this example for a volume conserving oceanic Boussinesq fluid.

Two columns with equal mass yet different densities

Consider two adjacent columns of seawater with equal mass but with distinct density; assume the density in each column is constant throughout the respective columns; and assume the atmospheric pressure has equal mass above the two water columns. Figure 23.2 offers a schematic, where we make the additional assumption that the two columns sit on a flat bottom. We can imagine setting up this configuration by starting with uniform density water, then warming the water in column B more than column A while maintaining constant mass in the two columns. This process sets up a horizontal density gradient with an associated horizontal gradient in the hydrostatic pressure. Furthermore, the less dense water in column B occupies more volume so that its free surface sits higher

$$\rho_B < \rho_A \implies \eta_B > \eta_A. \quad (23.50)$$

What is the sign of the horizontal pressure gradient? As we show in the following, column B (the low density column) has larger hydrostatic pressure than column A (the high density column) for every point in the column, except at the bottom where the two bottom pressures are identical since the two columns have equal mass.

Computing pressure starting from the equal bottom pressures

Since the two columns have equal mass and equal cross-sectional area, the hydrostatic pressures (weight per unit area) at the bottom of the two columns are equal and given by

$$p_{\text{bot}} = g \rho_A (\eta_A - \eta_b) = g \rho_B (\eta_B - \eta_b), \quad (23.51)$$

where $z = \eta_b(x, y)$ is the vertical position at the bottom, $z = \eta_A(x, y, t)$ is the top of column A, and $z = \eta_B(x, y, t)$ top of column B. Since the bottom pressures are identical, there is no horizontal pressure gradient at the bottom so that all pressure gradients exist above the bottom.

The hydrostatic pressure at an arbitrary position within column A is given by

$$p_A(z) = g \rho_A (\eta_A - z) = p_{\text{bot}} - g \rho_A (-\eta_b + z). \quad (23.52)$$

The second equality arose by substituting the bottom pressure from equation (23.51) to eliminate the surface height η_A . Likewise, the pressure in column B is given by

$$p_B(z) = g \rho_B (\eta_B - z) = p_{\text{bot}} - g \rho_B (-\eta_b + z). \quad (23.53)$$

We can now take the difference between the two hydrostatic pressures to find

$$p_B(z) - p_A(z) = g (-\eta_b + z) (\rho_A - \rho_B) > 0. \quad (23.54)$$

Since $\rho_A > \rho_B$ and $z \geq \eta_b$ we see that at any point above the bottom, the hydrostatic pressure in column B (the lighter column) is greater than that in column A (the denser column). This horizontal difference in the hydrostatic pressure renders a force pointing from column B to column A. Vertically integrating this pressure difference over the thickness of column A leads to the net force per horizontal length

$$F_{\text{pressure B to A}} = \int_{\eta_b}^{\eta_A} [p_B(z) - p_A(z)] dz = (g/2) (\rho_A - \rho_B) (\eta_A - \eta_b)^2 > 0. \quad (23.55)$$

Inferring pressure gradients starting from the top

Another way to understand why the pressure force points from column B to column A is to note that at the top of both columns the pressures are the same (and equal to the uniform atmospheric pressure). However, since column B sits higher than column A , as we move down from $z = \eta_B$ the pressure increases in column B immediately, whereas the pressure in column A remains at the atmospheric pressure until entering the water column at $z = \eta_A < \eta_B$. So it is clear that the pressure in column B is greater than A starting from the surface and moving down. And since the two bottom pressures are equal, then one can infer the pressure isolines as drawn in Figure 23.2.

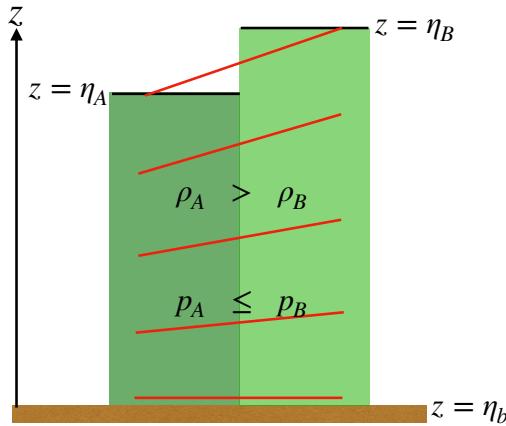


FIGURE 23.2: Two seawater columns on a flat bottom with equal mass but different densities with $\rho_A > \rho_B$. We assume the atmosphere above the columns has the same pressure over both columns, thus offering zero horizontal pressure force. Furthermore, the cross-sectional area of the two columns are the same so that the less dense water in column B has more volume and thus a greater thickness: $\eta_B > \eta_A$. Since the column masses are the same, the hydrostatic pressures (weight per horizontal area) at the bottom of the two columns are equal: $p_A(z = \eta_b) = p_B(z = \eta_b) = p_{\text{bot}}$. In oceanographic parlance, the bottom offers a “level of no motion” from which to reference the pressure field. However, at any position z above the bottom, equation (23.54) shows that the hydrostatic pressure in column B is greater than A : $p_B(z) - p_A(z) = g(-\eta_b + z)(\rho_A - \rho_B) > 0$. The horizontal gradient in hydrostatic pressure thus points from column B towards column A . The red lines show lines of constant pressure (isobars), which are horizontal next to the bottom but which slope upward to the right moving towards the surface. This configuration provides salient points about hydrostatic pressure relevant for the slightly more complex reduced gravity example in Figure 31.5. Also, it is useful to compare this schematic to Figure 26.4, which discusses the depth dependence of the horizontal gradient in hydrostatic pressure as per $\partial(\nabla_z p)/\partial z = -g \nabla_z \rho$.

23.3.7 Ocean dynamic topography

There are occasions in oceanography where it is useful to study the thickness of a layer bounded by isobars. For example, it is often assumed in dynamical oceanography that there is a pressure at which baroclinic currents vanish (*level of no motion*), with the reduced gravity model of Section 31.3 studying the associated dynamical implications.⁴ We are thus led to consider the thickness of fluid extending from the ocean free surface to a chosen pressure level in the ocean interior as given by

$$D(p) = \eta - z(p). \quad (23.56)$$

⁴ Baroclinic refers to a fluid state with pressure and density surfaces distinct, whereas these surfaces are parallel in a *barotropic* fluid state. We discuss *baroclinicity* in the context of vorticity in Section 36.5.

Assuming a hydrostatic balance for each fluid column allows us to relate this expression to the vertical integral between two pressure surfaces of the specific volume, ρ^{-1}

$$D(p) = \int_{z(p)}^{\eta} dz = g^{-1} \int_{p_a}^p \frac{dp'}{\rho}, \quad (23.57)$$

where the second step used the hydrostatic balance and absorbed a minus sign by swapping integral limits. We refer to the thickness $D(p)$ as the *dynamic topography* with respect to a reference pressure p . Note that it is sometimes also called the *steric sea level*. Evolution of the dynamic topography arises from changes in the pressure applied to the free surface as well as changes in the specific volume

$$g \frac{\partial D(p)}{\partial t} = -\frac{1}{\rho(\eta)} \frac{\partial p_a}{\partial t} + \int_{p_a}^p \frac{\partial \rho^{-1}}{\partial t} dp. \quad (23.58)$$

Note that the time derivative acts on the specific volume when computed on surfaces of constant pressure. If the depth $z(p)$ of the constant pressure surface is static, then the evolution of layer thickness, $D(p)$, is identical to the sea surface, η . In general, there is no such static pressure level, thus making the time tendencies differ, though certain situations warrant this approximation.

23.3.8 Surfaces of atmospheric geopotential height and pressure

In Section 21.9.10 we computed the geopotential height within an exact hydrostatic and ideal gas atmosphere. We here apply those results to the case of approximate hydrostatic and ideal gas columns, making use of equation (21.149) for the difference in geopotential height between two isobars

$$Z_2 - Z_1 = -(R^{\text{atm}} \langle T \rangle / g) \ln(p_2/p_1). \quad (23.59)$$

In this equation, $\langle T \rangle$ is the mean temperature within the column as computed according equation (21.147), and R^{atm} is the specific gas constant for air given by equation (21.118). The geopotential thickness of a column is positive when the isobars have $p_2 < p_1$; i.e., p_2 is higher in the atmosphere than p_1 . Furthermore, the geopotential thickness of a column is directly proportional to the column mean temperature so that a warmer column is thicker. This result is expected since for a given mass of air, a warmer column is less dense and so isobars are higher over warmer hydrostatic air columns than cooler columns. Correspondingly, when moving horizontally along a constant geopotential surface, we encounter higher pressure when moving into a region of warmer air. This situation is entirely analogous to that in Figure 23.2 when studying the horizontal pressure difference between two hydrostatic and equal mass columns of seawater.

23.3.9 Concerning vertical motion

Unbalanced vertical accelerations still exist in an approximate hydrostatic fluid. Yet these vertical accelerations are not needed for the prognostic equations since the vertical momentum equation reduces to a local hydrostatic balance. Hence, rather than compute vertical motion prognostically, vertical motion in a hydrostatic fluid is diagnosed through constraints on the fluid motion.

For example, an important constraint on the large-scale ocean circulation arises from vorticity and potential vorticity conservation, which are topics considered in Part VI of this book. Mass continuity discussed in Section 24.1.4 provides another constraint. In particular, horizontal velocity divergence in an incompressible flow is balanced by vertical velocity convergence as per equation (24.17). The vertical pressure forces required to produce the vertical motion are those precisely needed to maintain volume continuity. In a hydrostatic fluid, we do not directly compute these

forces for the purpose of prognosing vertical accelerations. Rather, the vertical acceleration is inferred through the kinematic constraint. The associated forces, if desired, can be diagnosed given the velocity and the accelerations.

23.3.10 Further study

Section 2.7.4 in [Vallis \(2017\)](#) provides examples of scales over which the hydrostatic relation remains a useful approximation in geophysical fluids. Our discussion of the hydrostatic pressure in Section 23.3.6 is motivated by similar considerations presented in Chapter 2 of the descriptive oceanography text from [Tomczak and Godfrey \(1994\)](#). Further discussion of dynamic topography is given in Appendix B.4 of [Griffies et al. \(2014\)](#) as well as in [Tomczak and Godfrey \(1994\)](#). Finally, this 8-minute video from Prof. Hogg offers an introduction to hydrostatic pressure.

23.4 Homogeneous fluid in a rotating tank

As an application of the ideas developed in this chapter and in earlier chapters, we develop the equations for a homogeneous fluid in a rotating tank such as occurs in laboratory studies of rotating fluids. One point of departure from planetary applications concerns the choice of vertical coordinate. Recall we introduced geopotential surfaces in Section 10.12.4, on which the effective gravitational force (sum of central gravity plus planetary centrifugal) is constant. Correspondingly, we introduced geopotential coordinates in Section 10.13.3 to simplify the equations for planetary fluid dynamics. In contrast, for the rotating tank we do not make use of geopotential coordinates. Instead, we expose the centrifugal acceleration (due to rotation of the tank), which allows for a clear display of the parabolic shape for the free surface when the fluid is in solid-body motion.

23.4.1 What about the planet's rotation?

Do we need to worry about the planet's rotation? To answer this question, consider a typical record player with an angular speed of 45 revolutions per minute

$$\Omega_{\text{record}} = 0.75 \text{ s}^{-1}. \quad (23.60)$$

This angular speed is roughly 10^4 times faster than the earth's angular speed of $7.29 \times 10^{-5} \text{ s}^{-1}$ (equation (10.20)). For a tank rotating at a rate on the same order as a record player, we are justified ignoring the rotating earth in comparison to the rotating tank. That is, we can safely ignore planetary Coriolis and planetary centrifugal accelerations, allowing us to instead focus on non-inertial accelerations just from the rotating tank.

23.4.2 Formulating the equations of motion

In an inertial reference frame, a fluid element feels the gravitational force, pressure force, and friction, thus leading to the Cartesian coordinate equations of motion

$$\frac{Du_1}{Dt} = -\frac{1}{\rho} \frac{\partial p}{\partial x} + F^x \quad (23.61)$$

$$\frac{Dv_1}{Dt} = -\frac{1}{\rho} \frac{\partial p}{\partial y} + F^y \quad (23.62)$$

$$\frac{Dw_1}{Dt} = -\frac{1}{\rho} \frac{\partial p}{\partial z} - g_e + F^z, \quad (23.63)$$

where \mathbf{v}_I is the inertial velocity, ρ is the constant density, and we orient the coordinates so that the z -axis extends vertically upward from the center of the tank and parallel to the gravity acceleration. Correspondingly, the rotation vector for the tank is

$$\boldsymbol{\Omega} = \Omega \hat{\mathbf{z}} = (f/2) \hat{\mathbf{z}}. \quad (23.64)$$

To derive the rotating frame equations, return to some of the kinematics from Chapter 10, in which we write the position of a fluid particle as

$$\mathbf{X}(t) = X \hat{\mathbf{x}} + Y \hat{\mathbf{y}} + Z \hat{\mathbf{z}}. \quad (23.65)$$

We assume that the Cartesian unit vectors are fixed in the rotating frame and thus move as a solid body with the rotating tank. The inertial velocity is thus given by

$$\frac{d\mathbf{X}}{dt} = \left[\frac{dX}{dt} \right] \hat{\mathbf{x}} + \left[\frac{dY}{dt} \right] \hat{\mathbf{y}} + \left[\frac{dZ}{dt} \right] \hat{\mathbf{z}} + \boldsymbol{\Omega} \wedge \mathbf{X}. \quad (23.66)$$

Correspondingly, the acceleration is given by

$$\frac{d^2\mathbf{X}}{dt^2} = \left[\frac{d^2X}{dt^2} \right] \hat{\mathbf{x}} + \left[\frac{d^2Y}{dt^2} \right] \hat{\mathbf{y}} + \left[\frac{d^2Z}{dt^2} \right] \hat{\mathbf{z}} + 2\boldsymbol{\Omega} \wedge \mathbf{v} + \boldsymbol{\Omega} \wedge (\boldsymbol{\Omega} \wedge \mathbf{X}), \quad (23.67)$$

where we defined the rotating frame Cartesian velocity as

$$\mathbf{v} = \left[\frac{dx}{dt} \right] \hat{\mathbf{x}} + \left[\frac{dy}{dt} \right] \hat{\mathbf{y}} + \left[\frac{dz}{dt} \right] \hat{\mathbf{z}}. \quad (23.68)$$

Setting the inertial acceleration equal to the inertial force per mass leads to the equations of motion in the rotating frame

$$\frac{Du}{Dt} - 2\Omega v = -\frac{1}{\rho} \frac{\partial p}{\partial x} + \Omega^2 x + F^x \quad (23.69)$$

$$\frac{Dv}{Dt} + 2\Omega u = -\frac{1}{\rho} \frac{\partial p}{\partial y} + \Omega^2 y + F^y \quad (23.70)$$

$$\frac{Dw}{Dt} = -\frac{1}{\rho} \frac{\partial p}{\partial z} - g_e + F^z, \quad (23.71)$$

which take on the vector form

$$\frac{D\mathbf{v}}{Dt} + f \hat{\mathbf{z}} \wedge \mathbf{v} = -\nabla [p/\rho + g_e z - \Omega^2 (x^2 + y^2)/2] + \mathbf{F}. \quad (23.72)$$

As expected, we encounter both a Coriolis and centrifugal acceleration due to the rotation of the tank.

23.4.3 Solid-body rotation and free surface shape

Consider a fluid at rest in a non-rotating tank, and then start the tank rotating. As in our discussion of Couette flow in Section 19.7.2, viscous effects transfer motion from the outside tank wall (where a no-slip boundary condition makes the fluid move with the wall) into the interior of the fluid. Given sufficient time and a constant rotation rate, the fluid will move as a solid-body. As an application of the above equations of motion, we here determine the shape of the upper free surface for this solid-body motion, offering two related derivations. Note that when the fluid reaches solid-body motion, all strains vanish within the fluid so that frictional stresses vanish (see Section 19.7). Hence, the steady force balance is fully inviscid although the steady state required viscosity to reach it.

Component equations of motion

The velocity and acceleration in the rotating frame are zero when the fluid is in solid-body rotation. The vertical momentum equation (23.71) thus reduces to hydrostatic balance

$$\frac{\partial p}{\partial z} = -\rho g_e. \quad (23.73)$$

In general we do not have hydrostatic balance for motion in a tank that deviates from solid-body. However, when that motion is close to a solid-body rotation, then the fluid is in an approximate hydrostatic balance. As seen in Section 23.3, this situation corresponds to the large-scale ocean and atmosphere.

Hydrostatic balance with a constant density means that the pressure is a linear function of depth

$$p(x, y, z) = \rho g_e (\eta - z), \quad (23.74)$$

where $z = \eta(x, y)$ is the vertical position of the free surface. The horizontal momentum equations (23.69)-(23.70) reduce to a balance between the pressure gradient and centrifugal accelerations

$$\frac{\partial p}{\partial x} = \rho x \Omega^2 \quad \text{and} \quad \frac{\partial p}{\partial y} = \rho y \Omega^2. \quad (23.75)$$

Pressure thus increases when moving radially away from the center. Substituting in the pressure as given by the hydrostatic relation (23.74) leads to relations satisfied by the solid-body free surface

$$g_e \frac{\partial \eta}{\partial x} = x \Omega^2 \quad \text{and} \quad g_e \frac{\partial \eta}{\partial y} = y \Omega^2. \quad (23.76)$$

Integration leads to the quadratic expression for the free surface

$$\eta = \eta(0) + \frac{\Omega^2 (x^2 + y^2)}{2 g_e}, \quad (23.77)$$

where $\eta(0)$ is the free surface at the center of the tank where $x = y = 0$. The solid-body rotating fluid thus has a quadratic free surface with the height of the surface increasing away from the center. Notice how the fluid density dropped out from the problem, so that this parabolic shape holds for any fluid in solid-body motion.

Vector force balance

A more telescopic means to determine the free surface shape is to set the forces to zero on the right hand side of the vector equation of motion (23.72) so that

$$p/\rho + g_e z - \Omega^2 (x^2 + y^2)/2 = p_0/\rho, \quad (23.78)$$

where p_0 is a constant pressure to be specified below. Furthermore, we set friction to zero since the fluid is in solid-body motion. Everywhere along the free surface, with $z = \eta$, the pressure equals to that applied to the free surface by the overlying media, $p = p_a$ (e.g., atmospheric pressure). Hence, setting $z = \eta$ in equation (23.78) and solving for η yields

$$\eta = \frac{p_0 - p_a}{\rho g_e} + \frac{\Omega^2 (x^2 + y^2)}{2 g_e}. \quad (23.79)$$

For simplicity, assume the applied pressure is spatially constant. Hence, setting p_0 according to the free surface at $x = y = 0$ brings the free surface to the parabolic form in equation (23.77)

$$\frac{p_0 - p_a}{\rho g_e} = \eta(0) \implies \eta = \eta(0) + \frac{\Omega^2 (x^2 + y^2)}{2 g_e}. \quad (23.80)$$

23.4.4 Further study

We study the angular momentum for the shallow water version of this system in Section 32.7. Refer to Section 6.6.4 of *Marshall and Plumb* (2008) for more discussion of laboratory rotating tank experiments.

23.5 Exercises

EXERCISE 23.1: PRIMITIVE EQUATIONS AND AXIAL ANGULAR MOMENTUM

The axial angular momentum of a fluid element satisfying the primitive equations is given by

$$L^z = (\rho \delta V) R_{\perp} (u + R_{\perp} \Omega) \equiv (\rho \delta V) l^z \quad (23.81)$$

where

$$R_{\perp} = R_e \cos \phi \quad (23.82)$$

is the distance from the polar rotation axis to a point on the sphere with radius R_e , and

$$l^z = R_{\perp} (u + R_{\perp} \Omega) \quad (23.83)$$

is the angular momentum per unit mass.

- (a) Consider a constant mass fluid element in the absence of friction. Show that the primitive equation zonal momentum equation (23.7) implies that the material evolution of axial angular momentum per mass is given by

$$\frac{Dl^z}{Dt} = -\frac{1}{\rho} \frac{\partial p}{\partial \lambda}. \quad (23.84)$$

- (b) Assume the zonal pressure gradient vanishes. Move the fluid element vertically while maintaining a fixed latitude. What happens to the zonal momentum of this primitive equation fluid element? Hint: be sure to remain within the “world” of the primitive equations.
- (c) Give a very brief symmetry argument for why the axial angular momentum is materially conserved when $\partial p / \partial \lambda = 0$. Hint: recall the discussion of Noether’s Theorem in Section 11.2.1.
- (d) Consider the material evolution of primitive equation axial angular momentum per mass in the case where the zonal momentum equation retains the unapproximated form of the Coriolis acceleration. Discuss the resulting material evolution equation. Does this equation make sense based on the symmetry argument given in the previous part of this exercise?

EXERCISE 23.2: MASS BALANCE FOR A HYDROSTATIC OCEAN COLUMN

Return to Exercise 15.2. Show that for a hydrostatic fluid the mass balance for a fluid column (equation (15.99)) takes the form

$$\frac{1}{g} \frac{\partial (p_b - p_a)}{\partial t} = -\nabla \cdot \left[\int_{\eta_b}^{\eta} \mathbf{u} \rho dz \right] + Q_m, \quad (23.85)$$

where

$$p_b = p_a + g \int_{\eta_b}^{\eta} \rho dz \quad (23.86)$$

is the hydrostatic pressure at the ocean bottom, and p_a is the pressure applied to the ocean surface from the overlying atmosphere or sea ice. Hint: this exercise simply involves the substitution of equation (23.86) into (15.99).



Oceanic Boussinesq fluid

In some areas of fluid mechanics the speed of fluid particles can approach the speed of sound. In that case, the *Mach number*, which is the ratio of the fluid speed to the sound speed, reaches unity or larger.¹ In such flows it is important to consider the influence of compressibility since the large flow speeds can lead to nontrivial local density changes through the convergence of advective fluxes. Geophysical flows of interest in this book have Mach numbers well below unity. Nonetheless, even with a small Mach number the impacts of pressure can play a nontrivial role in affecting density changes. Such pressure effects are particularly important in a gas such as the atmosphere, in which case realistic studies of atmospheric flow generally maintain full compressibility.

In contrast, the flow in many liquids is close to incompressible whereby the velocity is well approximated as non-divergent thus allowing for the kinematics from Chapter 17. Even so, a fully incompressible fluid is one where the density is uniform throughout, with this approximation not suitable for the ocean when aiming to capture physical processes driving large-scale ocean circulation. In this chapter we study how the *oceanic Boussinesq approximation* finds a middle ground by allowing for density changes while maintaining the kinematics of an incompressible flow. In essence, the oceanic Boussinesq approximation results in a quasi-compressible fluid, so that density can vary in space and time, and yet whose flow field is non-divergent and thus incompressible. In part, it achieves this middle ground by measuring the mass of a fluid element by multiplying the volume of the element times a constant reference density, ρ_0 , whereas the weight of the fluid element is computed with the space-time dependent *in situ* density, $\rho(\mathbf{x}, t)$.

READER'S GUIDE TO THIS CHAPTER

The oceanic Boussinesq approximation comprises a quasi-compressible fluid in which density materially changes, yet with an incompressible flow whereby the prognostic velocity is non-divergent. In this chapter we derive the oceanic Boussinesq approximation and explore many of its physical properties including energetics. As part of this discussion we study the available potential energy, which measures that portion of the potential energy available for converting to kinetic energy. Available potential energy is a concept that holds for compressible non-Boussinesq fluids, and yet we find it most convenient to study it within the Boussinesq fluid.

The oceanic Boussinesq fluid has a much broader application than just to the ocean since many of the physical and mathematical features also hold for atmospheric flow satisfying the anelastic approximation (see Section 2.5 of [Vallis \(2017\)](#)). Furthermore, the oceanic Boussinesq

¹When a jet airplane or rocket moves at a speed greater than Mach one, such super-sonic motion generates a spectacularly loud and powerful sonic boom.

fluid provides the starting point for many of the geophysical fluid models found later in this book.

24.1	The oceanic Boussinesq approximation	512
24.1.1	Isolating the dynamically active pressure field	513
24.1.2	Boussinesq momentum equation	514
24.1.3	A vorticity motivation to set ρ_0 constant	515
24.1.4	Mass continuity	515
24.1.5	Dependence on reference density	516
24.1.6	Version I of the Boussinesq equations	516
24.1.7	Version II of the oceanic Boussinesq equations	517
24.1.8	Summary points	518
24.1.9	Further study	520
24.2	Scaling for the hydrostatic approximation	520
24.2.1	Stratified non-rotating Boussinesq equations	520
24.2.2	Non-dimensionalization	521
24.2.3	Specifying scales	521
24.2.4	The role of the Froude number	522
24.2.5	Horizontal hydrostatic pressure gradient	523
24.2.6	Evolution of hydrostatic pressure	524
24.3	How pressure enforces non-divergent flow	524
24.3.1	Poisson equation for pressure	525
24.3.2	Characterizing the pressure sources	525
24.3.3	Comments and further study	527
24.4	Tracer budgets in Eulerian regions	527
24.4.1	Formulating the budget equation	528
24.4.2	Interpreting advective tracer contributions	528
24.4.3	A rectangular region example	529
24.4.4	Comments	530
24.5	Subgrid scale processes and boundary conditions	530
24.5.1	Material evolution of buoyancy	530
24.5.2	Mechanical forcing and dissipation	531
24.6	Mechanical energy analysis part I	532
24.6.1	Kinetic energy	532
24.6.2	Gravitational and mechanical energies	533
24.7	Mechanical energy analysis part II	534
24.7.1	Boussinesq dynamic enthalpy	534
24.7.2	Regionally integrated Boussinesq dynamic enthalpy	536
24.7.3	Comments	537
24.8	Available potential energy	538
24.8.1	Analytic continuation of buoyancy surfaces	539
24.8.2	The dual relation between height and buoyancy	542
24.8.3	Exact expression for APE	543
24.8.4	Approximate expression for APE	544
24.8.5	Comments	546
24.9	Exercises	546

24.1 The oceanic Boussinesq approximation

For the ocean, density deviates no more than a few percent relative to the volume mean density. Although small, ocean density deviations act over large distances and are crucial for driving

large-scale circulations. Such *thermohaline circulations* derive their driving force from variations in temperature and salinity that affect density and, in turn, modify the pressure. A key reason that small density changes can be so pivotal is that the density variations are multiplied by the relatively large gravitational acceleration when computing pressure. The oceanic Boussinesq approximation provides a systematic means to ignore small density deviations where it is safe to do so dynamically, while retaining density variations where they are critical such as when multiplied by gravity. In brief, the oceanic Boussinesq approximation makes use of (a slightly modified) compressible thermodynamics plus an incompressible kinematics. The use of compressible thermodynamics allows for thermohaline processes to modify density and thus pressure, while the incompressible kinematics removes sound waves and renders the volume of a fluid element materially invariant in the fluid interior.

24.1.1 Isolating the dynamically active pressure field

Pressure in a vertically stratified fluid can be decomposed into a static background hydrostatic pressure plus a deviation from the background pressure. We made use of this decomposition in Section 23.3.4 when developing the scaling for the hydrostatic approximation. However, the decomposition holds even when the fluid is non-hydrostatic. We consider the background pressure to be a function just of depth and as such it is determined by a static and horizontally homogeneous background density field. We are motivated to introduce this decomposition given that the background hydrostatic pressure field (again, it is just a function of depth) is dynamically inactive (we show this below). It is thus convenient to remove it from the dynamical equations. This decomposition is exact and motivated by the desire to isolate the dynamically active part of the equations of motion.

To achieve the pressure decomposition, start by decomposing density according to

$$\rho(\mathbf{x}, t) = \rho_0(z) + \rho'(\mathbf{x}, t) \quad (24.1)$$

where the deviation density is much smaller than the reference density

$$\rho' \ll \rho_0. \quad (24.2)$$

The following formulation for the momentum equation holds for the general case of $\rho_0(z)$. However, we note in Section 24.1.2 that setting $\rho_0(z) = \rho_0$ is motivated for studies of potential vorticity in Boussinesq fluids. Indeed, a space-time constant reference density is generally synonymous with the oceanic Boussinesq approximation.²

The corresponding decomposition of pressure is given by the sum of a static and depth dependent background pressure, $p_0(z)$, and a deviation pressure, $p'(\mathbf{x}, t)$,

$$p(\mathbf{x}, t) = p_0(z) + p'(\mathbf{x}, t). \quad (24.3)$$

The background pressure is assumed to be in hydrostatic balance with the reference density

$$\frac{dp_0}{dz} = -\rho_0 g, \quad (24.4)$$

with p_0 and ρ_0 both static. We offer the following points to clarify the decomposition of pressure in equation (24.3).

²Exercise 24.3 offers a modest means to generalize this assumption.

- Assuming the background pressure, $p_0(z)$, to be hydrostatic does *not* imply that the full pressure, $p(\mathbf{x}, t)$, is also hydrostatic. Rather, the decomposition merely serves to remove that portion of the pressure field that plays no role in establishing motion (we see this property below). So this decomposition holds whether the full pressure is approximately hydrostatic or fully non-hydrostatic.
- Furthermore, if $p(\mathbf{x}, t)$ is in an approximate hydrostatic balance (Section 23.3), the decomposition (24.3) does *not* remove all of the hydrostatic pressure from $p(\mathbf{x}, t)$. Rather, $p'(\mathbf{x}, t)$ is generally nonzero whether $p(\mathbf{x}, t)$ is in an approximate hydrostatic balance or fully non-hydrostatic.

With the above density and pressure decompositions, the momentum equation

$$\rho \left[\frac{D\mathbf{v}}{Dt} + 2\boldsymbol{\Omega} \wedge \mathbf{v} \right] = -\nabla p - \hat{\mathbf{z}} g \rho \quad (24.5)$$

takes the equivalent form

$$(\rho_0 + \rho') \left[\frac{D\mathbf{v}}{Dt} + 2\boldsymbol{\Omega} \wedge \mathbf{v} \right] = -\nabla p' - g \rho' \hat{\mathbf{z}} - \left[\frac{dp_0}{dz} + \rho_0 g \right] \hat{\mathbf{z}} \quad (24.6a)$$

$$= -\nabla p' - g \rho' \hat{\mathbf{z}}, \quad (24.6b)$$

where we used the hydrostatic balance (24.4) for the second equality. We thus see that the background hydrostatic pressure, $p_0(z)$, leads to no motion since it drops out from the momentum equation. Hence, it is a “do nothing” portion of the pressure field. The pressure force is thus determined solely by the gradient of p' , which serves as the dynamically relevant portion of the pressure field.

24.1.2 Boussinesq momentum equation

To develop the Boussinesq momentum equation, divide the momentum equation (24.6b) by the density, $\rho = \rho_0 + \rho'$, and write the pressure and gravity terms as

$$\frac{\nabla p' + g \rho' \hat{\mathbf{z}}}{\rho_0 + \rho'} = \frac{\nabla p' + g \rho' \hat{\mathbf{z}}}{\rho_0 + \rho'} \left[\frac{\rho_0 - \rho'}{\rho_0 - \rho'} \right] = \frac{\nabla p' + g \rho' \hat{\mathbf{z}}}{\rho_0^2 - (\rho')^2} (\rho_0 - \rho') \approx \frac{\nabla p' + g \rho' \hat{\mathbf{z}}}{\rho_0}, \quad (24.7)$$

where the final approximation results from dropping all terms that are second order in deviation quantities. This approximation then leads to the Boussinesq momentum equation

$$\frac{D\mathbf{v}}{Dt} + 2\boldsymbol{\Omega} \wedge \mathbf{v} = -\frac{1}{\rho_0} \nabla p' + b \hat{\mathbf{z}}, \quad (24.8)$$

where we introduced the globally referenced *Archimedean buoyancy* (Chapter 25) as defined relative to the constant background density

$$b = -\frac{g \rho'}{\rho_0} = -\frac{g (\rho - \rho_0)}{\rho_0}. \quad (24.9)$$

Hence, the globally referenced Archimedean buoyancy is positive when the *in situ* density is less than the reference density so that $\rho' = \rho - \rho_0 < 0$. That is, $b > 0$ when the fluid element is lighter (more buoyant) than the background reference density. Buoyancy is the product of the gravitational acceleration, which is a relatively large term, and the small number ρ'/ρ_0 . Their

product is generally not small so that their product cannot be neglected from the momentum equation.

In the special case of a space-time constant reference density, $\rho_0(z) = \rho_0$, it is convenient to introduce the shorthand for the deviation pressure normalized by the reference density

$$\varphi = \frac{p'}{\rho_0} = \frac{p - p_0(z)}{\rho_0}. \quad (24.10)$$

In this case the Boussinesq momentum equation takes the form

$$\frac{D\mathbf{v}}{Dt} + 2\boldsymbol{\Omega} \wedge \mathbf{v} = -\nabla\varphi + b\hat{\mathbf{z}}. \quad (24.11)$$

24.1.3 A vorticity motivation to set ρ_0 constant

We now anticipate a later discussion of vorticity and potential vorticity to motivate setting $\rho_0(z)$ to a constant. This paragraph is not critical for the remainder of this chapter, but worth returning to after studying vorticity. For that purpose, we note that the form of the pressure gradient acceleration found in equation (24.11) is particularly useful given that the curl of the right hand side eliminates pressure from the vorticity equation (Section 36.8.1). In contrast, for the more general form with $\rho_0(z)$ in equation (24.8), the Boussinesq baroclinicity vector (equation (36.100)) has a contribution from both pressure and buoyancy (we derive this equation in Exercise 36.1)

$$\mathbf{B} = \nabla \left[b - \frac{p'}{\rho_0^2} \frac{d\rho_0}{dz} \right] \wedge \hat{\mathbf{z}}. \quad (24.12)$$

The additional pressure contribution complicates the development of potential vorticity whereby we wish to have $\mathbf{B} \cdot \nabla b = 0$ (see Section 38.5.1). We are thus motivated to use a space-time constant reference density so that $d\rho_0/dz = 0$. Following this motivation, we generally assume ρ_0 is a constant in this book. Even so, in Exercise 24.3 we show how to split the difference by defining a slightly more general buoyancy field while retaining a constant ρ_0 .

24.1.4 Mass continuity

When decomposing density according to equation (24.1), the mass continuity equation

$$\frac{D\rho}{Dt} = -\rho \nabla \cdot \mathbf{v}, \quad (24.13)$$

takes the form

$$\frac{D\rho'}{Dt} = -(\rho_0 + \rho') (\nabla_z \cdot \mathbf{u} + \partial_z w). \quad (24.14)$$

For many geophysical flows, the material time derivative on the left hand side is much smaller than either of the two terms appearing on the right hand side. To help formalize this observation it is useful to introduce a time scale for the various terms in this equation

$$\left| \frac{1}{\rho} \frac{D\rho'}{Dt} \right| \sim \frac{1}{T_\rho} \quad |\partial_x u| \sim T_u^{-1} \quad |\partial_y v| \sim T_v^{-1} \quad |\partial_z w| \sim T_w^{-1}. \quad (24.15)$$

Quite often we find flows in which the time scales associated with the spatial deformations of the flow, in the direction of the flow, are much smaller than time scales for the material changes in density, whereby

$$T_u^{-1}, T_v^{-1}, T_w^{-1} \gg T_\rho^{-1}. \quad (24.16)$$

In this case the only way for the mass balance equation (24.14) to hold is for the three terms contributing to the divergence to balance one another

$$\partial_x u + \partial_y v + \partial_z w \approx 0. \quad (24.17)$$

Note that for stratified flows we generally find the horizontal divergence of the horizontal velocity balancing the vertical convergence of the vertical velocity

$$\nabla_z \cdot \mathbf{u} \approx -\partial_z w. \quad (24.18)$$

Taking this balance to the limit motivates setting the velocity field for a Boussinesq fluid to be non-divergent

$$\nabla \cdot \mathbf{v} = 0. \quad (24.19)$$

24.1.5 Dependence on reference density

How does the solution compare when considering two distinct reference densities, say $\rho_0 \neq \rho_1$? To answer this question, write the inviscid Boussinesq velocity equation in the form

$$(\partial_t + \mathbf{v}_0 \cdot \nabla) \mathbf{v}_0 + 2\boldsymbol{\Omega} \wedge \mathbf{v}_0 = -\nabla \varphi_0 + b_0 \hat{\mathbf{z}} \quad (24.20a)$$

$$(\partial_t + \mathbf{v}_1 \cdot \nabla) \mathbf{v}_1 + 2\boldsymbol{\Omega} \wedge \mathbf{v}_1 = -\nabla \varphi_1 + b_1 \hat{\mathbf{z}}. \quad (24.20b)$$

Consider an initial condition in which

$$\rho_0 \mathbf{v}_0 = \rho_1 \mathbf{v}_1, \quad (24.21)$$

and with pressure and buoyancy initialized so that

$$\rho_0 (-\nabla \varphi_0 + b_0 \hat{\mathbf{z}}) = \rho_1 (-\nabla \varphi_1 + b_1 \hat{\mathbf{z}}) = -\nabla p - g \rho \hat{\mathbf{z}}. \quad (24.22)$$

Hence, the difference between equations (24.20a) and (24.20b) takes the form

$$\partial_t (\rho_0 \mathbf{v}_0 - \rho_1 \mathbf{v}_1) = -(\rho_0 \mathbf{v}_0 \cdot \nabla) \mathbf{v}_0 + (\rho_1 \mathbf{v}_1 \cdot \nabla) \mathbf{v}_1 \quad (24.23a)$$

$$= -[(\rho_1 - \rho_0)/\rho_0] (\mathbf{v}_1 \cdot \nabla) \mathbf{v}_1. \quad (24.23b)$$

The difference in reference densities allows for the nonlinear self-advection to evolve the difference $\rho_0 \mathbf{v}_0 - \rho_1 \mathbf{v}_1$. This property of the Boussinesq equations must be kept in mind if comparing numerical solutions using distinct reference densities. Namely, oceanic Boussinesq fluid is dependent on the reference density through the nonlinear advection term.

24.1.6 Version I of the Boussinesq equations

The first form of the oceanic Boussinesq equations emphasizes the role of buoyancy computed relative to a reference state of constant density, $\rho = \rho_0$. This form facilitates a focus on that portion of the pressure field giving rise to internal or baroclinic pressure gradients; i.e., those pressure gradients that generate motion independent of free surface undulations. The oceanic

Boussinesq equations thus take the form

$$\frac{D\mathbf{v}}{Dt} + 2\boldsymbol{\Omega} \wedge \mathbf{v} = -\nabla\varphi + b\hat{\mathbf{z}} + \mathbf{F} \quad \text{velocity equation} \quad (24.24a)$$

$$\nabla \cdot \mathbf{v} = 0 \quad \text{continuity equation} \quad (24.24b)$$

$$\frac{Db}{Dt} = \dot{b} \quad \text{buoyancy equation} \quad (24.24c)$$

$$b = -\frac{g\rho'}{\rho_0} = -\frac{g(\rho - \rho_0)}{\rho_0} \quad \text{buoyancy defined} \quad (24.24d)$$

$$\varphi = \frac{p'}{\rho_0} = \frac{p - p_0(z)}{\rho_0} \quad \text{dynamic pressure defined} \quad (24.24e)$$

$$\rho = \rho_0(1 - \alpha\Theta + \beta S) \quad \text{linear equation of state} \quad (24.24f)$$

$$\frac{dp_0}{dz} = -\rho_0 g \quad \text{background hydrostatic pressure.} \quad (24.24g)$$

We offer the following comments on these equations.

- **EQUATION OF STATE:** The equation of state, (24.24f), is written as a linear function of salinity and Conservative Temperature, with the thermal expansion coefficient, α , and haline contraction coefficient, β , assumed constant. This form of the equation of state eliminates processes such as cabbeling and thermobaricity (see Section 50.4). These processes are important for certain features of the ocean, thus prompting the more general equation set in Section 24.1.7. However, for many studies the linear equation of state is sufficient, and we will at times make use of it, as well as the further simplified form with zero haline contraction, $\beta = 0$.
- **HYDROSTATIC APPROXIMATION:** Most numerical models of the large-scale ocean circulation are hydrostatic. Making the hydrostatic approximation in the velocity equation (24.24a) leads to the split into a horizontal velocity equation plus a vertical hydrostatic balance

$$\frac{Du}{Dt} + 2\boldsymbol{\Omega} \wedge \mathbf{v} = -\nabla_z\varphi + \mathbf{F}^h \quad \text{horizontal velocity equation} \quad (24.25a)$$

$$\frac{\partial\varphi}{\partial z} = b \quad \text{hydrostatic balance,} \quad (24.25b)$$

where ∇_z is the horizontal gradient operator and \mathbf{F}^h is the horizontal frictional acceleration.

24.1.7 Version II of the oceanic Boussinesq equations

The non-hydrostatic Boussinesq equations (24.24a)-(24.24g) are suited for many purposes in this book. However, we find the following form more convenient when examining the equations used in a realistic non-hydrostatic Boussinesq ocean model, with such models solving prognostic equations for salinity and Conservative Temperature rather than for density or buoyancy. We are thus motivated to consider the equations in the form

$$\frac{D\mathbf{v}}{Dt} + 2\boldsymbol{\Omega} \wedge \mathbf{v} = -(1/\rho_0)(\nabla p + \rho\nabla\Phi) + \mathbf{F} \quad \text{velocity equation} \quad (24.26a)$$

$$\nabla \cdot \mathbf{v} = 0 \quad \text{continuity equation} \quad (24.26b)$$

$$\frac{DS}{Dt} = \dot{S} \quad \text{salinity equation} \quad (24.26c)$$

$$\frac{D\Theta}{Dt} = \dot{\Theta} \quad \text{Conservative Temperature equation} \quad (24.26d)$$

$$\rho = \rho(S, \Theta, \Phi) \quad \text{equation of state.} \quad (24.26e)$$

We make the following comments concerning these equations.

- **GEOPOTENTIAL:** The geopotential is here considered to be a function of space and time, as relevant when studying the role of atmospheric tidal forcing or changes to the mass distribution of the planet (see Chapter 29)

$$\Phi = \Phi(\mathbf{x}, t). \quad (24.27)$$

- **EQUATION OF STATE:** The equation of state is a function of salinity, Conservative Temperature, and geopotential, thus allowing for processes such as cabelling and thermobaricity (see Section 50.4). Even so, the pressure dependence in the equation of state is computed as per a homogeneous and resting hydrostatic fluid

$$\rho(S, \Theta, \Phi) = \rho(S, \Theta, p = -\rho_0 \Phi). \quad (24.28)$$

In Section 24.7 we see why it is appropriate to take this functional form rather than the more general form discussed in Section 25.3, in which density is a function of the full *in situ* pressure rather than the geopotential: $\rho = \rho(S, \Theta, p)$.

- **HYDROSTATIC APPROXIMATION:** As for version I of the oceanic Boussinesq equations, we here list the equations when making the hydrostatic approximation in the vertical momentum equation, in which case the velocity equation (24.26a) splits into a horizontal velocity equation and hydrostatic balance

$$\frac{D\mathbf{u}}{Dt} + 2\boldsymbol{\Omega} \wedge \mathbf{v} = -(1/\rho_0) (\nabla_z p + \rho \nabla_z \Phi) + \mathbf{F}^h \quad \text{horizontal velocity equation} \quad (24.29a)$$

$$\frac{\partial p}{\partial z} = -g \rho \quad \text{hydrostatic balance.} \quad (24.29b)$$

24.1.8 Summary points

We close this section by summarizing a number of conceptual points characterizing the oceanic Boussinesq fluid. It is useful to return to this list to help avoid common conceptual confusions.

Divergent and non-divergent velocity components

The velocity that results from the Boussinesq momentum equation (i.e., the prognostic Boussinesq velocity) is non-divergent. This is the velocity used for transport as per the material time derivative operator. Additionally, there is a divergent velocity component, \mathbf{v}^d , that balances the material evolution of density

$$\frac{1}{\rho'} \frac{D\rho'}{Dt} = -\nabla \cdot \mathbf{v}^d \neq 0. \quad (24.30)$$

The divergent velocity is not used for any of the Boussinesq dynamical equations. Nonetheless, $\mathbf{v}^d \neq 0$, as its divergence balances the material evolution of density according to equation (24.30).

Concerning density evolution and thermohaline circulation

The use of a non-divergent velocity for the Boussinesq equations does not mean that the material time evolution of ρ vanishes. Instead, the scaling in Section 24.1.4 focuses just on the mass continuity equation. We must additionally acknowledge that as temperature and salinity evolve, so too does *in situ* density as determined through the equation of state. Changes to density translate into changes in pressure, which in turn drive the large-scale *thermohaline circulation*.

The thermodynamic equation for Conservative Temperature or potential temperature is needed to determine density. There are various forms for the relation between temperature and density that depend on thermodynamic assumptions. We discuss the flavors for density in Section 25.3. For purposes of realistic ocean modeling, an accurate expression for density is critical, whereas for idealized modeling it is common to assume density equals to a constant times the temperature.

Buoyancy

We note the rather trivial point that there is identically zero buoyancy in a fully homogeneous fluid where density is constant everywhere. Hence, for an exactly incompressible fluid, where density is a fixed and uniform constant, there are no buoyancy forces. Such fluids serve many purposes, as exemplified by studies of a single layer of shallow water fluid in Part V of this book. However, buoyancy forces are of primary importance for many other purposes in geophysical fluid mechanics. The oceanic Boussinesq approximation accounts for buoyancy forces, and the changes arising from processes such as heating and freshening, while making use of the more convenient kinematics of an incompressible flow. We have more to say concerning buoyancy in Section 25.7.

Distinction from traditional Boussinesq

The oceanic Boussinesq approximation is more general than the Boussinesq approximation considered in other areas of fluid mechanics (e.g., [Chandrasekhar, 1961](#)). In particular, the traditional Boussinesq approximation typically assumes a linear equation of state. However, as we saw in Section 24.1.7, the oceanic Boussinesq approximation does not generally assume a linear equation of state, thus making it suitable for realistic ocean circulation studies.

Mass and weight of a Boussinesq fluid element

The oceanic Boussinesq fluid element conserves its volume, with this property forming the basis for Boussinesq kinematics. In turn, we measure the mass of an oceanic Boussinesq fluid element by multiplying its volume times the constant reference density, ρ_0 . However, whenever there is a gravitational acceleration multiplying density we retain the full *in situ* density to measure the weight of the fluid element. So in summary, the mass and weight for an oceanic Boussinesq fluid element are given by

$$\text{mass of Boussinesq fluid element} = \rho_0 \delta V \quad (24.31\text{a})$$

$$\text{weight of Boussinesq fluid element} = g \rho \delta V. \quad (24.31\text{b})$$

This convention enables a realistic calculation of the hydrostatic pressure (Section 24.2.6) and the gravitational potential energy (Sections 24.6 and 24.7), while maintaining the convenience of an incompressible flow.

Connection to anelastic atmosphere

The atmosphere is far more compressible than the ocean, so that density variations cannot be neglected and the divergent nature of the velocity is important. However, there are some cases in which an atmospheric analog to the Boussinesq approximation can be useful. The analog is known as the *anelastic approximation* and it is mathematically isomorphic to the oceanic Boussinesq approximation. Section 2.5 of [Vallis \(2017\)](#) discusses the atmospheric anelastic approximation.

24.1.9 Further study

Section 2.4 of [Vallis \(2017\)](#) offers more details to show that density variations are small within the ocean. Further discussion of the oceanic Boussinesq approximation also can be found in Section 9.3 of [Griffies and Adcroft \(2008\)](#). This video from [SciencePrimer](#) provides a concise summary of ocean circulation arising from differences in density created by both thermal and haline (salinity) processes.

24.2 Scaling for the hydrostatic approximation

Section 23.3 provided the rudiments of a scale analysis justifying the hydrostatic approximation for large-scale ocean and atmospheric flow. However, in that discussion we noted the need to remove a dynamically inactive hydrostatic pressure before addressing the question of whether the dynamically active pressure field is approximately hydrostatic. That decomposition of pressure was performed in Section 24.1.1 as part of deriving the Boussinesq equations. We thus here return to the question of hydrostatic scaling within the context of the perfect non-rotating Boussinesq fluid equations.

24.2.1 Stratified non-rotating Boussinesq equations

The stratified non-rotating perfect Boussinesq equations are

$$\frac{Du}{Dt} = -\nabla_z \varphi \quad \text{and} \quad \frac{Dw}{Dt} = -\frac{\partial \varphi}{\partial z} + b \quad \text{and} \quad \nabla \cdot \mathbf{v} = 0 \quad \text{and} \quad \frac{Db}{Dt} = 0. \quad (24.32)$$

To help isolate the dynamically important portion of pressure, we proceed much like in Section 24.1.1 whereby buoyancy is written

$$b = b'(x, y, z, t) + \tilde{b}(z). \quad (24.33)$$

The static buoyancy $\tilde{b}(z)$ encompasses a background stratification that is in hydrostatic balance with its corresponding portion of the pressure field

$$\frac{d\tilde{\varphi}}{dz} = \tilde{b}(z). \quad (24.34)$$

The Boussinesq equations thus take the form

$$\frac{Du}{Dt} = -\nabla_z \varphi' \quad \text{and} \quad \frac{Dw}{Dt} = -\frac{\partial \varphi'}{\partial z} + b' \quad \text{and} \quad \nabla \cdot \mathbf{v} = 0 \quad \text{and} \quad \frac{Db'}{Dt} = -w N^2, \quad (24.35)$$

where

$$N^2 = \frac{d\tilde{b}}{dz} \quad (24.36)$$

is the squared buoyancy frequency from the background vertical stratification. The decomposition into a background stratification helps to isolate the dynamical portion of the horizontal pressure gradient by removing a static depth dependent background. It also allows us to consider the dynamically interesting, but simpler, case in which the background stratification dominates those perturbations around it.

24.2.2 Non-dimensionalization

Now introduce the dimensional scales (in uppercase) and corresponding non-dimensional quantities (with hats)

$$(x, y) = L(\hat{x}, \hat{y}) \quad z = H\hat{z} \quad \mathbf{u} = U\hat{\mathbf{u}} \quad w = W\hat{w} \quad (24.37)$$

$$t = T\hat{t} \quad \varphi' = \Phi\hat{\varphi}' \quad b' = B\hat{b}' \quad N^2 = \bar{N}^2\hat{N}^2, \quad (24.38)$$

which yields the equations of motion

$$\frac{U}{T}\frac{\partial\hat{u}}{\partial\hat{t}} + \frac{U^2}{L}\hat{u}\frac{\partial\hat{u}}{\partial\hat{x}} + \frac{U^2}{L}\hat{v}\frac{\partial\hat{u}}{\partial\hat{y}} + \frac{UW}{H}\hat{w}\frac{\partial\hat{u}}{\partial\hat{z}} = -\frac{\Phi}{L}\frac{\partial\hat{\varphi}'}{\partial\hat{x}} \quad (24.39a)$$

$$\frac{U}{T}\frac{\partial\hat{v}}{\partial\hat{t}} + \frac{U^2}{L}\hat{u}\frac{\partial\hat{v}}{\partial\hat{x}} + \frac{U^2}{L}\hat{v}\frac{\partial\hat{v}}{\partial\hat{y}} + \frac{UW}{H}\hat{w}\frac{\partial\hat{v}}{\partial\hat{z}} = -\frac{\Phi}{L}\frac{\partial\hat{\varphi}'}{\partial\hat{y}} \quad (24.39b)$$

$$\frac{W}{T}\frac{\partial\hat{w}}{\partial\hat{t}} + \frac{UW}{L}\hat{u}\frac{\partial\hat{w}}{\partial\hat{x}} + \frac{UW}{L}\hat{v}\frac{\partial\hat{w}}{\partial\hat{y}} + \frac{WW}{H}\hat{w}\frac{\partial\hat{w}}{\partial\hat{z}} = -\frac{\Phi}{H}\frac{\partial\hat{\varphi}'}{\partial\hat{z}} + B\hat{b}' \quad (24.39c)$$

$$\frac{B}{T}\frac{\partial\hat{b}'}{\partial\hat{t}} + \frac{UB}{L}\hat{u}\frac{\partial\hat{b}'}{\partial\hat{x}} + \frac{UB}{L}\hat{v}\frac{\partial\hat{b}'}{\partial\hat{y}} + \frac{WB}{H}\hat{w}\frac{\partial\hat{b}'}{\partial\hat{z}} = -W\bar{N}^2\hat{w}\hat{N}^2 \quad (24.39d)$$

$$\frac{U}{L}\frac{\partial\hat{u}}{\partial\hat{x}} + \frac{U}{L}\frac{\partial\hat{v}}{\partial\hat{y}} + \frac{W}{H}\frac{\partial\hat{w}}{\partial\hat{z}} = 0. \quad (24.39e)$$

24.2.3 Specifying scales

We now impose the following choices based on the flow regimes of interest.

- **TIME SCALE:** Assume that the time scale is determined by the horizontal velocity and the horizontal length scale

$$T = L/U. \quad (24.40)$$

- **VERTICAL VELOCITY:** It is common to assume the vertical velocity scales according to the continuity equation

$$\nabla_z \cdot \mathbf{u} + \frac{\partial w}{\partial z} = 0 \implies W = U\frac{H}{L} \equiv U\alpha_{\text{aspect}}, \quad (24.41)$$

where the final equality introduced the vertical to horizontal aspect ratio, α_{aspect} . However, vertical stratification acts to suppress vertical motion so that we introduce a non-dimensional number ϵ so that

$$w = W\hat{w} = \epsilon \left[\frac{HU}{L} \right] \hat{w}. \quad (24.42)$$

In Section 24.2.4 we motivate choosing ϵ to be the squared Froude number.

- **PRESSURE:** Scale the pressure according to the non-rotating balance of the material time change in horizontal velocity and the horizontal pressure gradient

$$\frac{U}{T} + \frac{UU}{L} = \frac{\Phi}{L} \implies \Phi = U^2. \quad (24.43)$$

Note that for rotating flows in near geostrophic balance, we will see how pressure scales with the Coriolis acceleration as in Section 41.1.

- BUOYANCY: Scale buoyancy according to the hydrostatic balance

$$B = \frac{\Phi}{H} = \frac{U^2}{H}. \quad (24.44)$$

With these choices, equations of motion (24.39a)-(24.39e) take on form

$$\frac{D\hat{\mathbf{u}}}{Dt} = -\hat{\nabla}\hat{\varphi}' \quad (24.45)$$

$$\epsilon \alpha_{\text{aspect}}^2 \frac{D\hat{w}}{Dt} = -\frac{\partial \hat{\varphi}'}{\partial \hat{z}} + \hat{b}' \quad (24.46)$$

$$\left[\frac{U^2}{N^2 H^2} \right] \frac{D\hat{b}'}{Dt} + \epsilon \hat{N}^2 \hat{w} = 0 \quad (24.47)$$

$$\hat{\nabla} \cdot \hat{\mathbf{u}} + \epsilon \frac{\partial \hat{w}}{\partial \hat{z}} = 0 \quad (24.48)$$

where we introduced the non-dimensional material time derivative

$$\frac{D}{Dt} = \frac{\partial}{\partial \hat{t}} + \hat{\mathbf{u}} \cdot \hat{\nabla}_z + \epsilon \hat{w} \frac{\partial}{\partial \hat{z}}. \quad (24.49)$$

24.2.4 The role of the Froude number

At this point we make a choice for the parameter ϵ and there are many choices that one could consider. For our interests it is suitable to set ϵ equal to the squared Froude number

$$\epsilon = \text{Fr}^2 = \frac{U^2}{N^2 H^2}. \quad (24.50)$$

The Froude number measures the relative strength of vertical shears of the horizontal velocity, U/H , versus the buoyancy stratification, N . Alternatively, it measures the ratio of the horizontal speed for a fluid particle, U , to an internal gravity wave speed, NH . Large Froude numbers indicate large fluid particle speeds relative to wave speeds, with $\text{Fr} > 1$ a common indicator of hydraulic instability (see Exercise 31.2 for a shallow water example). In contrast, a relatively strong stratification (N^2 large) corresponds to a small Froude number and to flow that is stabilized by vertical stratification. Note that the squared Froude number is the inverse of the Richardson number

$$\text{Ri} = \text{Fr}^{-2} = \frac{\overline{N}^2 H^2}{U^2}. \quad (24.51)$$

It is a matter of taste whether one works with Fr or Ri .

The choice (24.50) leads to the vertical velocity scale

$$W = \text{Fr}^2 \left[\frac{HU}{L} \right]. \quad (24.52)$$

For $\text{Fr} < 1$, which is the case for stratified fluids, this result means that stratification reduces the scale for the vertical velocity. The corresponding non-dimensional Boussinesq equations take the

form

$$\frac{D\hat{u}}{Dt} = -\hat{\nabla}_z \hat{\varphi}' \quad (24.53)$$

$$Fr^2 \alpha_{\text{aspect}}^2 \frac{D\hat{w}}{Dt} = -\frac{\partial \hat{\varphi}'}{\partial \hat{z}} + \hat{b}' \quad (24.54)$$

$$\frac{\hat{D}\hat{b}'}{Dt} + \hat{N}^2 \hat{w} = 0 \quad (24.55)$$

$$\hat{\nabla} \cdot \hat{\mathbf{u}} + Fr^2 \frac{\partial \hat{w}}{\partial \hat{z}} = 0. \quad (24.56)$$

The condition for hydrostatic balance in a stratified fluid thus takes the form

$$Fr^2 \alpha_{\text{aspect}}^2 \ll 1. \quad (24.57)$$

This result supports our initial suspicion that stratification suppresses vertical motion, thus reducing the vertical acceleration terms that break hydrostatic balance. That is, hydrostatic balance is more readily achieved for a stratified flow than for an unstratified flow. Note also that the horizontal divergence of the horizontal flow is reduced by the presence of stratification, which thus leads to a nearly horizontally non-divergent flow

$$|\hat{\nabla} \cdot \hat{\mathbf{u}}| = \left| Fr^2 \frac{\partial \hat{w}}{\partial \hat{z}} \right| \ll \left| \frac{\partial \hat{w}}{\partial \hat{z}} \right|. \quad (24.58)$$

Finally, this scaling reveals how the hydrostatic approximation becomes less accurate when $Fr^2 \alpha_{\text{aspect}}^2 \sim 1$, which occurs when stratification is weak and/or the aspect ratio large.

24.2.5 Horizontal hydrostatic pressure gradient

In Section 24.2.5 we studied the horizontal pressure gradient between two columns of constant density for a hydrostatic fluid. In that example, the two columns were assumed to have equal mass and the fluid was non-Boussinesq. Here, we reconsider that example for a Boussinesq fluid. The expression for the hydrostatic pressure at a point within the fluid takes on the same form as that for a compressible fluid (Section 23.3.1)

$$p_h(x, y, z, t) = p_a(x, y, t) + g \int_z^\eta \rho(x, y, z', t) dz', \quad (24.59)$$

and the horizontal pressure gradient is thus given by

$$\nabla_z p_h = \nabla_z p_a + g \rho(\eta) \nabla_z \eta + g \int_z^\eta \nabla_z \rho dz'. \quad (24.60)$$

For many studies with Boussinesq fluids we are interested in the horizontal pressure gradients in the presence of a rigid lid ocean surface whereby $\eta = 0$. In this case we compute the *internal pressure gradient*

$$\nabla_z p_h = g \int_z^0 \nabla_z \rho dz' \quad \text{rigid lid ocean.} \quad (24.61)$$

Hence, the internal horizontal pressure gradient at a vertical position z equals to the horizontal density gradient vertically integrated above that point. For example, if density increases poleward, then so too does the internal hydrostatic pressure.

24.2.6 Evolution of hydrostatic pressure

In Section 23.3.2 we developed the Eulerian evolution equation for hydrostatic pressure in a mass conserving non-Boussinesq fluid. Here we discuss the evolution in a Boussinesq fluid, starting with the expression (23.27) for hydrostatic pressure at a point within the ocean

$$p(x, y, z, t) = p_a(x, y, t) + g \int_z^\eta \rho(x, y, z', t) dz', \quad (24.62)$$

and then taking the Eulerian time tendency

$$\partial_t p = \partial_t p_a + g \rho(\eta) \partial_t \eta + \int_z^\eta \partial_t \rho(x, y, z', t) dz'. \quad (24.63)$$

These expressions hold for both the Boussinesq and non-Boussinesq fluids. However, it is only for the mass conserving non-Boussinesq fluid that we can use mass continuity (15.6) to set $\partial_t \rho = -\nabla \cdot (\mathbf{v} \rho)$. Use of Leibniz's rule then reveals that the hydrostatic pressure evolves due to the convergence of mass onto the fluid column above that point (see Section 23.3.2 for the derivation). This result is expected since the hydrostatic pressure at a point is the weight per area of fluid above that point.

For a Boussinesq fluid, the volume conserving kinematics means that we cannot replace $\partial_t \rho$ with $-\nabla \cdot (\mathbf{v} \rho)$. Correspondingly, the weight of a fluid column can change merely through *in situ* density changes, so that the weight can change even if the matter content remains fixed; e.g., through heating. We expose details by noting that the energetic consistency ideas from Section 24.7 mean that the *in situ* density in an oceanic Boussinesq fluid has the functional dependence, $\rho(S, \Theta, \Phi)$, with Φ the geopotential. Hence, the Eulerian time derivative of density is given by

$$\partial_t \rho = (\partial \rho / \partial S) \partial_t S + (\partial \Theta / \partial t) \partial_t \Theta + (\partial \rho / \partial \Phi) \partial_t \Phi. \quad (24.64)$$

With a simple geopotential, $\Phi = g z$, we have $\partial_t \Phi = 0$ since the Eulerian time derivative is computed at fixed (x, y, z) . This result leads to the time changes in the hydrostatic pressure for a Boussinesq fluid

$$\partial_t p = \partial_t p_a + g \rho(\eta) \partial_t \eta + \int_z^\eta [(\partial \rho / \partial S) \partial_t S + (\partial \Theta / \partial t) \partial_t \Theta] dz'. \quad (24.65)$$

This equation reveals the direct dependence of the hydrostatic pressure on changes in S and Θ . Hence, heating and freshening, which alter the *in situ* density, directly alter the hydrostatic pressure in a Boussinesq fluid by altering the fluid's weight per area. This result contrasts to the mass conserving non-Boussinesq fluid, whose weight per area changes only through changes to its mass per area.

24.3 How pressure enforces non-divergent flow

We return now to the case of a non-hydrostatic fluid and consider the Boussinesq velocity equation (24.26a) written in the form

$$\rho_0 (\partial_t + \mathbf{v} \cdot \nabla) \mathbf{v} + f \mathbf{z} \wedge \rho_0 \mathbf{v} = -\nabla p - \rho g \hat{\mathbf{z}} + \rho_0 \mathbf{F}. \quad (24.66)$$

The non-divergence constraint on the velocity, $\nabla \cdot \mathbf{v} = 0$, must be maintained at each point in the fluid and at each time instance. How is that constraint maintained? As we show in this section, pressure is the force that enforces non-divergence of the velocity. Furthermore, pressure

is determined through solving an elliptic boundary value problem. As we discussed in Section 3.5, elliptic partial differential equations transfer information instantaneously. Physically, this situation corresponds to the transition from a compressible fluid, in which pressure signals propagate via acoustic modes, to an incompressible flow such as for the oceanic Boussineq fluid, in which pressure adjusts instantaneously as if the acoustic waves travel at infinite speed.

24.3.1 Poisson equation for pressure

To derive the pressure equation, we expose Cartesian tensor labels on the momentum equation (24.66) to yield

$$\partial_t v_m + v_n \partial_n v_m + f \epsilon_{mnp} \delta_{3n} v_p = -\partial_m p / \rho_0 - \delta_{3m} g \rho / \rho_0 + F_m. \quad (24.67)$$

We eliminate the time derivative by taking the divergence through contracting with the operator ∂_m , thus leading to

$$-\nabla^2 p / \rho_0 = \partial_m (v_n \partial_n v_m + f \epsilon_{m3p} v_p + \delta_{3m} g \rho / \rho_0 - F_m), \quad (24.68)$$

where the Laplacian operator is

$$\nabla^2 = \partial_m \partial_m \quad (24.69)$$

and the non-divergence constraint eliminated the time derivative

$$\partial_m \partial_t v_m = \partial_t \partial_m v_m = 0. \quad (24.70)$$

Equation (24.68) can be written as the Poisson equation

$$-\nabla^2 p = \rho_0 \nabla \cdot \mathbf{G}, \quad (24.71)$$

with the vector \mathbf{G} given by

$$G_m = v_n \partial_n v_m + f \epsilon_{m3p} v_p + \delta_{3m} g \rho / \rho_0 - F_m \quad (24.72a)$$

$$\mathbf{G} = (\mathbf{v} \cdot \nabla) \mathbf{v} + f \hat{\mathbf{z}} \wedge \mathbf{v} + g (\rho / \rho_0) \hat{\mathbf{z}} - \mathbf{F}. \quad (24.72b)$$

As when studying Green's functions in Chapter 4, it is useful to maintain the minus sign on the left hand side of equation (24.71) so that a positive divergence ($\nabla \cdot \mathbf{G} > 0$) represents a positive source for p . Equivalently, for a wavelike pressure perturbation we have

$$-\nabla^2 p \propto p, \quad (24.73)$$

so that a positive source, $-\nabla^2 p = \rho_0 \nabla \cdot \mathbf{G} > 0$, leads to a local positive pressure anomaly, and conversely for a negative source. Finally, we can integrate the pressure equation (24.71) over the domain and use Gauss's Law to render the Neumann boundary condition

$$-\hat{\mathbf{n}} \cdot \nabla p = \rho_0 \hat{\mathbf{n}} \cdot \mathbf{G}. \quad (24.74)$$

where $\hat{\mathbf{n}}$ is the outward normal at the domain boundary.

24.3.2 Characterizing the pressure sources

There are four sources for the pressure field via the right hand side of the Poisson equation (24.71), and linearity of the Poisson equation allows us to separately study the sources. We here summarize the basics of these sources, borrowing from the treatment given in Section 34.5 for the kinematically simpler non-divergent barotropic fluid. As we will see, three of the pressure sources contribute to non-hydrostatic pressures and one to hydrostatic pressure. Note that these pressure sources are diagnostically related to pressure perturbations. That is, the sources are *associated* with the pressure perturbations rather than *causing* the perturbations. Such is part of the nuance of pressure in a non-divergent flow.

Divergence of self-advection

The first source is given by the divergence of self-advection, which can be written

$$\nabla \cdot \mathbf{G}_{\text{self-advect}} = \partial_n(v_m \partial_m v_n) = (\partial_n v_m)(\partial_m v_n) + v_m(\partial_n \partial_m v_n) = (\partial_n v_m)(\partial_m v_n), \quad (24.75)$$

where we set

$$v_m \partial_n \partial_m v_n = v_m \partial_m \partial_n v_n = 0 \quad (24.76)$$

since $\partial_n v_n = 0$ follows from the non-divergent nature of the flow. Furthermore, introducing the rate of strain tensor (equation (14.33a)) and rotation tensor (equation (14.33b)), renders

$$2 \mathbb{S}_{mn} \mathbb{S}_{mn} = (\partial_m v_n)(\partial_m v_n) + (\partial_m v_n)(\partial_n v_m) \quad (24.77)$$

$$2 \mathbb{A}_{mn} \mathbb{A}_{mn} = (\partial_m v_n)(\partial_m v_n) - (\partial_m v_n)(\partial_n v_m), \quad (24.78)$$

so that

$$\nabla \cdot \mathbf{G}_{\text{self-advect}} = (\partial_n v_m)(\partial_m v_n) = \underbrace{\mathbb{S}_{mn} \mathbb{S}_{mn}}_{\text{splat}} - \underbrace{\mathbb{A}_{mn} \mathbb{A}_{mn}}_{\text{spin}}. \quad (24.79)$$

In this equation, \mathbb{S}_{mn} are components to the rate of strain tensor and \mathbb{A}_{mn} are components to the rotation tensor. The vorticity or spin source provides a negative source to $-\nabla^2 p$. In contrast, strain term, sometimes referred to as *splat*, provides a positive source.³ As detailed in Sections 34.5.2 and 34.5.4 for the non-divergent barotropic fluid, we can understand the presence of a negative pressure source from the spin motion due to the need for pressure to counteract the centrifugal acceleration associated with curved fluid motion. Likewise, the positive pressure source from straining motion counteracts the convergent acceleration induced by the strains.

Divergence of the Coriolis acceleration

The divergence of the Coriolis acceleration introduces a pressure source given by

$$-\nabla \cdot \nabla p_{\text{coriolis}} = \nabla \cdot \mathbf{G}_{\text{coriolis}} = \nabla \cdot (f \hat{z} \wedge \mathbf{v}) = \beta u - f \zeta, \quad (24.80)$$

where $\zeta = \partial_x v - \partial_y u$ is the vertical component to the relative vorticity, and $\beta = \partial_y f$ is the planetary vorticity gradient. As an example, consider two-dimensional flow oriented cyclonically around a point as for geostrophic flow around a low pressure center. The cyclonic flow feels an associated centrifugal acceleration directed “outward”. To maintain a non-divergent two-dimensional flow requires an oppositely directed inward pressure force. Hence, such rotating flow induces a low pressure source as a means to counteract the centrifugal acceleration

$$-\nabla \cdot \nabla p_{\text{coriolis}} < 0. \quad (24.81)$$

See section 34.5.5 for more discussion of this source as found in the non-divergent barotropic fluid.

Divergence of the gravitational force per volume

The divergence of the gravitational force per volume is given by

$$-\nabla \cdot \nabla p_{\text{gravity}} = \rho_0 \nabla \cdot \mathbf{G}_{\text{gravity}} = \nabla \cdot (g \rho \hat{z}) = g \partial_z \rho, \quad (24.82)$$

³The splat source is so-named since it is large when a fluid element is squashed in a manner increasing fluid strains, akin to how strains appear when a fluid impacts or “splats” against a solid obstacle.

with this pressure source absent in the depth-independent barotropic fluid of Section 34.5. The associated pressure gradient is that arising from the local hydrostatic component of the pressure field in which

$$\nabla \cdot (-\nabla p_{\text{gravity}} + g \rho \hat{\mathbf{z}}) = 0. \quad (24.83)$$

In regions where density decreases upward, $\partial_z \rho < 0$, a compressible fluid element that conserves its mass will expand when moving upward into less dense fluid. For a non-divergent flow that conserves the volume of fluid elements, there must be a counteracting force from pressure to halt the expansion of a fluid element. This counteracting force arises from the hydrostatic component to the pressure field that acts inward to squeeze the fluid element, with this pressure force originating from the negative pressure source, $-\nabla^2 p_{\text{gravity}} = g \partial_z \rho < 0$.

Divergence of the friction vector

The third source arises from the divergence of friction,

$$-\nabla \cdot \nabla p_{\text{friction}} = \rho_0 \nabla \cdot \mathbf{G}_{\text{friction}} = -\rho_0 \nabla \cdot \mathbf{F}. \quad (24.84)$$

With interior fluid friction arising from a nonzero rate of strain (Section 19.7), we expect this pressure source to be most important in regions of large strains. Indeed, for an incompressible flow feeling only molecular viscosity, the friction operator is given by

$$\mathbf{F} = \nu \nabla^2 \mathbf{v}, \quad (24.85)$$

where $\nu > 0$ is a constant molecular viscosity. In this case $\nabla \cdot \mathbf{F} = 0$, so that viscous friction does not contribute a pressure source. More general cases are considered in applications where the molecular viscosity is replaced by a flow dependent eddy viscosity so that $\nabla \cdot \mathbf{F} \neq 0$. For the case of a converging frictional acceleration, $\nabla \cdot \mathbf{F} < 0$, friction then leads to a positive pressure source to counteract the friction to thus maintain a non-divergent flow.

24.3.3 Comments and further study

The gravitational source contributes a local hydrostatic component to the pressure field, whereas the other three sources contribute non-hydrostatic pressure sources. In many applications, such as general circulation modeling of the ocean and atmosphere, the fluid is assumed to be approximately hydrostatic. In this case vertical motion is diagnosed rather than prognosed, and the non-hydrostatic component of pressure is never needed to evolve the fluid motion. Even so, vertical derivatives in the non-hydrostatic pressure provide the vertical force needed for vertical accelerations. We have more to say on vertical motion in Section 25.7.

[Markowski and Richardson \(2010\)](#) provide lucid discussions of pressure forces acting in geophysical fluids. In particular their Section 2.5 provided inspiration for much of the current section.

24.4 Tracer budgets in Eulerian regions

We are commonly interested in the tracer budget for a fluid region, and we examined a variety of regions in Section 16.3 for a compressible fluid. Here, we expose issues that arise for tracer budgets in a Boussinesq fluid, whereby the flow is non-divergent. We specialize to the study of an Eulerian region, \mathcal{R} , and emphasize how the non-divergent flow constrains how the advective tracer transport affects changes to the volume integrated tracer content.

24.4.1 Formulating the budget equation

Consider a tracer concentration, C , and compute its net content over an Eulerian region, \mathcal{R}

$$\rho_0 \int_{\mathcal{R}} C dV = \rho_0 V \langle C \rangle, \quad (24.86)$$

where C satisfies the tracer equation

$$\frac{\partial(\rho_0 C)}{\partial t} + \nabla \cdot (\rho_0 C \mathbf{v} + \mathbf{J}) = 0, \quad (24.87)$$

and

$$\langle C \rangle = \frac{\int_{\mathcal{R}} C dV}{\int_{\mathcal{R}} dV} = \frac{1}{V} \int_{\mathcal{R}} C dV \quad (24.88)$$

is the volume averaged tracer concentration within the region with fixed volume, $V = \int_{\mathcal{R}} dV$. Following from the discussion of tracer budgets in Section 16.3, we have

$$\rho_0 \frac{d(V \langle C \rangle)}{dt} = \rho_0 V \frac{d\langle C \rangle}{dt} \quad (24.89a)$$

$$= \rho_0 \frac{d}{dt} \int_{\mathcal{R}} C dV \quad (24.89b)$$

$$= \int_{\mathcal{R}} \frac{\partial(\rho_0 C)}{\partial t} dV \quad (24.89c)$$

$$= - \int_{\mathcal{R}} \nabla \cdot (\rho_0 C \mathbf{v} + \mathbf{J}) dV \quad (24.89d)$$

$$= - \oint_{\partial\mathcal{R}} (\rho_0 C \mathbf{v} + \mathbf{J}) \cdot \hat{\mathbf{n}} d\mathcal{S}, \quad (24.89e)$$

where we used the divergence theorem on the final equality with $\partial\mathcal{R}$ the boundary of \mathcal{R} and the outward normal $\hat{\mathbf{n}}$. Changes in the total tracer contained within the region arise from the convergence of boundary fluxes due to the non-advective flux, \mathbf{J} , plus convergence of the advective tracer flux, $\rho_0 C \mathbf{v}$. Since the region volume is constant, changes in the total tracer content directly affect the volume averaged tracer concentration, $\langle C \rangle$.

At any point along the boundary, the tracer content is modified if there is a non-advective flux, \mathbf{J} , directed across the boundary. There is no *a priori* constraint on \mathbf{J} , with local properties determining its sign and magnitude. In contrast, contributions from the boundary advective flux are constrained due to the non-divergent nature of the Boussinesq velocity, which we discuss next.

24.4.2 Interpreting advective tracer contributions

As seen by equation (24.89e), any advective flux, $\rho_0 \mathbf{v} C$, that is directed into the region adds tracer to the region, whereas a flux directed outward reduces the region's tracer content. However, because the velocity is non-divergent, the tracer contained within the region is unaffected if we modify the advective tracer flux along the boundary by adding a constant. We see this property for any closed Eulerian region by writing⁴

$$0 = \int_{\mathcal{R}} \nabla \cdot \mathbf{v} dV = \oint_{\partial\mathcal{R}} \mathbf{v} \cdot \hat{\mathbf{n}} d\mathcal{S}. \quad (24.90)$$

⁴We considered property (24.90) in Exercise 17.3.

Hence, the velocity is, at each time instance, constrained so that the non-divergent flow cannot lead to the accumulation of fluid within any closed and static region. Correspondingly, the amount of fluid entering \mathcal{R} exactly and instantaneously balances the amount of fluid leaving \mathcal{R} . We can thus add any constant, k , to the advective flux without affecting the net tracer content change

$$\oint_{\partial\mathcal{R}} (C + k) \mathbf{v} \cdot \hat{\mathbf{n}} d\mathcal{S} = \oint_{\partial\mathcal{R}} C \mathbf{v} \cdot \hat{\mathbf{n}} d\mathcal{S} + k \oint_{\partial\mathcal{R}} \mathbf{v} \cdot \hat{\mathbf{n}} d\mathcal{S} = \oint_{\partial\mathcal{R}} C \mathbf{v} \cdot \hat{\mathbf{n}} d\mathcal{S}. \quad (24.91)$$

Boundary advection occurring with $C = k$ has no affect on the net tracer within a region since an equal amount of fluid enters as leaves.

To help interpret the role of advective fluxes on integrated tracer content, we find it useful to set the arbitrary constant to be $k = -\langle C \rangle$. In this manner, the advective contribution to the tracer budget takes the form

$$V \left[\frac{d\langle C \rangle}{dt} \right]_{\text{advective}} = - \oint_{\partial\mathcal{R}} C \mathbf{v} \cdot \hat{\mathbf{n}} d\mathcal{S} = - \oint_{\partial\mathcal{R}} (C - \langle C \rangle) \mathbf{v} \cdot \hat{\mathbf{n}} d\mathcal{S}. \quad (24.92)$$

Hence, advective transport through the region boundary changes the region integrated C , and thus the volume mean $\langle C \rangle$, only if the boundary transport occurs with C values that differ from the region average, $\langle C \rangle$.

24.4.3 A rectangular region example

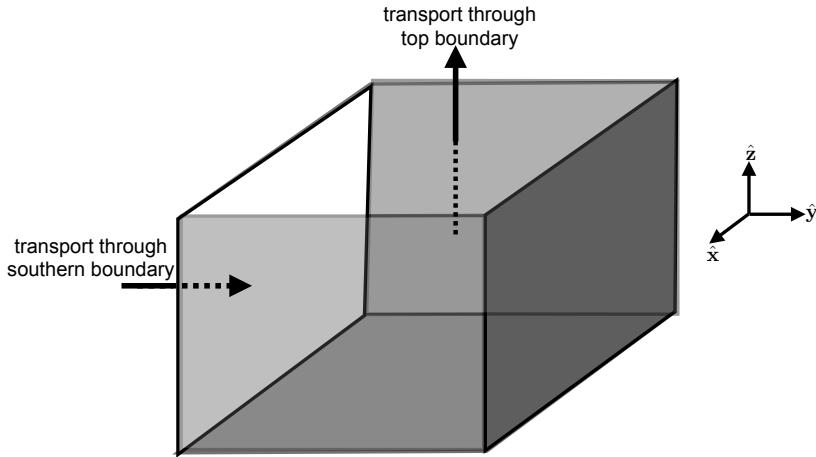


FIGURE 24.1: A rectangular ocean region for considering the tracer budget in a Boussinesq fluid. We assume the only open boundaries of the region are along the southern boundary and top boundary, with the remaining boundaries closed. Note that the top boundary can generally be within the ocean interior, and so does not need to be the top of the ocean. The non-divergent nature of the flow means that any fluid entering through the southern boundary must leave through the top boundary, and conversely. This constraint on the flow impacts on how the advective tracer transport affects changes to the volume integrated tracer within the region.

Consider the rectangular region of Figure 24.1 that is closed along its bottom, northern, eastern, and western boundaries, yet that is open along its top boundary and southern boundary. Specify flow along the southern boundary to be northward and let it bring fluid into the region with $C_{\text{south}} > \langle C \rangle$, thus acting to increase $\langle C \rangle$. Due to the non-divergent nature of the velocity, the northward transport of fluid through the southern boundary must be precisely balanced by a vertically upward transport of fluid out of the top boundary. What does that vertical transport

of fluid imply about changes to $\langle C \rangle$? The answer depends on the tracer concentration on the top boundary.

- If $C_{\text{top}} > \langle C \rangle$, then the relatively high values of tracer that leave through the top boundary acts to decrease the net C within the region, thus counteracting the contribution of $C_{\text{south}} > \langle C \rangle$ that enters through the southern boundary. For the special case of equal C transports through the two boundaries, then there is no accumulation of C within the region so that $\langle C \rangle$ remains unchanged.
- If $C_{\text{top}} = \langle C \rangle$, then $\langle C \rangle$ increases due to the transport of C through the southern boundary, with no net transport across the top boundary.
- If $C_{\text{top}} < \langle C \rangle$, then the vertical transport of $C_{\text{top}} < \langle C \rangle$ increases $\langle C \rangle$ acting just like the $C_{\text{south}} > \langle C \rangle$ fluid that enters through the southern boundary. That is, we can increase $\langle C \rangle$ in the region by bringing fluid into the domain with C greater than $\langle C \rangle$, or by exporting fluid with C less than $\langle C \rangle$.

24.4.4 Comments

Constraints introduced by the non-divergent nature of the Boussinesq fluid render subtleties to the physical interpretation of how tracer fluxes affect the budget of tracer within a region. These constraints are absent from the compressible fluid, whose flow is generally divergent. The discussion in this section was motivated by Appendix B of [Gregory \(2000\)](#), who studied heat (enthalpy) budgets within a Boussinesq ocean circulation model.

24.5 Subgrid scale processes and boundary conditions

The Boussinesq equations from Section 24.1 must be supplemented with boundary conditions as well as subgrid scale parameterizations. We examine boundary stresses in Chapter 19.9; study tracer subgrid scale parameterizations in Chapter 50; and detail the variety of boundary buoyancy fluxes in Section 51.4. Even so, there is far more to each of these stories than provided in this book. In this section we summarize facets sufficient to help work through exercises at the end of this chapter. One key point to emphasize here is that for all the subgrid scale and boundary conditions, appearances of the *in situ* density are converted to the Boussinesq reference density, ρ_0 .

24.5.1 Material evolution of buoyancy

The term \dot{b} appearing in the buoyancy equation (24.24c) is a shorthand for any process leading to a material change in buoyancy, such as a boundary flux, parameterized turbulent mixing of salinity or Conservative Temperature, or vertical motion through a vertically stratified fluid. With a linear equation of state, we can write the material evolution as the convergence of a subgrid scale buoyancy flux

$$\frac{Db}{Dt} = -\nabla \cdot \mathbf{F}^b. \quad (24.93)$$

A particularly common form for this flux is given by

$$\mathbf{F}^b = -\kappa \frac{\partial b}{\partial z} \hat{z} + \mathbf{v}^* b. \quad (24.94)$$

The first term is a downgradient vertical diffusive flux with the vertical *eddy diffusivity*, $\kappa > 0$, a function of the flow state so that

$$\kappa = \kappa(x, y, z, t). \quad (24.95)$$

The second term is an advective flux, where the advective velocity, $\mathbf{v}^* = (\mathbf{u}^*, w^*)$, is assumed to be non-divergent

$$\nabla \cdot \mathbf{v}^* = \nabla_z \cdot \mathbf{u}^* + \frac{\partial w^*}{\partial z} = 0. \quad (24.96)$$

The velocity \mathbf{v}^* is commonly termed the *eddy-induced* velocity, with a particular choice for its parameterization discussed in Exercise 24.10 and further examined in Section 50.1.

Boundary conditions for buoyancy are commonly incorporated through the boundary conditions on the subgrid scale fluxes, with the following representing one form for these boundary conditions

$$\hat{\mathbf{n}} \cdot \mathbf{F}^b = Q_{\text{surf}} \quad \text{at } z = \eta(x, y, t) \quad (24.97a)$$

$$(-\hat{\mathbf{n}}) \cdot \mathbf{F}^b = Q_{\text{bot}} \quad \text{at } z = \eta_b(x, y) \quad (24.97b)$$

$$\mathbf{v}^* \cdot \hat{\mathbf{n}} = 0 \quad \text{all boundaries,} \quad (24.97c)$$

where Q_{surf} is the surface buoyancy flux, Q_{bot} is the bottom buoyancy flux (e.g., geothermal heating), and $\hat{\mathbf{n}}$ is the outward normal at the boundaries. Both Q_{surf} and Q_{bot} are positive upward. In Section 51.4 we detail the plethora of processes leading to boundary fluxes of buoyancy.

24.5.2 Mechanical forcing and dissipation

The ocean is a forced-dissipative system, with mechanical and buoyant forcing predominantly at the surface and bottom boundaries and mechanical dissipation via molecular viscosity. We must parameterize mechanical dissipation in the ocean interior arising from subgrid scale processes. A common form for this parameterization is via a vertical transfer of momentum through vertical shears that are weighted by a vertical *eddy viscosity*, $\nu^{\text{eddy}} > 0$, in which case the horizontal frictional acceleration in the ocean interior is written as the vertical divergence of a turbulent stress

$$\mathbf{F}^h = \frac{1}{\rho_0} \frac{\partial \boldsymbol{\tau}}{\partial z} = \frac{\partial}{\partial z} \left[\nu^{\text{eddy}} \frac{\partial \mathbf{u}}{\partial z} \right], \quad (24.98)$$

where $\boldsymbol{\tau}$ is a horizontal turbulent stress whose vertical derivative contributes to the vertical transfer of horizontal momentum. The eddy viscosity can be many orders of magnitude larger than the molecular viscosity in regions of strong turbulent mixing. Whereas the molecular viscosity is a function of the fluid composition (Section 19.7), the eddy viscosity is a function of the flow state, which in turn means that

$$\nu^{\text{eddy}} = \nu^{\text{eddy}}(\mathbf{x}, t). \quad (24.99)$$

We parameterize the turbulent mechanical forcing at the ocean boundaries via a boundary stress. This stress is introduced to the governing equations through the following surface and bottom boundary conditions placed on the turbulent stress

$$\boldsymbol{\tau} = \boldsymbol{\tau}^{\text{wind}} \quad \text{at } z = \eta(x, y, t) \quad (24.100)$$

$$\boldsymbol{\tau} = \boldsymbol{\tau}^{\text{bott}} \quad \text{at } z = \eta_b(x, y) \quad (24.101)$$

At the ocean surface, $z = \eta$, we introduce the turbulent stress, $\boldsymbol{\tau}^{\text{winds}}$, arising from the transfer of momentum between the ocean and atmosphere (or the ocean and ice). In practice, the stress is

computed by a boundary layer parameterization that ingests the momentum from the atmosphere and computes a stress that is transferred to the ocean through these boundary conditions. As per Newton's third law (the action/reaction law; see Section 19.9), the stress imparted to the ocean is equal and opposite the stress imparted to the atmosphere at its lower boundary. At the ocean bottom, $z = \eta_b$, we often parameterize subgrid scale interactions with bottom topography via a quadratic bottom drag

$$\boldsymbol{\tau}^{\text{bott}} = -C_D \mathbf{u} |\mathbf{u}|, \quad (24.102)$$

where $C_D > 0$ is a dimensionless *drag coefficient* that is sometimes assumed to be a function of the bottom topographic roughness. This bottom stress acts to drag the ocean bottom velocity towards a state of rest. It is equal and opposite to the frictional stress transferred to the solid earth from the ocean.

24.6 Mechanical energy analysis part I

The volume of an fluid element in a non-divergent flow remains materially invariant even as pressure acts on the element, which means that pressure can perform no mechanical work. In turn, non-divergent flow supports no pressure work conversion between internal energy and kinetic energy. We are thus led to formulate a mechanical energy budget for the Boussinesq system that is closed for a perfect fluid with a time-independent geopotential and in the absence of boundary effects. Formulating a closed budget is the topic of this section as well as Section 24.7.

We here develop the mechanical energy budget for an unforced non-hydrostatic Boussinesq fluid assuming density is a function only of salinity and Conservative Temperature,

$$\rho = \rho(S, \Theta). \quad (24.103)$$

The more general pressure-dependent equation of state is considered in Section 24.7. Furthermore, we choose to work with the velocity, density, and continuity equations in the form

$$\rho_0 \left[\frac{D\mathbf{v}}{Dt} + 2\boldsymbol{\Omega} \wedge \mathbf{v} \right] = -\nabla p - \rho \nabla \Phi \quad (24.104a)$$

$$\frac{D\rho}{Dt} = \frac{\partial \rho}{\partial S} \frac{DS}{Dt} + \frac{\partial \rho}{\partial \Theta} \frac{D\Theta}{Dt} = \dot{\rho} \quad (24.104b)$$

$$\nabla \cdot \mathbf{v} = 0. \quad (24.104c)$$

Note that the geopotential, Φ , is generally a function of space and time, $\Phi = \Phi(\mathbf{x}, t)$, which is appropriate when studying astronomical tides or mass inhomogeneities creating spatial variations in the gravity field (see Chapter 29).

24.6.1 Kinetic energy

To obtain a kinetic energy equation, start by taking the dot product of the velocity, \mathbf{v} , and the momentum equation (24.179a) and note that the Coriolis acceleration drops out since it is orthogonal to the velocity

$$\mathbf{v} \cdot (2\boldsymbol{\Omega} \wedge \mathbf{v}) = 0. \quad (24.105)$$

The material time derivative takes the form

$$v_i \left[\frac{\partial v_i}{\partial t} + v_j \partial_j v_i \right] = \frac{\partial \mathcal{K}}{\partial t} + \mathbf{v} \cdot \nabla \mathcal{K} = \frac{D\mathcal{K}}{Dt}, \quad (24.106)$$

where we introduced the kinetic energy per mass

$$\mathcal{K} = \mathbf{v} \cdot \mathbf{v}/2. \quad (24.107)$$

The equation for the Boussinesq kinetic energy per volume thus takes the form

$$\rho_0 \frac{D\mathcal{K}}{Dt} = -\mathbf{v} \cdot \nabla p - \rho \mathbf{v} \cdot \nabla \Phi \quad (24.108)$$

Alternatively, we can write this equation in the Eulerian flux-form

$$\frac{\partial(\rho_0 \mathcal{K})}{\partial t} + \nabla \cdot [\mathbf{v} (\rho_0 \mathcal{K} + p)] = -\rho \mathbf{v} \cdot \nabla \Phi, \quad (24.109)$$

where we used $\nabla \cdot \mathbf{v} = 0$ to write $\mathbf{v} \cdot \nabla p = \nabla \cdot (\mathbf{v} p)$. Note that in Exercise 24.4 we show that the kinetic energy evolution derived here for the non-hydrostatic fluid holds also for the hydrostatic fluid, yet with the kinetic energy in the hydrostatic fluid determined solely by the horizontal velocity.

The term $-\rho \mathbf{v} \cdot \nabla \Phi$ in the kinetic energy equation (24.109) is a source/sink that arises from fluid motion crossing surfaces of constant geopotential. Moving a fluid element down the geopotential gradient ($\mathbf{v} \cdot \nabla \Phi < 0$) increases the kinetic energy, and conversely when the fluid moves up the geopotential gradient. We sometimes refer to this process as *buoyancy work*, particularly when considered in the context of a vertically stratified fluid. We can further exemplify this term by taking the simplified form of the geopotential, $\Phi = g z$, in which $\mathbf{v} \cdot \nabla \Phi = \rho g w$.

24.6.2 Gravitational and mechanical energies

A fluid element has a gravitational potential energy per mass given by the geopotential, Φ , which has a material time derivative

$$\frac{D\Phi}{Dt} = \partial_t \Phi + \mathbf{v} \cdot \nabla \Phi. \quad (24.110)$$

The time-dependent geopotential provides an external source of potential energy to the system. Additionally, motion moving up the gradient of the geopotential ($\mathbf{v} \cdot \nabla \Phi > 0$) increases the potential energy per mass, and conversely for motion down the geopotential gradient. Adding to the kinetic energy equation (24.108) renders the material evolution

$$\rho_0 \frac{D\mathcal{K}}{Dt} + \rho \frac{D\Phi}{Dt} = -\nabla \cdot (\mathbf{v} p) + \rho \partial_t \Phi. \quad (24.111)$$

Note how the buoyancy work source, $\rho \mathbf{v} \cdot \nabla \Phi$, dropped out from this budget. Consequently, this term provides a reversible transfer of mechanical energy between gravitational potential energy per volume and the kinetic energy per volume. We saw the same transfer in Section 22.3.3 when studying the mechanical energy budget for a compressible non-Boussinesq fluid.

Equation (24.111) has nearly the same form as that for the non-Boussinesq fluid given by equation (22.36c). However, for the Boussinesq fluid it does not lead to a flux-form Eulerian conservation law for mechanical energy, even for the perfect fluid. Operationally, the derivations diverge at this point since for the non-Boussinesq fluid we make use of the mass continuity equation (15.6) to write the material evolution of density. In contrast, material density evolution in a Boussinesq fluid is determined by material changes in temperature, salinity, and pressure.

To develop a closed Boussinesq mechanical energy budget, add $\Phi D\rho/Dt = \Phi \dot{\rho}$ to both sides of equation (24.111) to render the material evolution for the mechanical energy *per volume*

$$\frac{D}{Dt} [\rho_0 \mathcal{K} + \rho \Phi] = -\nabla \cdot (\mathbf{v} p) + \rho \partial_t \Phi + \dot{\rho} \Phi, \quad (24.112)$$

which has the Eulerian flux-form expression

$$\partial_t (\rho_0 \mathcal{K} + \rho \Phi) + \nabla \cdot [\mathbf{v} (\rho_0 \mathcal{K} + \rho \Phi + p)] = \rho \partial_t \Phi + \dot{\rho} \Phi. \quad (24.113)$$

In the absence of boundary processes we have

$$\rho_0 \frac{d}{dt} \left[\int m \, dV \right] = \int [\rho \partial_t \Phi + \dot{\rho} \Phi] \, dV, \quad (24.114)$$

where the Boussinesq mechanical energy per volume is given by

$$\rho_0 m = \rho_0 \mathcal{K} + \rho \Phi. \quad (24.115)$$

Recall we are assuming $\rho = \rho(S, \Theta)$ in this section, so that

$$\dot{\rho} = (\partial \rho / \partial S) \dot{S} + (\partial \rho / \partial \Theta) \dot{\Theta}. \quad (24.116)$$

Hence, for a time-independent geopotential and in the absence of irreversible mixing then the global integrated Boussinesq mechanical energy is constant.

24.7 Mechanical energy analysis part II

In this section we build on the analysis from Section 24.6, here allowing density to be a function of salinity, Conservative Temperature, and pressure so that

$$\frac{D\rho}{Dt} = \frac{\partial \rho}{\partial S} \frac{DS}{Dt} + \frac{\partial \rho}{\partial \Theta} \frac{D\Theta}{Dt} + \frac{\partial \rho}{\partial p} \frac{Dp}{Dt}. \quad (24.117)$$

Density thus materially evolves even in the absence of mixing or diabatic processes since Dp/Dt is nonzero whenever flow crosses isobars. Hence, the Eulerian flux-form mechanical energy equation (24.113) now takes on the form

$$\partial_t (\rho_0 \mathcal{K} + \rho \Phi) + \nabla \cdot [\mathbf{v} \cdot (\rho_0 \mathcal{K} + \rho \Phi + p)] = \rho \partial_t \Phi + \Phi \left[\frac{\partial \rho}{\partial S} \dot{S} + \frac{\partial \rho}{\partial \Theta} \dot{\Theta} + \frac{\partial \rho}{\partial p} \dot{p} \right]. \quad (24.118)$$

The right hand side terms provide interior sources that contribute to the evolution of mechanical energy. In the absence of mixing, diabatic processes, and with a time independent geopotential, the sources reduce to a term arising from motion across pressure surfaces. Such motion can occur for either reversible or irreversible processes. This pressure source term is rather awkward since it means the mechanical energy budget is not closed even when the flow is reversible and with time independent astronomical forces. We now follow the approach of [Young \(2010\)](#) to recover a closed Boussinesq mechanical energy budget by making use of a modified form of the gravitational potential energy.

24.7.1 Boussinesq dynamic enthalpy

In this section we introduce a new thermodynamic function that, in effect, provides us with an integrating factor to render a closed form to the Boussinesq mechanical energy budget. This function is referred to as the *Boussinesq dynamic enthalpy*. Before considering that function we do a brief warm-up exercise to refamiliarize ourselves with the necessary thermodynamic formalism from Chapter 21.

Material time changes to an integral of density over pressure

Consider a thermodynamic potential, $\tilde{\Pi}(S, \Theta, p | p_r)$, defined according to the pressure integral of the *in situ* density

$$\tilde{\Pi}(S, \Theta, p | p_r) \equiv \int_{p_r}^p \rho(S, \Theta, p') dp' \implies \left[\frac{\partial \tilde{\Pi}}{\partial p} \right]_{S, \Theta} = \rho(S, \Theta, p), \quad (24.119)$$

where p_r is an arbitrary constant reference pressure. The notation $\tilde{\Pi}(S, \Theta, p | p_r)$ emphasizes that p_r is a specified parameter whereas S, Θ, p are coordinates in thermodynamic configuration space. The integration in equation (24.119) is taken over pressure in a thermodynamic configuration space rather than an integral over a region in x -space.⁵ Accordingly, the infinitesimal increment of $\tilde{\Pi}$ is given by

$$\delta \tilde{\Pi} = \delta S \left[\frac{\partial \tilde{\Pi}}{\partial S} \right]_{\Theta, p} + \delta \Theta \left[\frac{\partial \tilde{\Pi}}{\partial \Theta} \right]_{S, p} + \delta p \left[\frac{\partial \tilde{\Pi}}{\partial p} \right]_{S, \Theta} \quad (24.120a)$$

$$= \delta S \left[\frac{\partial \tilde{\Pi}}{\partial S} \right]_{\Theta, p} + \delta \Theta \left[\frac{\partial \tilde{\Pi}}{\partial \Theta} \right]_{S, p} + \rho(S, \Theta, p) \delta p. \quad (24.120b)$$

If the increment is computed following a moving fluid element then we are led to the material time derivative

$$\frac{D \tilde{\Pi}}{Dt} = \frac{DS}{Dt} \left[\frac{\partial \tilde{\Pi}}{\partial S} \right]_{\Theta, p} + \frac{D\Theta}{Dt} \left[\frac{\partial \tilde{\Pi}}{\partial \Theta} \right]_{S, p} + \rho(S, \Theta, p) \frac{Dp}{Dt}. \quad (24.121)$$

Material time changes to an integral of density over geopotential

Using the same formalism as above, now consider a thermodynamic potential that is a function of salinity, Conservative Temperature, and geopotential

$$\Pi(S, \Theta, \Phi | \Phi_r) \equiv \int_{\Phi_r}^{\Phi} \rho(S, \Theta, \Phi') d\Phi' \implies \left[\frac{\partial \Pi}{\partial \Phi} \right]_{S, \Theta} = \rho(S, \Theta, \Phi), \quad (24.122)$$

where Φ_r is an arbitrary constant reference geopotential. We offer the following three comments concerning Π .

- For density that is independent of the geopotential, then $\Pi dV = (\Phi - \Phi_r) \rho dV$, which is the gravitational potential energy relative to a reference state. We thus interpret $\Pi(S, \Theta, \Phi)$ as a generalized gravitational potential energy per volume.
- One might consider Π to be the difference in hydrostatic pressure between Φ and Φ_r as per equation (18.38). However, the integral in equation (18.38) occurs in x -space between two geopotentials, with that integration generally crossing surfaces of constant S and Θ . In contrast, the integration in equation (24.122) is taken over the geopotential in thermodynamic configuration space so that S and Θ are fixed while performing the integral. In this manner, the geopotential, rather than pressure, provides a coordinate within a Boussinesq thermodynamic configuration space along with S and Θ .

⁵This is an example where the discussion in Section 22.4.2 is key, whereby we must distinguish between functions in a thermodynamic configuration space versus functions as space-time fields.

- *Young (2010)* provides motivation for calling $\tilde{\Pi}$ the *Boussinesq dynamic enthalpy*.

Following the same formalism used to derive $D\tilde{\Pi}/Dt$ in equation (24.121), we here compute the material time derivative of the Boussinesq dynamic enthalpy

$$\frac{D\Pi}{Dt} = \frac{DS}{Dt} \left[\frac{\partial\Pi}{\partial S} \right]_{\Theta,\Phi} + \frac{D\Theta}{Dt} \left[\frac{\partial\Pi}{\partial\Theta} \right]_{S,\Phi} + \rho(S, \Theta, \Phi) \frac{D\Phi}{Dt}. \quad (24.123)$$

We now create a mechanical energy budget in the form

$$\frac{D}{Dt} [\rho_0 \mathcal{K} + \Pi] = -[\mathbf{v} \cdot \nabla p + \rho \mathbf{v} \cdot \nabla \Phi] + \dot{S} \left[\frac{\partial\Pi}{\partial S} \right]_{\Theta,\Phi} + \dot{\Theta} \left[\frac{\partial\Pi}{\partial\Theta} \right]_{S,\Phi} + \rho \frac{D\Phi}{Dt} \quad (24.124a)$$

$$= -[\mathbf{v} \cdot \nabla p + \rho \mathbf{v} \cdot \nabla \Phi] + \dot{S} \left[\frac{\partial\Pi}{\partial S} \right]_{\Theta,\Phi} + \dot{\Theta} \left[\frac{\partial\Pi}{\partial\Theta} \right]_{S,\Phi} + \rho (\partial_t \Phi + \mathbf{v} \cdot \nabla \Phi) \quad (24.124b)$$

$$= -\mathbf{v} \cdot \nabla p + \rho \partial_t \Phi + \dot{S} \left[\frac{\partial\Pi}{\partial S} \right]_{\Theta,\Phi} + \dot{\Theta} \left[\frac{\partial\Pi}{\partial\Theta} \right]_{S,\Phi}, \quad (24.124c)$$

whose Eulerian flux-form expression is given by

$$\partial_t (\rho_0 \mathcal{K} + \Pi) + \nabla \cdot [\mathbf{v} \cdot (\rho_0 \mathcal{K} + \Pi + p)] = \rho \partial_t \Phi + \dot{S} \left[\frac{\partial\Pi}{\partial S} \right]_{\Theta,\Phi} + \dot{\Theta} \left[\frac{\partial\Pi}{\partial\Theta} \right]_{S,\Phi}. \quad (24.125)$$

We thus see that in the absence of irreversible effects, and with a time independent geopotential, we have succeeded in deriving a closed (i.e., flux-form) mechanical energy budget for a Boussinesq fluid, with

$$\rho_0 m = \rho_0 \mathcal{K} + \Pi \xrightarrow{\rho=\rho(S,\Theta)} \rho_0 \mathcal{K} + \rho (\Phi - \Phi_r) \quad (24.126)$$

the appropriate Boussinesq form of the mechanical energy per volume.

24.7.2 Regionally integrated Boussinesq dynamic enthalpy

Following the treatment for a compressible non-Boussinesq fluid in Section 22.2.3, we here study the evolution of the gravitational potential energy integrated over a finite region, \mathcal{R} , that is open to material mass transport. Rather than working with the geopotential as done for the non-Boussinesq fluid, we here follow the discussion in Section 24.7.1 by making use of the Boussinesq dynamic enthalpy, thus ensuring a closed mechanical energy budget.

For this purpose we make use the Leibniz-Reynolds transport theorem in the form of equation (16.34) to find

$$\frac{d}{dt} \int_{\mathcal{R}} \Pi dV = \int_{\mathcal{R}} \left[\frac{\partial\Pi}{\partial t} + \nabla \cdot (\Pi \mathbf{v}^{(b)}) \right] dV \quad (24.127a)$$

$$= \int_{\mathcal{R}} \left[\frac{\partial\Pi}{\partial t} + \nabla \cdot (\Pi \mathbf{v}) - \nabla \cdot (\Pi \mathbf{v}) + \nabla \cdot (\Pi \mathbf{v}^{(b)}) \right] dV \quad (24.127b)$$

$$= \int_{\mathcal{R}} \left[\frac{D\Pi}{Dt} + \Pi \nabla \cdot \mathbf{v} + \nabla \cdot [\Pi (\mathbf{v}^{(b)} - \mathbf{v})] \right] dV \quad (24.127c)$$

$$= \int_{\mathcal{R}} \left[\frac{D\Pi}{Dt} + \nabla \cdot [\Pi (\mathbf{v}^{(b)} - \mathbf{v})] \right] dV, \quad (24.127d)$$

where $\mathbf{v}^{(b)}$ is the velocity of a point on the boundary of the domain, $\partial\mathcal{R}$, and we set $\nabla \cdot \mathbf{v} = 0$ as per a Boussinesq fluid. We expose the contributions from irreversible S and Θ processes by making use of the identity (24.123)

$$\frac{d}{dt} \int_{\mathcal{R}} \Pi dV = \int_{\mathcal{R}} \left[\frac{DS}{Dt} \left[\frac{\partial \Pi}{\partial S} \right]_{\Theta, \Phi} + \frac{D\Theta}{Dt} \left[\frac{\partial \Pi}{\partial \Theta} \right]_{S, \Phi} + \rho \frac{D\Phi}{Dt} + \nabla \cdot [\Pi (\mathbf{v}^{(b)} - \mathbf{v})] \right] dV. \quad (24.128)$$

Note that a constant can be added to the dynamic enthalpy without altering the energetics, which is seen by noting that volume conservation means that (equation (17.67))

$$\frac{d}{dt} \int dV = - \int_{\mathcal{R}} \nabla \cdot (\mathbf{v} - \mathbf{v}^{(b)}) dV = - \int_{\partial\mathcal{R}} (\mathbf{v} - \mathbf{v}^{(b)}) \cdot \hat{\mathbf{n}} d\mathcal{S}. \quad (24.129)$$

If the region is a vertical column of fluid with fixed horizontal cross-section, extending from the ocean surface to the ocean bottom, then there is horizontal transport across the vertical boundaries, plus vertical transport of mass across the ocean free surface. For the free surface we make use of the surface kinematic boundary condition (15.73c) to write

$$\int_{z=\eta} (\Pi/\rho) \rho (\mathbf{v}^{(\eta)} - \mathbf{v}) \cdot \hat{\mathbf{n}} d\mathcal{S} = \int_{z=\eta} (\Pi/\rho) Q_m dA. \quad (24.130)$$

In this equation, Q_m is the mass per time per horizontal area of matter crossing the ocean free surface at $z = \eta$ where $Q_m > 0$ for matter entering the ocean domain, and $d\mathcal{S}$ is the area element on the free surface with dA its horizontal projection.

For the special case of an equation of state independent of pressure, $\rho = \rho(S, \Theta)$ (Section 24.6), we have $\Pi = \rho(\Phi - \Phi_r)$ so that equation (24.128) reduces to

$$\frac{d}{dt} \int_{\mathcal{R}} \rho (\Phi - \Phi_r) dV = \int_{\mathcal{R}} \left[(\Phi - \Phi_r) \frac{D\rho}{Dt} + \rho \frac{D\Phi}{Dt} + \nabla \cdot [\rho (\Phi - \Phi_r) (\mathbf{v}^{(b)} - \mathbf{v})] \right] dV. \quad (24.131)$$

Note that the reference geopotential, Φ_r , drops out since the Boussinesq form of the Leibniz-Reynolds transport theorem (i.e., equation (24.127d) with ρ replacing Π) leads to the identity

$$\frac{d}{dt} \int_{\mathcal{R}} \rho dV = \int_{\mathcal{R}} \left[\frac{D\rho}{Dt} + \nabla \cdot [\rho (\mathbf{v}^{(b)} - \mathbf{v})] \right] dV. \quad (24.132)$$

This equation means that the mass for a region of Boussinesq fluid (left hand side) changes through boundary terms, as for a non-Boussinesq fluid, plus processes that lead to material time changes in S and Θ

$$\rho = \rho(S, \Theta) \implies \frac{D\rho}{Dt} = \frac{\partial \rho}{\partial S} \frac{DS}{Dt} + \frac{\partial \rho}{\partial \Theta} \frac{D\Theta}{Dt}. \quad (24.133)$$

We understand the presence of the \dot{S} and $\dot{\Theta}$ terms by noting that irreversible processes, such as mixing, do not alter volume in a Boussinesq fluid. Hence, if irreversible processes change density then there must be a corresponding change in mass.

24.7.3 Comments

Decoupling mechanical energy from internal energy

There are further niceties required to unravel energetics of the oceanic Boussinesq fluid, with details provided by [Young \(2010\)](#). When encountering these details for the first time one may wonder why

bother since the compressible energetics discussed in Sections 22.3 and 22.5 are rather straightforward. However, the difficulty with compressible energetics arises from the internal energy. Namely, since many geophysical flows, particularly those in the ocean, have speeds that are tiny compared to molecular speeds (see Section 12.3), the mechanical energy associated with geophysical flow is tiny relative to the internal energy arising from molecular motions. So when studying the total energy budget for a compressible fluid, that energy is dominated by the internal energy. As detailed in [Young \(2010\)](#), the oceanic Boussinesq approximation allows one to focus on the Boussinesq mechanical energy arising just from the fluid flow, and it does so by decoupling mechanical energy from the internal energy.

General form of the geopotential

The treatments in [Young \(2010\)](#) and Section 2.4.3 of [Vallis \(2017\)](#) focus on the simple form of the geopotential, $\Phi = g z$, in which case it appears that density in an energetically consistent Boussinesq fluid can at most have the space-time dependence

$$\rho = \rho[S(\mathbf{x}, t), \Theta(\mathbf{x}, t), p = -\rho_0 g z]. \quad (24.134)$$

However, the formalism from this section allows for a general geopotential, including those that arise from astronomical tidal forcing and from mass redistributions such as near ice shelves. In these cases we retain a consistent Boussinesq energetics with density of the more general form

$$\rho = \rho[S(\mathbf{x}, t), \Theta(\mathbf{x}, t), p = -\rho_0 \Phi(\mathbf{x}, t)]. \quad (24.135)$$

Distinct manifestations of irreversible processes

It is notable that the irreversible terms from \dot{S} and $\dot{\Theta}$ that appear in the mechanical energy equations (24.113), (24.124c) and (24.125) are absent from the non-Boussinesq budget in equation (22.36c). Instead, for the non-Boussinesq fluid, the irreversible mixing processes manifest through their effects on flow convergence via the mass continuity equation

$$-\nabla \cdot \mathbf{v} = \rho^{-1} \dot{\rho}. \quad (24.136)$$

Since the Boussinesq fluid has a zero flow divergence, the role of mixing on the potential energy budget appears elsewhere within the mechanical energy budget.

24.8 Available potential energy

As we have seen, the gravitational potential energy per mass of a fluid element as measured relative to the $z = 0$ geopotential is given by $g \rho z$. But how much of that potential energy is convertible to kinetic energy? Clearly not all of it since a state of zero potential energy means all of the fluid sits at $z = 0$, which is not generally possible. In Section 24.6 we noted that $-g^{-1} b \Phi$ is the potential energy per mass relative to the constant density background state. Pursuing this idea one more step, consider a background buoyancy, $b_{\text{ref}}(z)$, that has a non-zero depth dependence but with no horizontal dependence. Without any horizontal buoyancy gradients there are zero horizontal pressure gradients so that an initially static fluid will remain static. As seen in the discussion of vorticity in Section 36.8.1, a fluid with zero baroclinicity does not generate vorticity since its pressure gradients are perpendicular to buoyancy gradients, which are vertical for a reference state with $b = b_{\text{ref}}(z)$ (see Section 24.1.2 for the Boussinesq baroclinicity vector). These ideas motivate us to compute the potential energy relative to a depth-dependent background buoyancy profile.

Lorenz (1955) suggested that a particularly relevant background buoyancy state is that obtained by a reversible rearrangement or “sorting” of the original buoyancy to a state that has zero baroclinicity. A reversible rearrangement means there is no mixing when moving between the original state and the background state. The difference in gravitational potential energy between these two states is termed the *available potential energy*, with the APE measuring the potential energy accessible for generating reversible motion. Figure 24.2 illustrates these notions. In the remainder of this section we provide further details to support this conceptual description of APE. We restrict attention to the perfect Boussinesq fluid with a linear equation of state, thus ignoring the role of mixing processes and subtleties associated with the nonlinear equation of state.

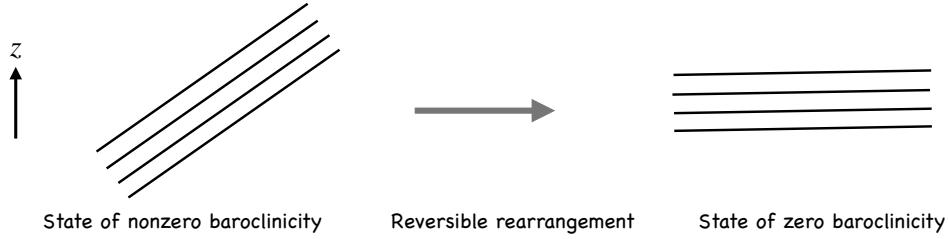


FIGURE 24.2: Isolines of constant buoyancy to illustrate the concept of available potential energy (APE) in a stably stratified Boussinesq fluid. The initial state (left panel) with non-zero baroclinicity is reversibly rearranged to have zero baroclinicity (right panel). The difference in gravitational potential energy between these two states defines the APE in the initial state. As shown in Section 24.8.3, the APE is a non-negative measure of the amount of gravitational potential energy that can, in principle, be reversibly converted to kinetic energy. Furthermore, note that the depth of a buoyancy surface in the background state shown in the right panel equals to the area average depth of the same buoyancy surface in the left panel (see Section 24.8.2).

24.8.1 Analytic continuation of buoyancy surfaces

We use the term *outcrop* to refer to the vanishing of a buoyancy surface at the upper boundary, and *incrop* when it vanishes at the lower boundary. Figure 24.3 illustrates such surfaces. When considering a fluid in a domain with geometric boundaries, and when describing properties of the fluid according to thermodynamic coordinates such as buoyancy, we must decide how to describe these surfaces in regions where they do not exist; i.e., where they are outcropped or incropped. We follow the *Lorenz convention* described in Lorenz (1955), Andrews (1983), Section 4 of Young (2012), and Appendix A of Ringler et al. (2017).

This goal might appear to be pointless; i.e., if the surface does not exist in a region, then why do we need to specify any of its properties. However, the “analytic continuation” of buoyancy surfaces is very useful when developing their kinematics, with particular use for available potential energy in Section 24.8.3. Such concerns have further applications for studies of water mass transformation arising from boundary buoyancy fluxes (e.g., Nurser et al. (1999) as well as Chapter 52). Additionally, as we will see, the buoyancy frequency along these surfaces is formally infinite since the extended buoyancy surfaces are squeezed into a zero thickness layer. Doing so thus creates an infinite potential vorticity, which corresponds to the potential vorticity *delta sheets* as discussed by Bretherton (1966) and Schneider et al. (2003). In the following we limit our attention to domains with flat bottoms and vertical side-walls to minimize the niceties that arise with more general domains.

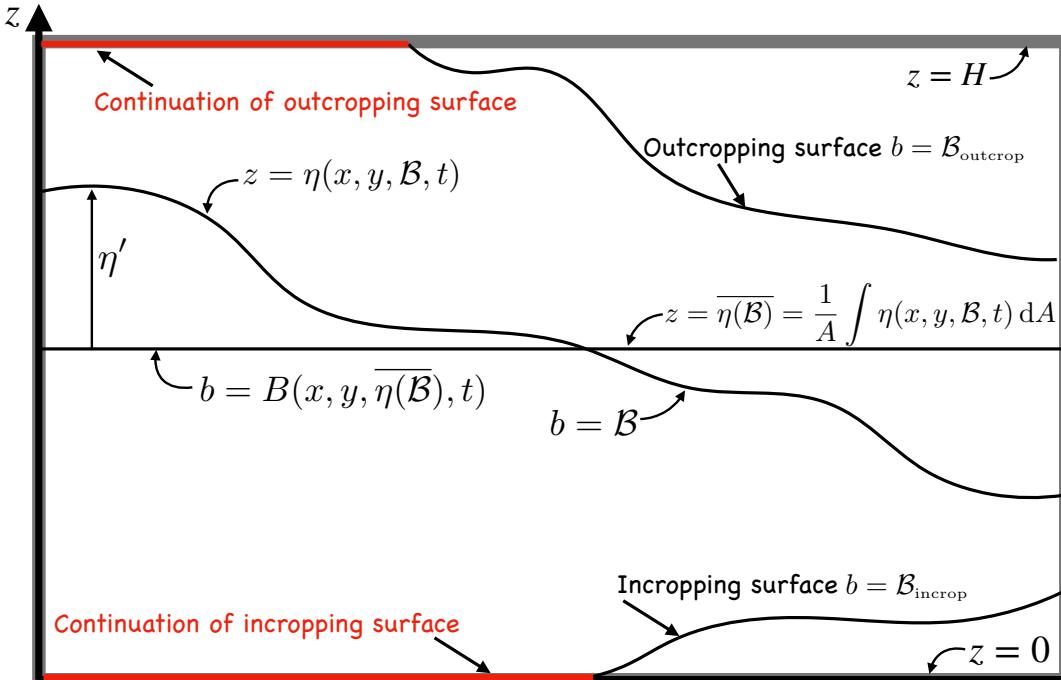


FIGURE 24.3: Geometry of buoyancy surfaces in a flat bottom box of perfect Boussinesq fluid with height H and horizontal area $\int dA = A$. Three representative buoyancy surfaces are shown: one that spans the full domain with $b = \mathcal{B}$, one that incrops at the bottom with $b = \mathcal{B}_{\text{incrop}}$, and one that outcrops at the surface with $b = \mathcal{B}_{\text{outcrop}}$. The vertical height of a buoyancy surface is $z = \eta(x, y, \mathcal{B}, t)$; its area average (which is time independent; see Section 24.8.2) is $\bar{\eta}(\mathcal{B}) = A^{-1} \int \eta(x, y, \mathcal{B}, t) dA$; and its corresponding anomalous height is $\eta'(x, y, \mathcal{B}, t) = \eta(x, y, \mathcal{B}, t) - \bar{\eta}(\mathcal{B})$. As a complement, a particular point on the $z = \bar{\eta}(\mathcal{B})$ height surface has buoyancy $b = B(x, y, \bar{\eta}(\mathcal{B}), t)$, which then leads to an anomalous buoyancy $b'(x, y, \bar{\eta}(\mathcal{B}), t) = B(x, y, \bar{\eta}(\mathcal{B}), t) - \mathcal{B}$, where $B(x, y, \bar{\eta}(\mathcal{B}), t) = \mathcal{B}$ (see equation (24.144)). To allow the formalism to be transparent across all buoyancy surfaces, we set $\eta(x, y, \mathcal{B}_{\text{outcrop}}, t) = H$ in regions where the surface has outcropped, and $\eta(x, y, \mathcal{B}_{\text{incrop}}, t) = 0$ where the surface has incropped (denoted by the red lines). As a complement, we set $\partial z / \partial b = 1/N^2 = 0$ for regions where the surface has either incropped or outcropped, thus formally imposing an infinitely stratified extension of the incropped and outcropped surfaces across the top and bottom domain boundaries. Through this analytic continuation of the buoyancy surfaces, we are ensured that the area mean height of all buoyancy surfaces forms a monotonic sequence from 0 to H , with $\bar{\eta}(\mathcal{B}_1) > \bar{\eta}(\mathcal{B}_2)$ if $\mathcal{B}_1 > \mathcal{B}_2$. When focused on a single buoyancy surface, we can reduce notational clutter by writing, for example, $\bar{\eta}$ rather than $\eta(\mathcal{B})$, as well as $b'(\bar{\eta}) = B(\bar{\eta}) - \mathcal{B}$, and $\eta' = \eta - \bar{\eta}$.

Buoyancy-area mean height of a buoyancy surface

A kinematic property that we make use of when formulating available potential energy is the area mean height of a buoyancy surface. One way to define this mean is to integrate the height over the area of the buoyancy surface and then divide by the area of the buoyancy surface

$$\bar{\eta}(\mathcal{B}, t)^{\text{buoyancy}} = \frac{\int_{\mathcal{B}} \eta(x, y, \mathcal{B}, t) dS}{\int_{\mathcal{B}} dS}, \quad (24.137)$$

where $\int_{\mathcal{B}} dS$ is the area integral over the \mathcal{B} buoyancy surface. There are two problems with this area mean. First, the area of a buoyancy surface is rather complicated to compute in practice. Second, the buoyancy surface area is time dependent thus making the area mean also time dependent.

Domain-area mean height of a buoyancy surface

An alternative method to compute the area mean height is to integrate over the area of the fluid domain

$$\overline{\eta(\mathcal{B})} = \frac{\int \eta(x, y, \mathcal{B}, t) dA}{\int dA}, \quad (24.138)$$

where

$$A = \int dA = \int_{\text{fluid domain}} dx dy \quad (24.139)$$

is the time-independent horizontal area of the fluid domain. However, in choosing the domain area we must specify the height of a buoyancy surface in those horizontal regions where the surface does not exist; i.e., where the surface outcrops or incrops. By doing so, we see in Section 24.8.2 that the area mean height is time-independent for all buoyancy surfaces. Furthermore, this area mean height satisfies the monotonicity property

$$\overline{\eta(\mathcal{B}_1)} > \overline{\eta(\mathcal{B}_2)} \quad \text{if} \quad \mathcal{B}_1 > \mathcal{B}_2. \quad (24.140)$$

We make use of both the time-independence of the mean height and the monotonicity property when formulating the available potential energy in Section 24.8.3.

Analytic continuation of surface height at outcrops and incrops

So how do we specify the height in outcrop regions? Let us motivate a specification by considering a buoyancy surface that sits near the top of the domain; i.e., its buoyancy is near the domain maximum, b_{\max} . Assume this surface is not horizontal, with the surface $b = \mathcal{B}_{\text{outcrop}}$ in Figure 24.3 an example. Furthermore, let it cover less horizontal area than the full domain area. If we horizontally integrate just over regions where the surface does not outcrop, but still normalize by the total horizontal area of the domain, then the area mean height will be less than certain other buoyancy surfaces whose buoyancy is less and yet whose horizontal area is more. As a result we will not satisfy the monotonicity property (24.140). A way to recover monotonicity is to analytically continue the buoyancy surface along the upper boundary so that its height in the outcropped region is set to $\eta(x_{\text{outcrop}}, y_{\text{outcrop}}, \mathcal{B}_{\text{outcrop}}, t) = H$. Doing so then ensures that the domain-area mean height for buoyancy surfaces will approach H as their buoyancy approaches the maximum buoyancy. We provide an analogous continuation of the surface within the bottom boundary so that

$$\eta(x, y, \mathcal{B}, t) = \begin{cases} H & \text{if } (x, y) \in \text{outcrop region} \\ 0 & \text{if } (x, y) \in \text{incrop region} \\ \eta(x, y, \mathcal{B}, t) & \text{otherwise.} \end{cases} \quad (24.141)$$

These two continuations of the buoyancy surfaces ensures that the domain-area mean height of all buoyancy surfaces forms a monotonic sequence and that the sequence extends from $0 \leq \overline{\eta(\mathcal{B})} \leq H$.

Analytic continuation of buoyancy stratification at outcrops and incrops

What does the analytic continuation (24.141) imply for buoyancy? As described, we allow all outcropped buoyancy surfaces to continue along the surface at $\eta = H$. All outcropped surfaces are thus squeezed into the infinitesimal upper fluid layer with buoyancy in that layer bounded above by the domain maximum buoyancy, b_{\max} . Likewise, for the bottom of the domain we squeeze all incropped buoyancy surfaces into an infinitesimal layer bounded below by b_{\min} , the minimum buoyancy in the domain. Consequently, the upper and lower boundaries are formally capped by infinitely stratified shells in which the inverse squared buoyancy frequency vanishes.

24.8.2 The dual relation between height and buoyancy

In deriving an expression for the APE in Section 24.8.3, we will find it useful to have relations between the unsorted and sorted buoyancy fields. We will also make use of the dual relation between the height of a constant buoyancy surface and the buoyancy of a constant height surface. For this purpose we examine certain kinematic properties of buoyancy surfaces in a stably stratified box of perfect Boussinesq fluid as in Figure 24.3.

Volume beneath a buoyancy surface using height coordinates

Making use of notation from Figure 24.3, the volume of fluid contained beneath an arbitrary buoyancy surface is

$$V(\mathcal{B}) = \int dA \int_0^{\eta(x,y,\mathcal{B},t)} dz = \int \eta(x, y, \mathcal{B}, t) dA = A \overline{\eta(\mathcal{B})}, \quad (24.142)$$

The following properties result from volume conservation in a perfect non-divergent flow (see Chapter 17).

- The volume of fluid beneath an arbitrary buoyancy surface is time-independent, as is the area mean height of this surface. This property allowed us to drop the time argument from $V(\mathcal{B})$ and $\eta(\mathcal{B})$ in equation (24.142).
- The area mean height of a buoyancy surface is identical to the height of the surface when it is reversibly rearranged to be horizontal.

To verify these properties, recall that buoyancy surfaces are material in a perfect fluid so that no fluid crosses them even as they fluctuate. It follows that the volume of fluid beneath an arbitrary buoyancy surface is time-independent. Since the horizontal area of the domain is time-independent, equation (24.142) also means that the area averaged height of the buoyancy surface is time-independent. Furthermore, any motion of a buoyancy surface in a perfect fluid is reversible, including motion that flattens the surface. Since its area mean height remains fixed, the area mean equals to the height of the surface when it is flat.

Area mean buoyancy on a constant depth surface

As a further realization of the dual relation between height and buoyancy, note that the area average buoyancy, $b = B(x, y, z, t)$ along a constant height surface is also constant in time

$$\overline{B(z)} = A^{-1} \int B(x, y, z, t) dA = \text{time independent.} \quad (24.143)$$

This property follows since both buoyancy and volume are material constants following a fluid parcel in a perfect Boussinesq fluid. Hence, a fluid parcel carries both its buoyancy and volume unchanged so that the volume integrated buoyancy within any fluid region remains constant. Correspondingly, the area integrated buoyancy along any fixed height surface remains constant. It also follows that the area mean buoyancy at $z = \overline{\eta(\mathcal{B})}$ is \mathcal{B}

$$\overline{B[\overline{\eta(\mathcal{B})}]} = A^{-1} \int B(x, y, z = \overline{\eta(\mathcal{B})}, t) dA = \mathcal{B}. \quad (24.144)$$

Volume beneath a buoyancy surface using buoyancy coordinates

Let us return to the volume beneath a buoyancy surface, only now use buoyancy coordinates to write

$$V(\mathcal{B}) = \int dA \int_0^{\eta(x,y,\mathcal{B},t)} dz = \int dA \int_{b(x,y,0,t)}^{\mathcal{B}} \frac{\partial z}{\partial b} db = \int dA \int_{b(x,y,0,t)}^{\mathcal{B}} \frac{db}{N^2} \quad (24.145)$$

where $b(x, y, 0, t)$ is the buoyancy at the bottom of the domain and

$$N^2 = \frac{\partial b}{\partial z} \quad (24.146)$$

is the squared buoyancy frequency. As noted in Section 24.8.1 and illustrated in Figure 24.3, we analytically continue the buoyancy surfaces into the surface and bottom boundaries so that

$$N^{-2}(x, y, \mathcal{B}) = \begin{cases} = 0 & \text{if } \mathcal{B} > b(x, y, H) \text{ (surface outcrop region)} \\ = 0 & \text{if } \mathcal{B} < b(x, y, 0) \text{ (bottom incrop region)} \\ = N^{-2}(x, y, \mathcal{B}) & \text{if } b(x, y, 0) \leq \mathcal{B} \leq b(x, y, H). \end{cases} \quad (24.147)$$

In this manner we can replace the lower limit in equation (24.145) with a constant buoyancy well below any buoyancy found in the domain, which we write as b_{\min} , so that

$$V(\mathcal{B}) = \int dA \int_{b_{\min}}^{\mathcal{B}} \frac{db}{N^2} = \int_{b_{\min}}^{\mathcal{B}} db \int dA \frac{1}{N^2}. \quad (24.148)$$

Being able to commute the area and buoyancy integrals proves useful in the following.

24.8.3 Exact expression for APE

In this section we develop an expression for the APE of the perfect stably stratified Boussinesq fluid in a box. To start, consider the gravitational potential energy of the fluid in Figure 24.3 relative to a constant density background state with $\rho = \rho_0$

$$\mathcal{P} = - \int dA \int_0^H b z dz = -\frac{1}{2} \int dA \int_0^H b d(z^2). \quad (24.149)$$

Integration by parts leads to the equivalent expression

$$\mathcal{P} = -\frac{1}{2} \int dA \int_0^H d(b z^2) + \frac{1}{2} \int dA \int_{b(x,y,0)}^{b(x,y,H)} \eta^2(x, y, b) db \quad (24.150)$$

$$= -\frac{A H^2}{2} \overline{b(H)} + \frac{1}{2} \int dA \int_{b(x,y,0)}^{b(x,y,H)} \eta^2(x, y, b) db, \quad (24.151)$$

where $\overline{b(H)}$ is the area averaged buoyancy at the top of the fluid domain, $z = H$. As discussed in Section 24.8.1, integration over the finite domain using a buoyancy coordinate leads us to set $\overline{b(H)} = b_{\max}$, the domain maximum buoyancy. Likewise, the second expression in equation (24.151) has its buoyancy integral range extended to b_{\min} and b_{\max} . By doing so we can swap the area and buoyancy integrals to render

$$\mathcal{P} = \frac{A}{2} \left[-H^2 b_{\max} + \int_{b_{\min}}^{b_{\max}} \overline{\eta^2(b)} db \right], \quad (24.152)$$

where $\overline{\eta^2(b)}$ is the area mean of the squared height of a buoyancy surface. The same calculation for the reference buoyancy, $b_{\text{ref}}(z)$, leads to

$$\mathcal{P}_{\text{ref}} = \frac{A}{2} \left[-H^2 b_{\max} + \int_{b_{\min}}^{b_{\max}} \overline{\eta(b)^2} db \right], \quad (24.153)$$

where we noted that the height of a reference buoyancy surface equals to the area mean of the corresponding buoyancy surface

$$\eta(b_{\text{ref}}) = \bar{\eta}(b = b_{\text{ref}}), \quad (24.154)$$

and the reference buoyancy at the surface boundary equals to the maximum buoyancy, $b_{\text{ref}}(H) = b_{\max}$.

Subtracting the gravitational potential energy of the reference/background state from the potential energy of the full state renders an expression for the available potential energy

$$\mathcal{P}_{\text{APE}} = \mathcal{P} - \mathcal{P}_{\text{ref}} = \frac{A}{2} \left[\int_{b_{\min}}^{b_{\max}} [\overline{\eta^2(b)} - \overline{\eta(b)}^2] db \right] = \int_{b_{\min}}^{b_{\max}} \overline{(\eta')^2} db \geq 0, \quad (24.155)$$

where (see Figure 24.3)

$$\eta(x, y, b, t) = \overline{\eta(b)} + \eta'(x, y, b, t). \quad (24.156)$$

The positive definite nature of the APE arises since either a positive or negative buoyancy surface undulation gives rise to motion.

Equation (24.155) is an exact expression for the APE of a perfect Boussinesq fluid in a flat bottom domain. We encounter the same expression when studying the APE in a shallow water fluid in Section 32.5.4. It is the natural expression when working in buoyancy coordinates, whereby the APE is determined by height variations of constant buoyancy surfaces.

24.8.4 Approximate expression for APE

When working with geopotential coordinates it is useful to obtain an approximate expression for the APE in terms of buoyancy variations on constant height surfaces. That is the subject of this subsection.

Approximate version of APE in terms of buoyancy fluctuations

To develop an approximate expression for the APE we write the height of a buoyancy surface $\eta(\mathcal{B})$ in the form (see caption to Figure 24.3 if confused by signs)

$$\eta(\mathcal{B}) \approx \overline{\eta(\mathcal{B})} + \left[\frac{\partial z}{\partial b} \right]_{z=\bar{\eta}} [\mathcal{B} - B(\bar{\eta})] = \bar{\eta} - \frac{b'}{N^2} \approx \bar{\eta} - \frac{b'}{N_{\text{ref}}^2}, \quad (24.157)$$

where the final step set $N^2(x, y, \bar{\eta}) \approx N_{\text{ref}}^2(\bar{\eta})$, which is valid to the same order as the approximation. We are thus lead to the approximate expression

$$\eta' = \eta - \bar{\eta} \approx -\frac{b'}{N_{\text{ref}}^2} \quad (24.158)$$

so that the APE is given approximately by

$$\mathcal{P}_{\text{APE}} \approx A \left[\int_0^H \frac{\overline{(b')^2}}{2 N_{\text{ref}}^2} dz \right], \quad (24.159)$$

This approximate expression is commonly used in practical calculations of APE, particularly when making use of field measurements (e.g., [Bishop et al. \(2020\)](#)).

Practical issues related to the sorted buoyancy profile

Figure 24.4 illustrates how to obtain the sorted buoyancy profile from a discretized version of a stably stratified fluid. The buoyancy of each cell is compared to that of all other cells and vertically stacked according to the relative buoyancy. The vertical position of the sorted grid cell is determined by accumulating the volume per horizontal area of the cell, starting from the bottom and moving up.

It is notable that cells with identical buoyancy lead to regions of zero vertical stratification in the sorted buoyancy profile. Such zero stratification regions commonly arise when sorting stratified fluid layers, where the buoyancy is constant within the layers. One is thus led to perform a vertical smoothing of the sorted profile to remove such unstratified regions, particularly if using the profile to define a background buoyancy frequency as required for the approximate APE calculation given by equation (24.159).

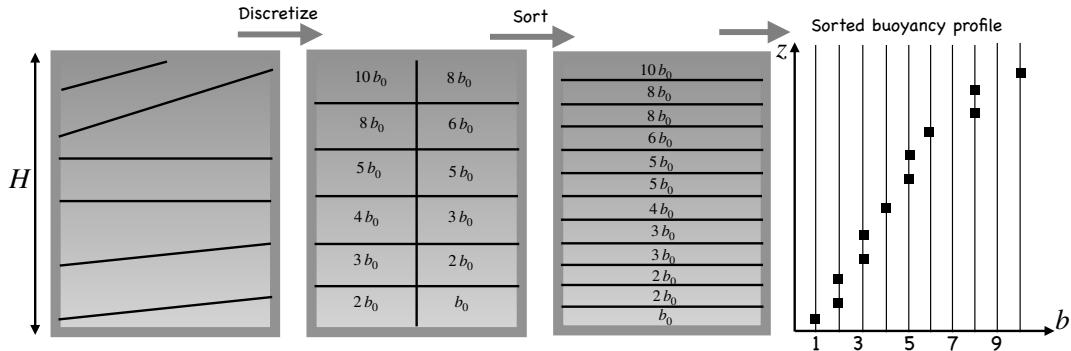


FIGURE 24.4: Illustrating how to sort buoyancy to determine the background or reference profile for computing APE. The first panel on the left shows a sample buoyancy field with black lines representing buoyancy isolines. The second panel shows a discretized version of the field, with b_0 a unit of buoyancy and each cell's buoyancy an integer multiple of b_0 . For simplicity we assume the horizontal area of the domain is depth independent and that each of the discrete grid cells has the same volume and horizontal area. The third panel shows the result of sorting the discrete buoyancy field, with the most buoyant fluid above the less buoyant fluid. During the sort, the cell's volume remains constant (Boussinesq fluid) and the accumulated volume per horizontal area determines the vertical position of the sorted cell. The final panel shows the sorted profile $b_{\text{ref}}(z)$. Note that regions of zero lateral buoyancy gradient in the unsorted buoyancy field lead to vertically unstratified regions in the sorted buoyancy.

Budget for approximatae APE

To develop a budget for the approximate form of APE, start by considering the budget for buoyancy as decomposed as

$$b(x, y, z, t) = b_{\text{ref}}(z) + b'(x, y, z, t), \quad (24.160)$$

so that the perfect fluid buoyancy equation takes on the form

$$\frac{Db}{Dt} = 0 \implies \frac{Db'}{Dt} = -w N_{\text{ref}}^2. \quad (24.161)$$

Multiplying by b' leads to

$$\frac{D[(b')^2/2]}{Dt} = -w b' N_{\text{ref}}^2, \quad (24.162)$$

and then dividing by N_{ref}^2 renders

$$\frac{\partial P_{\text{APE}}}{\partial t} + \mathbf{v} \cdot \nabla P_{\text{APE}} = -w b' \left[1 - \frac{b'}{2} \frac{\partial (1/N_{\text{ref}}^2)}{\partial z} \right], \quad (24.163)$$

where we defined the approximate APE per unit volume

$$P_{\text{APE}} = (1/2) [b'/N_{\text{ref}}]^2. \quad (24.164)$$

In the case of a depth-independent reference buoyancy frequency, we see that the APE per unit volume materially evolves with a source $-w b'$, which is analogous to the potential energy evolution where the source is $-w b$. Now adding equation (24.163) to the kinetic energy equation (24.109) with $\Phi = g z$ leads to

$$\frac{\partial(\mathcal{K} + P_{\text{APE}})}{\partial t} + \mathbf{v} \cdot \nabla(\mathcal{K} + P_{\text{APE}} + p/\rho_0) = -w \left[g - b_{\text{ref}} - \frac{(b')^2}{2} \frac{\partial(1/N_{\text{ref}}^2)}{\partial z} \right]. \quad (24.165)$$

Note that a global area average on any surface eliminates the $w(g - b_{\text{ref}})$ term since

$$\bar{w} = A^{-1} \int w \, dA = 0, \quad (24.166)$$

which follows from $\nabla \cdot \mathbf{v} = 0$ (see Exercise 17.4). Further simplifications arise with depth-independent N_{ref} , with the corresponding space-time spectra studied by [Bühler et al. \(2014\)](#).

24.8.5 Comments

Elements of this section follow from Section 3.11.1 of [Vallis \(2017\)](#). APE has remained a compelling notion throughout geophysical fluid studies. Unfortunately, for the ocean it has proven very difficult to extend its formalism beyond the perfect fluid Boussinesq system considered here. Particular difficulties arise from the nonlinear equation of state for seawater and the nontrivial ocean geometry with distinct basins and enclosed seas.



24.9 Exercises

EXERCISE 24.1: SYMMETRY UNDER A TIME-DEPENDENT TRANSLATION

In this exercise we consider the Euler equation in free space (no boundaries) where we focus only on the acceleration and the pressure gradient force. That is, we ignore any body forces from gravity and rotation so that the Euler equation takes on the form

$$\rho_0 \frac{D\mathbf{v}}{Dt} = -\nabla p, \quad (24.167)$$

where we assume a Boussinesq fluid so that $\nabla \cdot \mathbf{v} = 0$.

Consider a shift in the reference frame used to describe the flow so that a coordinate position shifts according to

$$\mathbf{x} \rightarrow \mathbf{x} + \mathbf{c}(t), \quad (24.168)$$

where the vector \mathbf{c} is time dependent but has the same value for all points in space.

- (a) Is the shift (24.168) a Galilean transformation? Hint: recall the discussion of Galilean transformation in Section 13.6.
- (b) What happens to pressure in the new reference frame? Hint: consider the elliptic problem for pressure as discussed in Section 24.3.1.

(c) Write the equation of motion (24.167) in this new reference frame.

EXERCISE 24.2: STEADY PARALLEL SHEARED FLOW

Consider the non-divergent parallel sheared flow on a tangent plane

$$\mathbf{v} = \hat{\mathbf{x}} u(y). \quad (24.169)$$

That is, the steady velocity is zonal and has a meridional dependence.

- (a) Show that this flow is an exact solution to the β -plane inviscid Boussinesq velocity equation.
- (b) What is the corresponding pressure gradient?
- (c) Show that for an f -plane the stationary velocity can have an arbitrary orientation. Hint: it is sufficient to show that $\mathbf{v} = \hat{\mathbf{y}} v(x)$ is a stationary solution.

EXERCISE 24.3: A GENERALIZED BOUSSINESQ APPROXIMATION

In this exercise we derive a mild generalization to the Boussinesq approximation. This generalization facilitates a more accurate decomposition of pressure by introducing a new reference density, $\bar{\rho}(z)$, that is a static function of depth. The new decomposition also leads to a slightly modified buoyancy field.

The derivation starts with the usual decomposition of density using ρ_0 as a space-time constant

$$\rho(\mathbf{x}, t) = \rho_0 + \rho'(\mathbf{x}, t), \quad (24.170)$$

thus leading to the Boussinesq momentum equation

$$\frac{D\mathbf{v}}{Dt} + 2\boldsymbol{\Omega} \wedge \mathbf{v} = -\rho_0^{-1} (\nabla p + \hat{\mathbf{z}} g \rho). \quad (24.171)$$

But rather than take the traditional decomposition of the pressure and gravitational terms, we now write

$$\nabla p + \hat{\mathbf{z}} g \rho = \nabla[p - \bar{\rho}(z) + \bar{p}(z)] + \hat{\mathbf{z}} g [\rho - \bar{\rho}(z) + \bar{\rho}(z)] \quad (24.172a)$$

$$= \nabla[p - \bar{\rho}(z)] + \hat{\mathbf{z}} g [\rho - \bar{\rho}(z)]. \quad (24.172b)$$

This step introduced the density, $\bar{\rho}(z)$, and the corresponding hydrostatically balanced pressure

$$\frac{d\bar{\rho}(z)}{dz} = -g \bar{\rho}(z). \quad (24.173)$$

- (a) Show that the pressure is decomposed as

$$p(\mathbf{x}, t) = [p(\mathbf{x}, t) - \bar{\rho}(z)] + \bar{p}(z) = \rho_0 \tilde{\varphi} + \bar{p}(z). \quad (24.174)$$

What is $\tilde{\varphi}$?

- (b) Introduce a buoyancy

$$\tilde{b} = -\frac{g(\rho - \bar{\rho})}{\rho_0}, \quad (24.175)$$

defined relative to the depth-dependent background density, $\bar{\rho}(z)$, rather than the globally constant density ρ_0 . Show that the momentum equation is given by

$$\frac{D\mathbf{v}}{Dt} + 2\boldsymbol{\Omega} \wedge \mathbf{v} = -\nabla \tilde{\varphi} + \hat{\mathbf{z}} \tilde{b}. \quad (24.176)$$

- (c) Show that the baroclinicity vector appearing in the Boussinesq vorticity equation takes the form

$$\tilde{\mathbf{B}} = \nabla \tilde{b} \wedge \hat{\mathbf{z}}, \quad (24.177)$$

which is mathematically the same as with the traditional Boussinesq approximation from Section 24.1.2. We have thus succeeded in generalizing the pressure decomposition and buoyancy field, yet without corrupting the familiar Boussinesq vorticity dynamics.

EXERCISE 24.4: KINETIC ENERGY FOR A PERFECT HYDROSTATIC BOUSSINESQ FLUID

Consider a perfect *hydrostatic* Boussinesq fluid. Show that the kinetic energy per mass contained in the horizontal velocity,

$$\mathcal{K}_{\text{horz}} = \frac{\mathbf{u} \cdot \mathbf{u}}{2}, \quad (24.178)$$

satisfies the same equation as that for a non-hydrostatic fluid given by equation (24.108). Assume the simple form for the geopotential, $\Phi = g z$.

EXERCISE 24.5: ENERGETICS FOR A PERFECT BOUSSINESQ FLUID

In Section 24.6 we developed the energetic balances for an ideal Boussinesq fluid in a closed domain. We here rederive the same energetics but using the momentum, buoyancy, and continuity equations in the form that exposes buoyancy

$$\frac{D\mathbf{v}}{Dt} + 2\boldsymbol{\Omega} \wedge \mathbf{v} = -\nabla\varphi + \hat{\mathbf{z}} b \quad (24.179a)$$

$$\frac{Db}{Dt} = 0 \quad (24.179b)$$

$$\nabla \cdot \mathbf{v} = 0 \quad (24.179c)$$

$$b = b(S, \Theta). \quad (24.179d)$$

Assume the simple form for the geopotential, $\Phi = g z$, and assume a closed and static domain (i.e., an Eulerian domain with no boundary contributions). To help physically interpret terms, remember to isolate the total divergence terms and the remainder.

- (a) Derive the material time evolution equation for the kinetic energy.
- (b) Derive the material time evolution equation for the gravitational potential energy. For this purpose it is convenient to introduce $P^b = -g^{-1} \Phi b$ with $\Phi = g z$. Interpret P^b .
- (c) Derive the mechanical energy equation written in its Eulerian flux-form. Discuss.
- (d) Integrate the mechanical energy equation to derive a global domain budget for mechanical energy. Discuss.

EXERCISE 24.6: ENERGETICS FOR A DISSIPATIVE BOUSSINESQ FLUID

We here examine the energy budget for a Boussinesq fluid in a rotating frame under the Traditional approximation and including diabatic forcing, momentum mixing from molecular viscosity, and buoyancy mixing from molecular diffusivity. The equations for this fluid system are

$$\frac{D\mathbf{v}}{Dt} + f \hat{\mathbf{z}} \wedge \mathbf{v} = -\nabla\varphi + \hat{\mathbf{z}} b + \nu \nabla^2 \mathbf{v} \quad (24.180a)$$

$$\frac{Db}{Dt} = Q^b + \kappa \nabla^2 b \quad (24.180b)$$

$$\nabla \cdot \mathbf{v} = 0. \quad (24.180c)$$

In the above, $\nu > 0$ is the viscosity leading to irreversible mixing of velocity, and $\kappa > 0$ is the diffusivity leading to irreversible mixing of buoyancy.

- Assume both ν and κ are constant for this exercise; e.g., they are molecular values.
- Assume a linear equation of state so that buoyancy is linearly related to temperature with a constant thermal expansion coefficient (equation (24.24f) with haline contraction coefficient $\beta = 0$).
- Assume the domain has static boundaries so that it has a fixed volume.
- When performing domain integrated budgets, ignore boundary terms that appear as a result of the divergence theorem. Even though boundary terms are generally important, we here are not focused on them.
- Let Q^b be a three-dimensional buoyancy source and/or boundary term; e.g., surface buoyancy fluxes, geothermal heating, interior sources. That is, we choose to isolate the boundary and source terms in Q^b so that $\hat{\mathbf{n}} \cdot \nabla b = 0$ along all boundaries. This construct allows us to contemplate placing buoyancy sources at various places within the fluid column and determining how those sources affect the energy budget.
- Hint: Elements of this exercise are discussed in Chapter 21 of [Vallis \(2017\)](#).
- To help physically interpret terms, remember to isolate the total divergence terms and the remainder.

- (a) Derive the material evolution of kinetic energy per mass.
(b) Derive the equation for domain integrated evolution of kinetic energy per mass. Make use of the identity

$$\nabla^2 \mathbf{v} = -\nabla \wedge \boldsymbol{\omega} \text{ if } \nabla \cdot \mathbf{v} = 0 \quad (24.181)$$

to express the viscosity contributions in terms of the vorticity

$$\boldsymbol{\omega} = \nabla \wedge \mathbf{v}. \quad (24.182)$$

- (c) Show that the domain integrated kinetic energy per mass is reduced (dissipated) by the viscosity in the presence of a nonzero vorticity. Discuss this result.
(d) Derive the material evolution equation for $P^b = -g^{-1} \Phi b$ with $\Phi = g z$. Interpret P^b .
(e) Derive the equation for domain integrated evolution of gravitational potential energy per mass.
- Discuss how downgradient vertical diffusion of buoyancy impacts on the domain integrated gravitational potential energy.
 - Discuss how diabatic heating impacts on the domain integrated gravitational potential energy.
- (f) Discuss the reversible conversion between kinetic and gravitational potential energy. For example, what happens to the kinetic and gravitational potential energies if a heavy fluid element moves upward?

- (g) Derive the equation for material evolution of mechanical energy per mass.
- (h) Derive the domain integrated evolution of mechanical energy per mass.
- (i) Over a closed volume (no boundary fluxes), show that the dissipation of domain integrated kinetic energy is balanced by the buoyancy source.
- (j) In a steady state (Eulerian time derivative vanishes) and absent buoyancy diffusion, show that diabatic heating from Q^b must occur at a lower level (lower gravitational geopotential) than cooling if a kinetic energy dissipating circulation is to be maintained.

EXERCISE 24.7: KINETIC ENERGY AND THE HYDROSTATIC BOUSSINESQ EQUATIONS

In this exercise we develop some properties of the kinetic energy for the hydrostatic Boussinesq equations listed in Section 24.1.6. We here assume the horizontal frictional acceleration is determined by vertical viscous friction in equation (24.98), and the stress boundary conditions are given by equations (24.100) and (24.102).

- (a) Derive the Eulerian flux-form expression for the kinetic energy budget.
- (b) Why does the kinetic energy only have contributions from the horizontal velocity components?
- (c) Discuss the role of vertical viscosity in transporting kinetic energy in the vertical.
- (d) Discuss the role of vertical viscosity in dissipating kinetic energy.
- (e) Discuss how wind stress and bottom drag impact the globally integrated kinetic energy. Assume bottom drag in the form of equation (24.102).

EXERCISE 24.8: POTENTIAL ENERGY AND THE HYDROSTATIC BOUSSINESQ EQUATIONS

In this exercise we develop some properties of the gravitational potential energy for the hydrostatic Boussinesq equations stated in Section 24.1.6.

- (a) Derive the Eulerian flux-form budget for gravitational potential energy written as $P^b = -g^{-1} \Phi b$ with $\Phi = g z$. Interpret P^b .
- (b) Discuss the role of the subgrid scale eddy-induced advection in this budget as given by equation (24.94). In particular, discuss its impact on the center of mass of the fluid.
- (c) Discuss the role of vertical diffusion in this budget as given by equation (24.94). In particular, discuss its impact on the center of mass of the fluid.
- (d) Integrate the gravitational potential energy budget over the global ocean. Discuss how the surface boundary buoyancy flux, Q_b , impacts on the global potential energy budget through impacts on the center of mass of the fluid. Ignore any bottom geothermal heating.

EXERCISE 24.9: BUOYANCY AND THE OCEAN MODEL EQUATIONS

In this exercise we develop some properties of the squared buoyancy for the hydrostatic Boussinesq equations stated in Section 24.1.6.

- (a) Write the Eulerian flux-form budget describing the evolution of b^2 , the squared buoyancy.
- (b) Discuss the impacts from vertical diffusion on the b^2 budget.

EXERCISE 24.10: PARAMETERIZED EDDY VELOCITY AND THE OCEAN MODEL EQUATIONS

In this exercise we develop some implications of assuming a specific form for the parameterized eddy velocity for the hydrostatic Boussinesq equations stated in Section 24.1.6. Namely, we consider the specific form for the parameterized eddy-induced velocity proposed by [Gent et al. \(1995\)](#)

$$\mathbf{u}^* = -\partial_z(B \mathbf{S}) \quad (24.183a)$$

$$w^* = \nabla_z \cdot (B \mathbf{S}) \quad (24.183b)$$

$$\mathbf{S} = -\frac{\nabla_z b}{N^2} \quad (24.183c)$$

$$\mathbf{v}^* \cdot \hat{\mathbf{n}} = 0 \quad \text{at all ocean boundaries.} \quad (24.183d)$$

In this expression, $B > 0$ is an eddy diffusivity. To ensure $\mathbf{v}^* \cdot \hat{\mathbf{n}} = 0$ at all domain boundaries requires that $B = 0$ along these boundaries. The horizontal vector $\mathbf{S} = (S^{(x)}, S^{(y)}, 0)$ measures the slope of the buoyancy surfaces relative to the horizontal. We assume the ocean is stably stratified in the vertical, so that $\partial b / \partial z = N^2 > 0$.

- (a) Determine the vector streamfunction Ψ^* such that

$$\mathbf{v}^* = \nabla \wedge \Psi^*. \quad (24.184)$$

Choose the gauge with $\hat{\mathbf{z}} \cdot \Psi^* = 0$.

- (b) Show that

$$\int_{-H}^{\eta} \mathbf{u}^* dz = 0. \quad (24.185)$$

That is, the parameterized horizontal flow has a zero depth integral.

- (c) At any chosen meridional position y , the meridional buoyancy transport from advection (resolved plus parameterized) is computed by

$$\mathcal{B}^{(y)}(y, t) = \int_{x_1}^{x_2} dx \int_{-H}^{\eta} b(v + v^*) dz. \quad (24.186)$$

The zonal and vertical integrals are over the full zonal and vertical extent of the ocean domain. Show that the effects from v^* are to reduce the meridional gradients of buoyancy. That is, if buoyancy decreases poleward, then v^* will flux buoyancy poleward to reduce the gradient.

- (d) How does the introduction of \mathbf{v}^* to the buoyancy equation (24.24c) affect the global integrated *gravitational potential energy*? Discuss.
- (e) How does the introduction of \mathbf{v}^* to the buoyancy equation (24.24c) affect the global integrated *available potential energy*? Discuss.



Buoyancy

A large portion of the vertical pressure force is balanced by the gravitational force, as per a hydrostatic fluid. If there is any unbalanced density-induced pressure force then a fluid element experiences a *buoyant* acceleration that causes vertical motion. More precisely, *buoyancy* is the vertical pressure force acting on a fluid element arising from density inhomogeneities. If these pressure forces are balanced by gravity, then the fluid element is said to be *neutrally buoyant*, whereas if they are unbalanced then the fluid element has a net buoyancy that causes vertical acceleration. Buoyancy is thus a packaging of the vertical forces from gravity and pressure that offers insights into the role of density inhomogeneities in creating vertical fluid motion within a gravitational field.

Our study of buoyancy starts by examining the buoyancy of a *test fluid element* where a test fluid element probes properties of the surrounding fluid environment without altering the environment.¹ We refer to the buoyant force acting on the test fluid element as the *Archimedean buoyancy*, since it offers the natural generalization of the more familiar buoyancy force acting, for example, to keep a ship floating. Examining the buoyancy of test fluid elements throughout a fluid then leads to the notions of gravitational stability and neutral directions.²

Most references to buoyancy in the geophysical literature refer to the Archimedean buoyancy. Even so, Archimedean buoyancy is an incomplete rendering of the forces acting in a fluid element arising from density inhomogeneities. A complete description is provided by the *effective buoyancy*, which accounts for the *static forces* creating vertical accelerations of a fluid element. That is, the effective buoyancy is defined as the vertical acceleration acting on a fluid element that remains when setting all velocity dependent accelerations to zero. The effective buoyancy has a contribution from both Archimedean buoyancy acting on a test fluid element, plus the vertical derivative of a pressure perturbation that depends only on the density field. This extra pressure force can be interpreted as the *back-reaction* on a fluid element due to the surrounding fluid environment; i.e., it accounts for the distinction between a test fluid element and a fluid element.

READER'S GUIDE TO THIS CHAPTER

This chapter builds from our study of pressure in Chapters 18, 19, and 20. We also make use of the equations for a Boussinesq fluid as derived in Chapter 24, in particular the Poisson equation satisfied by pressure in a non-divergent flow.

¹As defined in Section 13.2.5, a test fluid element is directly analogous to the point test mass particles that probe properties of the gravitational field and the point test electrical charges that probe the electromagnetic field, with neither of these test particles altering the fields where they are placed.

²Archimedes was a Greek mathematician, inventor, engineer, physicist, and astronomer who lived in Syracuse, Sicily from roughly 287 B.C.E to 212 B.C.E. Hence, Archimedean buoyancy of a test fluid element builds from perhaps the most ancient of physical concepts used in modern fluid mechanics.

25.1	Loose threads	554
25.2	Archimedes' Principle	554
25.2.1	Hydrostatic pressure force	555
25.2.2	Archimedean buoyancy force	556
25.3	Mass density and its flavors	557
25.3.1	Equation of state for the atmosphere and ocean	557
25.3.2	Modified temperature variables	557
25.3.3	Infinitesimal density increments and material time changes	558
25.3.4	Potential density	559
25.3.5	Linear equation of state for the ocean	560
25.3.6	Further study	560
25.4	Archimedean buoyancy of a test fluid element	560
25.4.1	Locally referenced Archimedean buoyancy	560
25.4.2	Globally referenced Archimedean buoyancy	561
25.5	Buoyancy stratification and neutral directions	561
25.5.1	General considerations	561
25.5.2	Comparing <i>in situ</i> densities	562
25.5.3	Neutral directions and the neutrality condition	563
25.5.4	Buoyancy frequency and gravitational stability	564
25.5.5	Locally referenced potential density	565
25.5.6	Gravitational stability of an ideal gas atmosphere	565
25.5.7	Comments and further study	567
25.6	Neutral helicity	568
25.6.1	Mathematical preliminaries	568
25.6.2	Helical nature of neutral displacements	568
25.6.3	Comments and further study	569
25.7	Effective buoyancy and vertical accelerations	569
25.7.1	Vertical motion in approximately hydrostatic fluids	570
25.7.2	Two decompositions of pressure and gravity forces	570
25.7.3	Vertical equation of motion	571
25.7.4	Effective buoyancy and motional vertical acceleration	571
25.7.5	Poisson equations for the pressures and accelerations	572
25.7.6	Thought experiments for effective buoyancy	574
25.7.7	Comments and further study	575
25.8	Exercises	575

25.1 Loose threads

- Provide some examples of b_{eff} fields in Section 25.7 as taken from [Tarshish et al. \(2018\)](#).
- More on the mathematics of neutral helicity in Section 25.6.

25.2 Archimedes' Principle

The Archimedean buoyancy acting on a massive body immersed in a static fluid is the vertical acceleration due to the gravitational acceleration of the body relative to the gravitational acceleration of the fluid displaced by the body. A body's Archimedean buoyancy is proportional to its density relative to that of the displaced fluid. Hence, there is no Archimedean buoyancy without gravity

nor is there a buoyant acceleration without density differences. These statements of *Archimedes' Principle* are a direct consequence of hydrostatics.

Archimedean buoyancy has wide applications for the study of bodies immersed within fluids; e.g. mechanics of marine organisms, ships, submarines, hot air balloons. Our interests concern the buoyancy of the fluid within itself, with this buoyancy fundamentally related to the density field. To extend the notions of Archimedean buoyancy to a fluid, we introduce the conceptual construct of a *test fluid element* originally defined in Section 13.2.5. A test fluid element does not alter the environment into which it is placed, so that it is an imaginary probe that allows us to map the Archimedean buoyancy of the fluid environment without modifying environment.³

To help build toward an understanding of the Archimedean buoyancy of a test fluid element, consider an arbitrary finite region, \mathcal{R} , contained within a static fluid such as shown in Figure 25.1. The region can be occupied by a rigid solid such as a submarine, a semi-rigid region such as a fish, or something else. For our studies, we are interested in an arbitrary region of the fluid itself, such as an infinitesimal fluid element or a finite fluid volume. The mass of the region is given by the integral of the fluid density over the region

$$M_{\text{fluid}} = \int_{\mathcal{R}} \rho dV. \quad (25.1)$$

We seek an understanding of the forces acting on this region from the surrounding fluid.

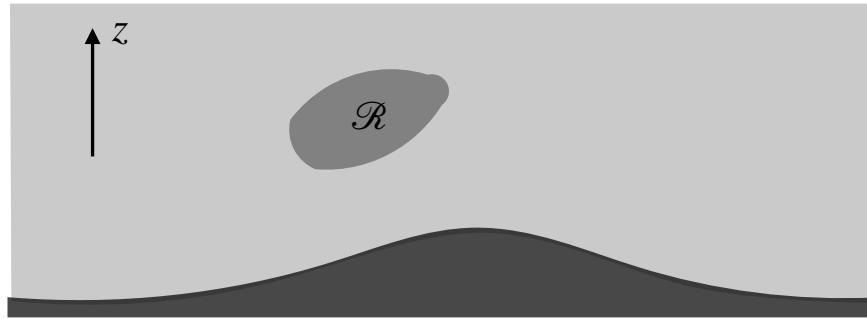


FIGURE 25.1: An arbitrary region, \mathcal{R} , contained within a fluid experiences a gravitational force acting down and a buoyant force acting up. The region considered here is arbitrary, such as a rigid solid, a swimming creature, an infinitesimal fluid element, or a finite region of the fluid. When the fluid is static, hydrostatics says that the fluid displaced by the body imparts a force equal to the weight of the displaced fluid. This result is *Archimedes' Principle*. The buoyant force acting on the body is determined by the difference in weight between the body and the displaced fluid. A positive buoyancy acts upward on a body that is less dense than the displaced fluid, and a negative buoyancy force acts downward on a body that is more dense than the displaced fluid.

25.2.1 Hydrostatic pressure force

As discussed in Chapters 19 and 20, any surface, even an imaginary surface, within a fluid experiences a contact stress due to interactions between the fluid and the surface. For a fluid at rest in a gravitational field, the only contact stress arises from pressure. Pressure is a compressive stress, acting in the direction determined by minus the outward normal along the surface. Integrating the pressure over the closed boundary, $\partial\mathcal{R}$, leads to the pressure force acting on the region

$$\mathbf{F}_{\text{pressure}} = - \oint_{\partial\mathcal{R}} p \hat{\mathbf{n}} dS, \quad (25.2)$$

³We dispense with the “does not alter” assumption when considering the effective buoyancy acting on a fluid element in Section 25.7.

where p is the pressure, $\hat{\mathbf{n}}$ is the outward normal on the boundary, and $d\mathcal{S}$ is the area element.

If the region is part of the fluid itself, then we can use Gauss's divergence theorem in the form of equation (2.67) to render the equivalent expression in terms of the volume integral of the pressure gradient

$$\mathbf{F}_{\text{pressure}} = - \int_{\mathcal{R}} \nabla p \, dV. \quad (25.3)$$

Furthermore, since the fluid is at rest, it is in an exact hydrostatic balance and thus only has a dependence on the vertical position within the fluid, $p = p(z)$. The hydrostatic pressure equals to the weight per horizontal area of fluid sitting above any point in the fluid so that its vertical derivative given by

$$\frac{dp}{dz} = -\rho g. \quad (25.4)$$

Hence, the force acting on the region, regardless the composition of the region, is

$$\mathbf{F}_{\text{pressure}} = - \int_{\mathcal{R}} \nabla p \, dV = - \int_{\mathcal{R}} \hat{\mathbf{z}} (dp/dz) \, dV = \hat{\mathbf{z}} g \int_{\mathcal{R}} \rho \, dV = \hat{\mathbf{z}} g M_{\text{fluid}}, \quad (25.5)$$

where we assumed a constant gravitational acceleration over \mathcal{R} . The hydrostatic pressure thus imparts a vertical upward force equal to the weight of the fluid displaced by the region. This result is a mathematical expression of *Archimedes' Principle*.

Note that there are no net horizontal forces acting over the region. The reason is that there are no horizontal pressure gradients within the fluid since, by assumption, the fluid is at rest and thus experiences no horizontal acceleration. So although the immersed body experiences horizontal compressive pressure forces, these forces balance when integrated over the body (assuming the body can withstand the forces!), thus leaving a zero net horizontal acceleration.

25.2.2 Archimedean buoyancy force

The vertical pressure force (25.6) acts to push the body vertically upward against gravity. Therefore, it reduces the vertical gravitational force when the body is within the fluid. We say that the massive body has a *buoyancy* imparted by the displaced fluid. If the body has a density less than the displaced fluid, then the body experiences a positive buoyancy force relative to the displaced fluid, in which case the body experiences an upward acceleration. The converse holds for a body more dense than the displaced fluid, in which case the body sinks downward. These buoyancy forces arise from pressure acting on the body. If the body is the fluid itself, then the forces arise from non-hydrostatic pressure forces since the vertical acceleration breaks the hydrostatic balance. We further examine these non-hydrostatic pressure forces in Section 25.7. However, for the present section we pursue a study of Archimedean buoyancy, which is computed as a relative force. Namely, we define a *buoyancy force* acting on a massive body relative to the displaced fluid according to

$$\mathbf{F}_{\text{buoyancy}} = \hat{\mathbf{z}} g \int_{\mathcal{R}} (\rho_{\text{fluid}} - \rho_{\text{body}}) \, dV = \hat{\mathbf{z}} g (M_{\text{fluid}} - M_{\text{body}}). \quad (25.6)$$

As so defined, the vertical buoyancy force is negative if the mass of the body is larger than that of the fluid it displaces, and conversely if the body is less massive. If the densities are equal, then the body is *neutrally buoyant* and thus experiences no net vertical force; i.e., it floats.

25.3 Mass density and its flavors

The density of a fluid element is central to determining its buoyancy. The *thermal equation of state*, or *equation of state* to be more brief (Section 21.9.1), provides an expression for the mass density as a function of pressure, temperature, and material tracer concentration (salinity in the ocean and humidity in the atmosphere). In this section we discuss the equation of state as well as the related flavors of mass density used to study stratified fluids.

25.3.1 Equation of state for the atmosphere and ocean

The atmosphere and ocean are commonly approximated as two-component fluids (air and water vapor for the atmosphere; freshwater and salt for ocean). We thus write the *in situ* density as a function

$$\rho = \rho(S, T, p). \quad (25.7)$$

This *equation of state* is a function of the *in situ* temperature, T , the *in situ* pressure, p , and the *in situ* salinity, S (ocean) or humidity (atmosphere).⁴ The term *in situ* refers to a property measured locally at a point in the fluid. Such *in situ* properties contrast to *potential* properties, which are based on referencing to a chosen pressure (e.g., potential temperature described in Section 21.8).

Liquids such as seawater have rather complex equations of states obtained from empirical fits to measurements. Part of the complexity arises from the multi-component nature of seawater (salt plus freshwater) as well as the nontrivial inter-molecular forces commonly found in liquids. In contrast, the dry atmosphere can, for many purposes, be well approximated as an ideal gas, which has a rather simple equation of state (see Section 21.9.1). Furthermore, even a moist atmosphere has an equation of state that can be massaged to look like that of an ideal gas (e.g., see Section 18.1 of [Vallis \(2017\)](#)). Hence, much of our discussion in this section is biased toward the ocean, where niceties of the equation of state are most important.

25.3.2 Modified temperature variables

As discussed in Section 21.8, the *in situ* temperature of a fluid element changes even if there is no heating applied to the element nor any changes to its material composition. Pressure changes provide the mechanical means for temperature to change even with processes that are adiabatic and constant composition. Is it possible to remove such pressure effects and still have a field that describes the “temperature” of a fluid element? That is, can we define a temperature field that is only modified by irreversible processes such as heating and mixing? This question is answered by defining *potential temperature*, θ as well as potential enthalpy or *Conservative Temperature*, Θ , with full discussions given for potential temperature in Section 21.8 and Conservative Temperature in Section 22.8. For now it is sufficient to note that the mass density can be written as a function salinity, potential temperature, and pressure

$$\rho = \rho(S, \theta, p), \quad (25.8)$$

or as a function of salinity, Conservative Temperature, and pressure

$$\rho = \rho(S, \Theta, p). \quad (25.9)$$

⁴For our purposes when discussing the ocean, we set $S = 1000 C$ with C the salt concentration, with this specification referred to as the *absolute salinity* by [IOC et al. \(2010\)](#). C generally has values around 0.035 so that S has values around 35.

For our purposes throughout this book, it is not important to distinguish between potential temperature and Conservative Temperature. Hence, we use θ for brevity when referring to either of the two temperature fields.

One comment on notation is key here. Namely, the functions $\rho(S, T, p)$, $\rho(S, \theta, p)$, and $\rho(S, \Theta, p)$ have distinct coefficients and as such they are distinct functions. Hence, one may wish to introduce notation to distinguish the functions by writing, say,

$$\rho = \mathcal{F}(S, T, p) = \mathcal{G}(S, \theta, p) = \mathcal{H}(S, \Theta, p). \quad (25.10)$$

However, we choose brevity in notation by allowing the functional dependence to signal the distinction. This overloaded notation is standard in the oceanography literature and yet care is needed to ensure clarity in understanding.

25.3.3 Infinitesimal density increments and material time changes

When computing the buoyancy of a fluid element, we compare the density of that element with the density of its local surrounding fluid environment. To support that comparison, we here consider how density changes under infinitesimal changes in spatial position. Given the functional dependence for the equation of state written in terms of S, θ, p (equation (25.8)), an infinitesimal density increment is given by

$$d\rho = \left[\frac{\partial \rho}{\partial S} \right] dS + \left[\frac{\partial \rho}{\partial \theta} \right] d\theta + \left[\frac{\partial \rho}{\partial p} \right] dp \equiv \rho \beta dS - \rho \alpha d\theta + c_s^{-2} dp. \quad (25.11)$$

The second equality introduced the following thermodynamic properties of the fluid

$$\beta = \frac{1}{\rho} \left[\frac{\partial \rho}{\partial S} \right]_{\theta, p} \quad \text{haline contraction coefficient} \quad (25.12)$$

$$\alpha = -\frac{1}{\rho} \left[\frac{\partial \rho}{\partial \theta} \right]_{S, p} \quad \text{thermal expansion coefficient} \quad (25.13)$$

$$c_s^2 = \left[\frac{\partial p}{\partial \rho} \right]_{S, \theta} \quad \text{squared sound speed.} \quad (25.14)$$

The haline contraction coefficient, β , is considered for the ocean, where *haline* refers to salinity.⁵

The infinitesimal density increment (25.11) leads to the expression for the material change in the *in situ* density moving along a trajectory

$$\frac{1}{\rho} \frac{D\rho}{Dt} = \beta \frac{DS}{Dt} - \alpha \frac{D\theta}{Dt} + \frac{1}{\rho c_s^2} \frac{Dp}{Dt}. \quad (25.15)$$

In the absence of mixing, the potential temperature and salinity are materially constant. In this case, the *in situ* density changes only through adiabatic processes that lead to pressure changes

$$\frac{D\rho}{Dt} = \frac{1}{c_s^2} \frac{Dp}{Dt} \quad \text{adiabatic and isohaline changes.} \quad (25.16)$$

⁵Note that in many chapters of this book, $\beta = \partial f / \partial y$ is the meridional derivative of the Coriolis parameter. We keep the two usages for β distinct so to avoid confusion.

25.3.4 Potential density

As discussed in Section 25.3.2, the reversible motion of a perfect fluid element generally occurs with materially constant potential temperature and materially constant tracer concentration. We thus find it convenient to combine the evolution of salinity and potential temperature into the evolution of a single variable. *Potential density* is one such combination, with it defined as the density a fluid element would have if reversibly moved to a reference pressure

$$\rho_{\text{pot}} = \rho(S, \theta, p = p_{\text{R}}). \quad (25.17)$$

Hence, potential density is found by evaluating the equation of state with the local value for S and θ , but with the pressure set to the reference pressure. As for the potential temperature, the reference pressure is often taken as that at sea level, though this is not necessary.

Material evolution of potential density

With the definition (25.17), the material evolution of potential density is given by

$$\frac{1}{\rho_{\text{R}}} \frac{D\rho_{\text{pot}}}{Dt} = \beta_{\text{R}} \frac{DS}{Dt} - \alpha_{\text{R}} \frac{D\theta}{Dt}, \quad (25.18)$$

where

$$\beta_{\text{R}} = \frac{1}{\rho_{\text{R}}} \left[\frac{\partial \rho(S, \theta, p_{\text{R}})}{\partial S} \right]_{\theta} \quad \text{haline contraction at } p = p_{\text{R}} \quad (25.19)$$

$$\alpha_{\text{R}} = -\frac{1}{\rho_{\text{R}}} \left[\frac{\partial \rho(S, \theta, p_{\text{R}})}{\partial \theta} \right]_S \quad \text{thermal expansion at } p = p_{\text{R}} \quad (25.20)$$

are the haline contraction and thermal expansion coefficients evaluated at the reference pressure $p = p_{\text{R}}$. Since pressure is fixed at the reference value, there is no pressure derivative on the right hand side of equation (25.18). Potential temperature and salinity are materially constant for reversible processes; i.e., adiabatic motion that also maintains constant matter content (e.g., isohaline) for fluid elements. By construction, potential density is also materially constant for reversible processes since both terms on the right hand side of equation (25.18) vanish. This behavior is in contrast to *in situ* density, whose evolution is affected by pressure changes as seen by equations (25.15) and (25.16).

Reference pressures for ρ_{pot} and θ

The reference pressure for the potential density is commonly assumed to be the same as for the potential temperature. This assumption is particularly the norm for the atmosphere, where the reference pressure is generally taken at the sea level. Likewise for the ocean, the potential temperature is generally computed using a sea level reference pressure. However, there are many occasions in the ocean to consider potential density with larger reference pressures, such as when considering processes within the ocean interior. Doing so is motivated by the rather strong nonlinear effects associated with the seawater equation of state. In this case, pressure effects prompt one to choose a reference pressure closer to the *in situ* pressure near to the region of analysis. Even though it is common to choose a potential density reference pressure distinct from the surface pressure, the potential temperature reference pressure generally remains at the surface. There is no fundamental problem with the use of distinct reference pressures for ρ_{pot} and θ . In particular, all of the above properties of potential density remain unchanged.

25.3.5 Linear equation of state for the ocean

For certain purposes, it is useful to approximate the equation of state used to study ocean fluid mechanics. One common idealization is to compute density as a linear function of potential temperature and salinity

$$\rho_{\text{pot}} = \rho_0 [1 - \alpha(\theta - \theta_0) + \beta(S - S_0)], \quad (25.21)$$

where α , β , θ_0 , and S_0 are positive constants. An even further simplification is to set salinity to a constant, so that density is just a linear function of potential temperature.

25.3.6 Further study

Chapter 1 of [Vallis \(2017\)](#) provides a pedagogical discussion of the equation of state for an ideal gas atmosphere and for seawater, as well as a discussion of the various flavors of density. See also Section 18.1 of [Vallis \(2017\)](#) for the equation of state for an ideal gas with water vapor. The seawater equation of state is detailed by [IOC et al. \(2010\)](#), with a pedagogical overview provided by [McDougall et al. \(2013\)](#).

25.4 Archimedean buoyancy of a test fluid element

We now return to the notions of Archimedean buoyancy (in brief, the fluid “buoyancy”) of a test fluid element. Again, buoyancy is the gravitational acceleration that acts on a massive body due to the difference between the density of the body and the density of the local fluid environment. For this section, we consider the massive body to be a test fluid element whose presence does not alter the flow field.

25.4.1 Locally referenced Archimedean buoyancy

Consider a local definition of fluid buoyancy according to

$$b_{\text{local}} = -g(\rho_{\text{test}} - \rho_{\text{environ}})/\rho_{\text{environ}} = g(1 - \rho_{\text{test}}/\rho_{\text{environ}}), \quad (25.22)$$

where ρ_{environ} is the local density of the fluid environment, and ρ_{test} is the density of the test fluid element within that environment. If the fluid element has a density greater than the environment, then it has a negative locally referenced buoyancy, and vice versa.

We determine ρ_{test} by specifying its point of origin and specifying how it is moved (e.g., with or without mixing?) to the environment point. Conventional approaches are specified later in this section. The key point is that buoyancy as defined by equation (25.22) is a function of the path that the test fluid element takes to reach the environment point. This subjectivity lends ambiguity in the definition of buoyancy. We remove this ambiguity by asking specific questions about local buoyancy. For example, if the fluid element moves an infinitesimal distance while mixing its temperature and salinity with the environment, what direction maintains a neutrally buoyant state for the fluid element? This question forms the basis for defining *neutral directions*.

Working with locally referenced Archimedean buoyancy requires a redefinition of a reference state when moving from point to point within a fluid. The continuum of reference states allows for an accounting of the gravitational stability and neutral directions. However, the re-referencing is not a process that can be seamlessly incorporated to the equations of motion since at each point one needs to redefine the reference state. As an alternative means to garner local information about forces associated with density gradients, we explicitly focus on the pressure force in our discussion of *effective buoyancy* in Section 25.7.

25.4.2 Globally referenced Archimedean buoyancy

The definition (25.22) accepts that buoyancy is a relative field. Hence, at each fluid point we redefine the environment to which to compare the density of the test fluid element. However, there are cases in which it is more convenient, and sufficient, to define a globally constant environment with a constant reference density, ρ_{ref} . In this case we consider the global buoyancy as

$$b_{\text{global}} = g(1 - \rho/\rho_{\text{ref}}), \quad (25.23)$$

where we compute ρ according to the local environmental density. This definition is particularly useful for idealized cases where the *in situ* density is not a function of pressure. In this case buoyancy is a function only of potential temperature and salinity so that we can make use of potential density to measure buoyancy (as explained below).

Although the numerical value of b_{global} is a function of the reference density, what is more relevant is the buoyancy of one fluid element relative to another

$$\Delta b_{\text{global}} = -\Delta\rho/\rho_{\text{ref}}. \quad (25.24)$$

The sign of this relative buoyancy does not depend on the reference density, with the sign all that we need to conclude whether one fluid element is more buoyant than another. Furthermore, with a globally constant environmental density, the buoyancy becomes a local function of space. That is, we no longer compare the fluid element density to a changing local density. Instead, we compute the local density and compare it to the reference density. We can thus determine b_{global} at a point through information available just at that point. Correspondingly, we can map b_{global} and determine the relative buoyancy of fluid elements anywhere in the fluid.

25.5 Buoyancy stratification and neutral directions

What are the Archimedean buoyant forces that act on a fluid element? Is the fluid gravitationally stable to small perturbations? We answer these questions by using a test fluid element to probe the local fluid environment without disturbing the environment. We do not consider time changes in this section, instead considering a snapshot of the fluid environment and using test fluid elements to probe the buoyant properties of the environment.

25.5.1 General considerations

As a fluid element moves through the fluid environment it is exposed to a suite of physical processes that can modify its thermal, material, and mechanical properties; i.e., its θ , S , and p . Modification of its pressure occurs through contact stresses with other fluid elements (Chapters 19 and 20). Modification of its thermal and material properties occurs through the mixing-induced exchange of heat and matter. The exchange of thermal and material properties occurs in the presence of irreversible processes whereas mechanical exchanges occur either reversibly (pressure exchange) or irreversibly (viscous exchange; Section 19.2).

The *in situ* density of a fluid element generally changes when it moves through the fluid, with the density change determined by how the element interacts with the surrounding environment. We conceive of two complementary interactions for the purpose of examining the local Archimedean buoyancy of the test fluid element.

- UNMIXED TEST FLUID ELEMENT: Displace the test fluid element without changing θ and S yet allowing p to equilibrate with the local environment. We imagine this adiabatic and iso-haline displacement to occur by surrounding the test element with a thermally and materially

impermeable sack that is elastic. In this case there is no mixing of the test element with the environment, yet it mechanically equilibrates its pressure with the surrounding fluid. Hence, the test fluid element's *in situ* density changes through the changing pressure of the local environment.

- **MIXED TEST FLUID ELEMENT:** Displace the test fluid element with θ , S , and p equilibrating with the local environment. Such equilibration requires a complete mixing of the test fluid element's θ and S with the local environment, as well as the mechanical equilibration of its pressure.

If we displace the unmixed test fluid element to a region where its *in situ* density differs from the local environment, then the fluid element feels a local Archimedean buoyant force. However, there are directions that the test fluid element can move that leave its local Archimedean buoyancy zero; i.e., where the test fluid element retains the same *in situ* density as the local environment and so remains neutrally buoyant. These directions are referred to as *neutral directions*, with a test fluid element displaced along a neutral direction maintaining zero local Archimedean buoyancy. It is as if the test fluid element floats along a neutral direction.

A complementary means to conceive of neutral directions is found by considering the mixed test fluid element that is displaced and equilibrates its θ , S , and p with the local environment. An arbitrary displacement leads to a change in the fluid element's *in situ* density due to changes in θ , S , and p . However, if the test element is displaced along the neutral direction, the mixing-induced changes from θ and S exactly balance (details given below). In this manner, the only change to the *in situ* density arises from changes to the pressure felt by the test fluid element, with these changes identical to those felt by the unmixed test fluid element.

25.5.2 Comparing *in situ* densities

To make the above ideas mathematically precise, consider an infinitesimal displacement, $d\mathbf{x}$, of a test fluid element and examine how its *in situ* density, $\rho = \rho(S, \theta, p)$, changes under the above two displacements. First consider the mixed fluid element so that it equilibrates θ , S , and p with the local environment. The *in situ* density of the test element at the new location equals to that of the local environment, $\rho(\mathbf{x} + d\mathbf{x})$. To leading order, the difference in density between the origin and the displaced position is computed according to the exact differential

$$d\rho = \rho(\mathbf{x} + d\mathbf{x}) - \rho(\mathbf{x}) \quad (25.25a)$$

$$= \rho[S(\mathbf{x} + d\mathbf{x}), \theta(\mathbf{x} + d\mathbf{x}), p(\mathbf{x} + d\mathbf{x})] - \rho[S(\mathbf{x}), \theta(\mathbf{x}), p(\mathbf{x})] \quad (25.25b)$$

$$= d\mathbf{x} \cdot \left[\frac{\partial \rho}{\partial \theta} \nabla \theta + \frac{\partial \rho}{\partial S} \nabla S + \frac{\partial \rho}{\partial p} \nabla p \right] \quad (25.25c)$$

$$= \rho d\mathbf{x} \cdot \left[-\alpha \nabla \theta + \beta \nabla S + \frac{1}{\rho c_s^2} \nabla p \right]. \quad (25.25d)$$

For the unmixed test fluid element there is no exchange (mixing) of θ and S with the surrounding environment. Hence, the test element's density changes just from pressure changes

$$(d\rho)_{(\text{no mix})} = \rho(\mathbf{x} + d\mathbf{x})_{(\text{no mix})} - \rho(\mathbf{x}) \quad (25.26a)$$

$$= \rho[S(\mathbf{x}), \theta(\mathbf{x}), p(\mathbf{x} + d\mathbf{x})] - \rho[S(\mathbf{x}), \theta(\mathbf{x}), p(\mathbf{x})] \quad (25.26b)$$

$$= \rho d\mathbf{x} \cdot \left[\frac{1}{\rho c_s^2} \nabla p \right]. \quad (25.26c)$$

Comparing the densities of the two displaced test fluid elements renders

$$\rho(\mathbf{x} + d\mathbf{x}) - \rho(\mathbf{x} + d\mathbf{x})_{(\text{no mix})} = [\rho(\mathbf{x} + d\mathbf{x}) - \rho(\mathbf{x})] - [\rho(\mathbf{x} + d\mathbf{x})_{(\text{no mix})} - \rho(\mathbf{x})] \quad (25.27\text{a})$$

$$= \rho[S(\mathbf{x} + d\mathbf{x}), \theta(\mathbf{x} + d\mathbf{x}), p(\mathbf{x} + d\mathbf{x})] - \rho[S(\mathbf{x}), \theta(\mathbf{x}), p(\mathbf{x} + d\mathbf{x})] \quad (25.27\text{b})$$

$$= \rho d\mathbf{x} \cdot [-\alpha \nabla \theta + \beta \nabla S]. \quad (25.27\text{c})$$

We discuss this result in the following.

25.5.3 Neutral directions and the neutrality condition

As noted in Figure 25.2, if the density of the displaced unmixed test fluid element is the same as the local environment, then the particular displacement defines a neutral direction. From equation (25.27c) we see a displacement along a neutral displacement satisfies

$$d\mathbf{x}_{\text{neutral}} \cdot [-\alpha \nabla \theta + \beta \nabla S] = d\mathbf{x}_{\text{neutral}} \cdot \hat{\gamma} | -\alpha \nabla \theta + \beta \nabla S | = 0, \quad (25.28)$$

where we introduced the *dianeutral direction*

$$\hat{\gamma} = \frac{-\alpha \nabla \theta + \beta \nabla S}{| -\alpha \nabla \theta + \beta \nabla S |}. \quad (25.29)$$

Infinitesimal displacements, $d\mathbf{x}_{\text{neutral}}$, that are orthogonal to the dianeutral direction, $\hat{\gamma}$, occur along a neutral direction

$$d\mathbf{x}_{\text{neutral}} \cdot \hat{\gamma} = 0 \implies \text{displacement along a neutral direction.} \quad (25.30)$$

These displacements of the unmixed test fluid element lead to no difference in the *in situ* density of the local environment and the fluid element. Hence, neutral displacements retain a vanishing local Archimedean buoyancy for the unmixed test fluid element. In contrast, displacements in the dianeutral direction alter the local buoyancy.

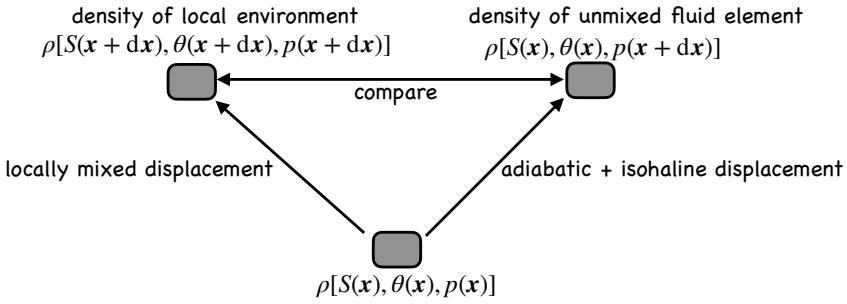
Let us write the neutral displacement in the form

$$d\mathbf{x}_{\text{neutral}} = \hat{\mathbf{t}}_{\text{neutral}} ds, \quad (25.31)$$

where ds is the arc-length along the displacement and $\hat{\mathbf{t}}_{\text{neutral}}$ is the unit vector pointing along the neutral displacement. We can thus write equation (25.28) as

$$\alpha \hat{\mathbf{t}}_{\text{neutral}} \cdot \nabla \theta = \beta \hat{\mathbf{t}}_{\text{neutral}} \cdot \nabla S, \quad (25.32)$$

which we refer to as the *neutrality condition*. The neutrality condition means that the derivatives of θ and S aligned along a neutral direction, and suitably weighted by α and β , are precisely balanced. So when considering neutral directions from the perspective of the mixed test fluid element, the mixing-induced changes in θ precisely compensate mixing-induced changes in S as per the neutrality condition (25.32). As a result, the *in situ* density of the mixed test fluid element changes only via changes to the pressure (since θ and S changes are compensated), which is precisely how the unmixed test fluid element changes its *in situ* density.



$$\rho[S(x+dx), \theta(x+dx), p(x+dx)] = \rho[S(x), \theta(x), p(x+dx)] \implies \text{neutral direction}$$

$$\alpha \hat{\mathbf{t}}_{\text{neutral}} \cdot \nabla \theta = \beta \hat{\mathbf{t}}_{\text{neutral}} \cdot \nabla S \implies \text{neutrality condition}$$

FIGURE 25.2: Schematic of the complementary thought experiments used to determine neutral directions. For the right pathway we imagine a test fluid element that is thermally and materially closed but is mechanically open. Hence, as the test fluid element is displaced from its original location, \mathbf{x} , to a new position, $\mathbf{x} + d\mathbf{x}$, it equilibrates to the local pressure but retains the θ and S of the origin. If the *in situ* density of the fluid element is displaced along a direction where it maintains the same density as the local environment, then the locally defined Archimedean buoyancy vanishes and this displacement is aligned along a neutral direction. The complement perspective is shown by the left pathway whereby we follow the mixed test fluid element. If the mixing-induced changes in θ precisely compensate mixing-induced changes in S , according to the neutrality condition (25.32), then the *in situ* density changes for this test fluid element arise only from pressure changes, just like for the unmixed test element on the right pathway.

25.5.4 Buoyancy frequency and gravitational stability

Thus far we have considered a general displacement, $d\mathbf{x}$, of two test fluid elements, and determined conditions for this displacement to keep the locally defined Archimedean buoyancy at zero. Here we specialize to the case of a vertical displacement (Figure 25.3) in which the difference between the environmental density and the unmixed test element's density is

$$\rho(z + dz) - \rho(z + dz)_{(\text{no mix})} = \rho dz \left[-\alpha \frac{\partial \theta}{\partial z} + \beta \frac{\partial S}{\partial z} \right]. \quad (25.33)$$

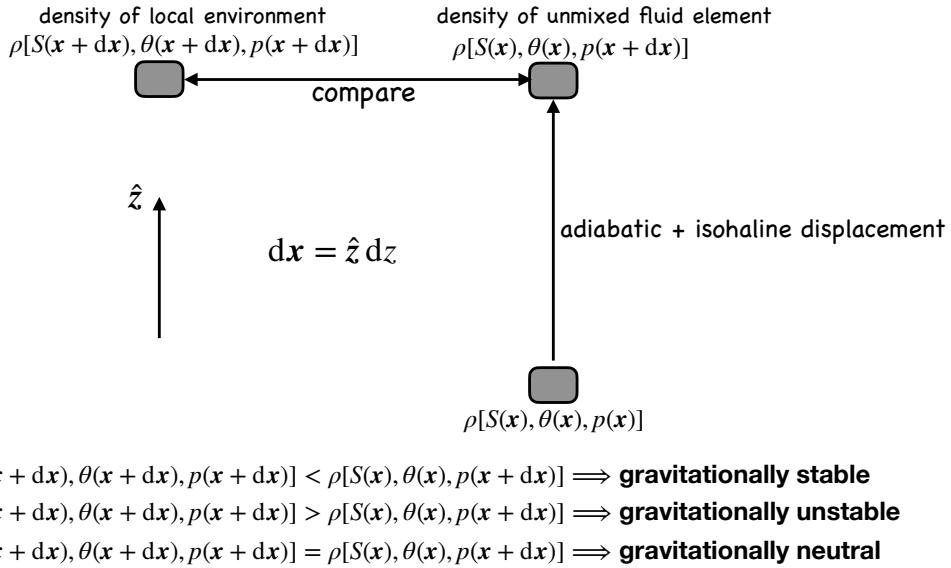
Now consider a vertically upward displacement so that $dz > 0$. If the surrounding environment has a lower density than the adiabatic and isohaline displaced test fluid element, $\rho(z + dz) < \rho(z + dz)_{(\text{no mix})}$, then the test element will feel a buoyancy force returning it to the original depth. The restorative buoyancy force per volume is written

$$g [\rho(z + dz) - \rho(z + dz)_{(\text{no mix})}] = g \rho dz \left[-\alpha \frac{\partial \theta}{\partial z} + \beta \frac{\partial S}{\partial z} \right] \equiv -N^2 \rho dz, \quad (25.34)$$

where we defined the squared buoyancy frequency

$$N^2 = g \left[\alpha \frac{\partial \theta}{\partial z} - \beta \frac{\partial S}{\partial z} \right]. \quad (25.35)$$

Stable vertical motion results from a background density profile with $N^2 > 0$. An unstable profile occurs when $N^2 < 0$, in which case motion of the test fluid element results in an exponential growth associated with a *gravitational instability*. That is, when the fluid column is unstably stratified in the vertical, an initially tiny vertical displacement of a test element will lead to an even larger displacement, thus causing the perturbation to grow unbounded, which is the sign of an instability. We illustrate these cases in Figure 25.4 for the case of linear density profiles.



$\rho[S(x + dx), \theta(x + dx), p(x + dx)] < \rho[S(x), \theta(x), p(x + dx)] \Rightarrow \text{gravitationally stable}$
 $\rho[S(x + dx), \theta(x + dx), p(x + dx)] > \rho[S(x), \theta(x), p(x + dx)] \Rightarrow \text{gravitationally unstable}$
 $\rho[S(x + dx), \theta(x + dx), p(x + dx)] = \rho[S(x), \theta(x), p(x + dx)] \Rightarrow \text{gravitationally neutral}$

FIGURE 25.3: Schematic of the calculation used to examine whether a fluid column is gravitationally stable under vertical displacement of a test fluid element, here realized by specializing the general displacement in Figure 25.2. A test fluid element is displaced from its original location at a position x to a position $x + dx$, with $dx = \hat{z} dz$ and $dz > 0$ in this figure. We compare the density of the displaced unmixed test fluid element with the local environment to determine whether the density stratification of the fluid environment is gravitationally stable, unstable, or neutral.

25.5.5 Locally referenced potential density

Equation (25.35) defines the squared buoyancy frequency in terms of the vertical temperature and salinity gradients. This expression is identical to the vertical gradient of the potential density (25.17), when the reference pressure for density is taken local to the point where the buoyancy frequency is computed. That is, the vertical gradient of the *locally referenced potential density* provides a measure of the vertical stratification

$$N^2 = -g \left[\frac{\partial \ln \rho_{\text{pot}}}{\partial z} \right]_{p_R=p} = g \left[\alpha \frac{\partial \theta}{\partial z} - \beta \frac{\partial S}{\partial z} \right]. \quad (25.36)$$

Note that at a point in the fluid, the locally referenced potential density equals to the *in situ* density. However, when probing nearby points by displacing test fluid elements, and thus to take spatial gradients into account, the two densities have distinct gradients. Namely, the *in situ* density is modified by pressure gradients, whereas spatial gradients of the locally referenced potential density remove pressure effects.

25.5.6 Gravitational stability of an ideal gas atmosphere

We introduced the adiabatic lapse rate in Section 21.7 as a measure of how temperature varies as a function of pressure or depth. For an ideal gas atmosphere (Section 21.9), the squared buoyancy frequency can be written (Exercise 25.2)

$$N^2 = \frac{g}{\theta} \frac{\partial \theta}{\partial z}. \quad (25.37)$$

The potential temperature for an ideal gas is given by equation (21.159)

$$\theta = T \left[\frac{p_R}{p} \right]^{\varphi} \quad (25.38)$$

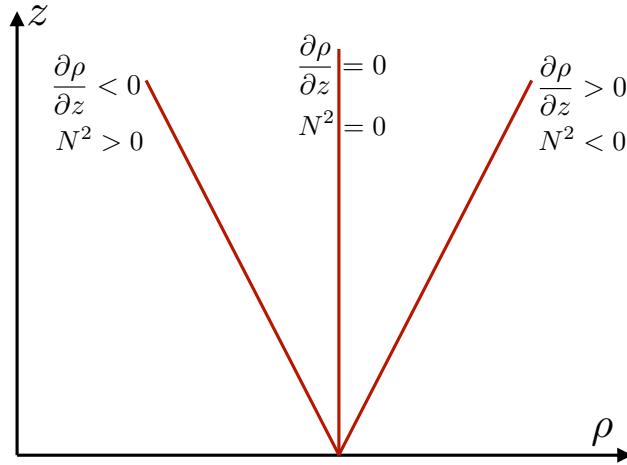


FIGURE 25.4: Sample vertical “soundings” of density in the atmosphere or density profiles in the ocean, with the density assumed to be a linear function of z . If the density decreases upward, the squared buoyancy frequency is positive, $N^2 > 0$, and the fluid is gravitationally stable. If the density does not change when moving vertically, then the fluid is said to be neutrally buoyant with $N^2 = 0$. If the density increases moving upward, then the fluid is gravitationally unstable, with this instability signaled by $N^2 < 0$.

where

$$\varphi = R^M/c_p \quad (25.39)$$

is a dimensionless constant for a simple ideal gas. Consequently, the squared buoyancy frequency takes the form

$$g^{-1} N^2 = \frac{\partial \ln \theta}{\partial z} = \frac{\partial \ln T}{\partial z} - \varphi \frac{\partial \ln p}{\partial z}. \quad (25.40)$$

For an exactly hydrostatic fluid with a constant gravitational acceleration, the vertical derivative of pressure is given by

$$\frac{dp}{dz} = -\rho g, \quad (25.41)$$

so that pressure at a point in the fluid equals to the weight per area above that point. Using this result leads to the squared buoyancy frequency

$$g^{-1} N^2 = \frac{d \ln T}{dz} + \frac{\varphi g \rho}{p} = \frac{1}{T} \frac{\partial T}{\partial z} + \frac{g}{c_p T}, \quad (25.42)$$

where we used the ideal gas relation $p = \rho T R^M$ for the final step.

A vanishing buoyancy frequency, or equivalently a vanishing vertical derivative of potential temperature, occurs when the vertical temperature gradient equals to the dry adiabatic lapse rate

$$N^2 = 0 \iff \frac{\partial T}{\partial z} = \Gamma_d, \quad (25.43)$$

where (see equation (21.141))

$$\Gamma_d = -\frac{g}{c_p} \approx -9.8 \text{ K/(1000 m)}. \quad (25.44)$$

That is, if the *in situ* temperature decreases upon ascent more quickly than the dry adiabatic lapse rate, then the vertical column is gravitationally unstable. In effect, the column becomes top heavy

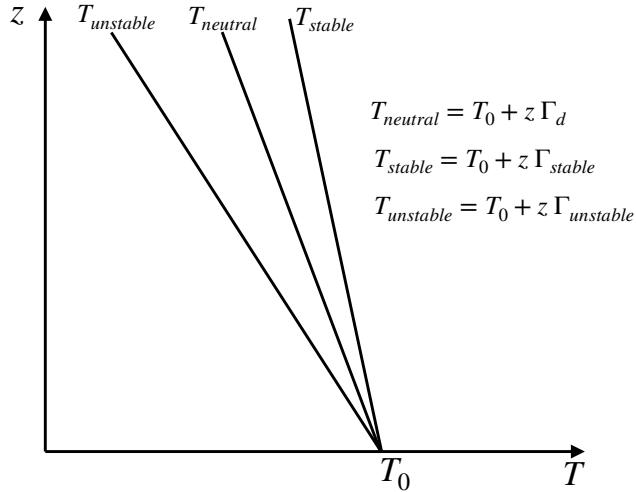


FIGURE 25.5: Three vertical profiles of *in situ* temperature in a dry ideal gas atmosphere. The neutrally stable profile has $T_{neutral} = T_0 + z \Gamma_d$, where $\Gamma_d = -g/c_p \approx -9.8 \text{ K}/(1000 \text{ m})$. A vertically unstable profile (heavy over light) has $T_{unstable} = T_0 + z \Gamma_{unstable}$ where $\Gamma_{unstable} < \Gamma_d = -g/c_p$. In contrast, the gravitationally stable atmosphere has $T_{stable} = T_0 + z \Gamma_{stable}$ where $\Gamma_{stable} > \Gamma_d = -g/c_p$.

and subject to overturning. We summarize this stability criteria as

$$\text{stable} \quad N^2 > 0 \iff -\frac{\partial T}{\partial z} < \frac{g}{c_p} \quad (25.45)$$

$$\text{unstable} \quad N^2 < 0 \iff -\frac{\partial T}{\partial z} > \frac{g}{c_p}, \quad (25.46)$$

with Figure 25.5 providing an illustration for three linear temperature profiles.

25.5.7 Comments and further study

Neutral directions were introduced to oceanography by [McDougall \(1987a\)](#) and [McDougall \(1987b\)](#), and they are the basis for how oceanographers think about buoyancy stratification. [McDougall et al. \(2014\)](#) offer another presentation of why neutral directions are relevant for the ocean, taking into account measurements of dianeutral mixing. Our discussion of neutral directions was inspired by the concise presentation in Section 2.7.2 of [Olbers et al. \(2012\)](#).

In deriving an algorithm to compute neutral directions, we have not presented an argument for dynamical processes favoring neutral displacements. That argument requires dynamical principles, such as the equations of motion and associated energetic constraints. Yet there is presently no dynamical argument for why the ocean prefers neutral displacements over arbitrary non-neutral displacements. However, robust empirical measures support the relevance of neutral directions for orienting tracer mixing within the ocean interior. More discussion of these points is given by [McDougall et al. \(2014\)](#).

Our study of vertical gravitational stability follows Section 3.6 of [Gill \(1982\)](#) as well as Section 2.10 of [Vallis \(2017\)](#). In an actual fluid, the movement of any fluid, even a tiny fluid element, modifies the surrounding fluid so that a perfect test fluid element is a fiction. We return to this point in Section 25.7 when studying effective buoyancy.

25.6 Neutral helicity

In our study of neutral directions in Section 25.5 we introduced the notion of a neutral displacement. For the test fluid element that mixes with its environment, a neutral displacement results when the mixing of θ and S are balanced according to the neutrality condition (25.32). We here expose a rather surprising result concerning the topological properties of the surface defined by a series of such neutral displacements.

25.6.1 Mathematical preliminaries

Consider a simply connected smooth surface with outward normal written in the form

$$\hat{\mathbf{n}} = |\mathbf{N}|^{-1} \mathbf{N}. \quad (25.47)$$

Likewise, consider a unit vector $\hat{\mathbf{t}}$ that lives within the surface and is directed tangent to an arbitrary closed loop. Since $\mathbf{N} \cdot \hat{\mathbf{t}} = 0$ by construction, we can integrate around an arbitrary closed loop within the surface and still maintain the trivial result

$$\oint_{\partial\mathcal{S}} \mathbf{N} \cdot \hat{\mathbf{t}} \, dl = 0. \quad (25.48)$$

Note that we chose a counterclockwise orientation of the loop around the boundary, with $\partial\mathcal{S}$ denoting the boundary of the area \mathcal{S} within the surface.

Now apply Stokes' Theorem to the loop integral (25.48) to yield

$$\int_{\mathcal{S}} (\nabla \wedge \mathbf{N}) \cdot \hat{\mathbf{n}} \, d\mathcal{S} = 0, \quad (25.49)$$

where $d\mathcal{S}$ is the area element in the surface \mathcal{S} with outward normal $\hat{\mathbf{n}}$. Since the closed path is arbitrary, the area integral (25.49) vanishes only if the integrand is identically zero. We conclude that for the surface to be simply connected requires that the helicity must vanish

$$\mathcal{H} = \mathbf{N} \cdot (\nabla \wedge \mathbf{N}) = 0 \implies \text{simply connected surface.} \quad (25.50)$$

The converse also holds so that

$$\mathcal{H} \neq 0 \implies \text{non-simply connected surface.} \quad (25.51)$$

25.6.2 Helical nature of neutral displacements

Now apply the above mathematical results towards the question of whether a surface with its outward normal parallel to the dianeutral direction,

$$\mathbf{N} = -\alpha \nabla \theta + \beta \nabla S \quad (25.52)$$

is simply connected. We thus consider the *neutral helicity*

$$\mathcal{H}_{\text{neutral}} = (-\alpha \nabla \theta + \beta \nabla S) \cdot \nabla \wedge (-\alpha \nabla \theta + \beta \nabla S), \quad (25.53)$$

which can be written

$$\mathcal{H}_{\text{neutral}} = (-\alpha \nabla \theta + \beta \nabla S) \cdot \nabla \wedge (-\alpha \nabla \theta + \beta \nabla S) \quad (25.54a)$$

$$= -\alpha \nabla \theta \cdot (\nabla \wedge \beta \nabla S) - \beta \nabla S \cdot (\nabla \wedge \alpha \nabla \theta) \quad (25.54b)$$

$$= -\alpha \nabla \theta \cdot (\nabla \beta \wedge \nabla S) - \beta \nabla S \cdot (\nabla \alpha \wedge \nabla \theta). \quad (25.54c)$$

Now expand the gradients of α and β according to

$$\nabla\alpha = \alpha_\theta \nabla\theta + \alpha_S \nabla S + \alpha_p \nabla p \quad (25.55a)$$

$$\nabla\beta = \beta_\theta \nabla\theta + \beta_S \nabla S + \beta_p \nabla p, \quad (25.55b)$$

where the subscripts are shorthands for the partial derivatives. In this case

$$-\mathcal{H}_{\text{neutral}} = \alpha \nabla\theta \cdot (\beta_\theta \nabla\theta + \beta_p \nabla p) + \beta \nabla S \cdot (\alpha_S \nabla S + \alpha_p \nabla p) \quad (25.56a)$$

$$= \alpha \nabla\theta \cdot (\nabla p \wedge \nabla S) \beta_p + \beta \nabla S \cdot (\nabla p \wedge \nabla\theta) \alpha_p \quad (25.56b)$$

$$= \nabla p \cdot (\nabla S \wedge \nabla\theta) (\alpha \beta_p - \beta \alpha_p). \quad (25.56c)$$

Introducing the thermobaricity parameter from Section 50.4

$$\mathcal{T} = \beta \partial_p(\alpha/\beta) \quad (25.57)$$

renders the tidy result

$$\mathcal{H}_{\text{neutral}} = \beta \mathcal{T} \nabla p \cdot (\nabla S \wedge \nabla\theta). \quad (25.58)$$

Consequently, a nonzero neutral helicity is fundamentally related to a nonzero thermobaricity parameter \mathcal{T} . It is also associated with the non-zero volume for a parallelopiped in (θ, S, p) space (see Section 1.5)

$$\nabla p \cdot (\nabla S \wedge \nabla\theta) = \nabla\theta \cdot (\nabla p \wedge \nabla S) = \nabla S \cdot (\nabla\theta \wedge \nabla p). \quad (25.59)$$

A nonzero $\mathcal{H}_{\text{neutral}}$ means that a path defined by the accumulation of neutral directions does not close. Rather, they possess a helical structure.

25.6.3 Comments and further study

Neutral helicity is a property of seawater that is revealed through the neutrally compensated mixing of θ and S . That is, neutral helicity is not a direct property of mixing though mixing is required to determine neutral directions. [McDougall and Jackett \(1988\)](#) first noted the nontrivial topology defined by neutral displacements, and [McDougall and Jackett \(2007\)](#) provide more elaboration and analysis from observational based measurements. [Bennett \(2019\)](#) connects the geometry defined by neutral paths with mathematical properties encountered in a variety of thermodynamic systems. [Stanley \(2019\)](#) offers insights into the topology of neutral directions.

25.7 Effective buoyancy and vertical accelerations

We extend our discussion of Archimedean buoyancy in this section by focusing on the vertical forces acting on a fluid element rather than a test fluid element. For this purpose, we find it useful to decompose these forces into *static forces* that remain even when imagining all motion to vanish, plus the *motional forces* that remain when imagining all density gradients to vanish. Motional forces are specific to details of the motion, whereas static forces can generally be deduced just from knowledge of the density field. In analogy with Archimedean buoyancy, we use the term *effective buoyancy* for the static forces acting on a fluid element. Although the effective buoyancy is in part comprised of Archimedean buoyancy, there are distinctions that arise through interactions between the fluid element and its surrounding environment.

25.7.1 Vertical motion in approximately hydrostatic fluids

As introduced in Section 23.3 and further detailed in Section 24.2, an approximate hydrostatic fluid is one in which the vertical pressure gradient locally balances the density, with the horizontal gradient of hydrostatic pressure contributing to horizontal accelerations. Although vertical motion can occur in the approximately hydrostatic fluid, that motion is diagnosed rather than prognosed since the vertical momentum equation is reduced to local hydrostatic balance. For example, a diagnostic evaluation of the vertical velocity in a Boussinesq fluid is performed through vertically integrating the continuity equation.

As we see in equation (25.67c) below, the vertical gradient of the non-hydrostatic pressure is the only inviscid force contributing to a vertical acceleration. So even if the fluid is approximately hydrostatic, it is the non-hydrostatic pressure force that enables vertical accelerations. This situation is directly analogous to the oceanic Boussinesq approximation as derived in Section 24.1, whereby the fluid flow is incompressible so that $\nabla \cdot \mathbf{v} = 0$, yet the fluid itself is compressible so that $D\rho/Dt \neq 0$. For the approximately hydrostatic fluid we have the local hydrostatic balance, $\partial p/\partial z = -\rho g$, yet nonzero vertical acceleration, $Dw/Dt \neq 0$.

25.7.2 Two decompositions of pressure and gravity forces

Consider the momentum equation with the geopotential, $\Phi = g z$,

$$\rho (\partial_t + \mathbf{v} \cdot \nabla) \mathbf{v} + f \mathbf{z} \wedge \rho \mathbf{v} = -\nabla p - \rho g \hat{\mathbf{z}} + \rho \mathbf{F}. \quad (25.60)$$

We are concerned with the pressure and gravity contributions on the right hand side, and we offer two means to organize these terms. The first follows that used to formulate the Boussinesq approximation in Section 24.1.1. Summarizing those steps for completeness, we introduce a constant reference density, ρ_0 , along with a corresponding hydrostatically balanced reference pressure

$$p = p' + p_0 \quad \text{with} \quad dp_0(z)/dz = -\rho_0 g. \quad (25.61)$$

This decomposition brings the pressure and gravity on the right hand side of the momentum equation (25.60) into the form

$$\nabla p + \rho g \hat{\mathbf{z}} = \nabla p' - \rho_0 b \hat{\mathbf{z}} \quad \text{with} \quad b = -g(\rho - \rho_0)/\rho_0, \quad (25.62)$$

where b is the globally referenced Archimedean buoyancy computed relative to the globally constant reference density, ρ_0 . The second decomposition splits pressure into a local hydrostatic pressure, p_h , and a non-hydrostatic pressure, p_{nh} ,

$$p = p_h + p_{nh} \quad \text{with} \quad \partial p_h / \partial z = -\rho g, \quad (25.63)$$

in which case

$$\nabla p + \rho g \hat{\mathbf{z}} = \nabla p_{nh} + \nabla_z p_h = \nabla_z p + \hat{\mathbf{z}} \partial_z p_{nh}. \quad (25.64)$$

The following comments compare and contrast the two decompositions.

- The static pressure $p_0(z)$ has no effect on the flow and is thus dynamically irrelevant. Hence, by decomposing pressure as $p = p_0 + p'$, we expose the dynamically active portion of the pressure field, p' , by removing the dynamically irrelevant pressure, p_0 , from the momentum equation. Notably, this decomposition results in a dynamical pressure, p' , that has both hydrostatic and non-hydrostatic contributions.

- The local hydrostatic plus non-hydrostatic decomposition serves to split the pressure into its two dynamically distinct portions. However, it does not remove the global hydrostatic background pressure from the forcing, which can be seen by the identity

$$\partial p_h / \partial z = dp_0(z) / dz - \rho' g. \quad (25.65)$$

We retain both the global and local pressure decompositions in the following as doing so offers complementary perspectives on how pressure forces lead to vertical accelerations.

25.7.3 Vertical equation of motion

To further study how pressure and gravity lead to vertical motion, we make the Boussinesq approximation so that the velocity equation (25.60) takes the form (24.66)

$$\rho_0 (\partial_t + \mathbf{v} \cdot \nabla) \mathbf{v} + f \hat{\mathbf{z}} \wedge \rho_0 \mathbf{v} = -\nabla p - \rho g \hat{\mathbf{z}} + \rho_0 \mathbf{F} \quad (25.66a)$$

$$= -\nabla p' + \rho_0 b \hat{\mathbf{z}} + \rho_0 \mathbf{F} \quad (25.66b)$$

$$= -\nabla p_{nh} - \nabla_z p_h + \rho_0 \mathbf{F}, \quad (25.66c)$$

with particular attention given to the vertical velocity equation

$$\rho_0 (\partial_t + \mathbf{v} \cdot \nabla) w = -\partial_z p - \rho g + \rho_0 \mathbf{F} \cdot \hat{\mathbf{z}} \quad (25.67a)$$

$$= -\partial_z p' + \rho_0 b + \rho_0 \mathbf{F} \cdot \hat{\mathbf{z}} \quad (25.67b)$$

$$= -\partial_z p_{nh} + \rho_0 \mathbf{F} \cdot \hat{\mathbf{z}}. \quad (25.67c)$$

Equation (25.67c) reveals that vertical acceleration has an inviscid contribution from pressure and gravity arising just from vertical derivatives in the non-hydrostatic pressure, p_{nh} . Equation (25.67b) has contributions from vertical gradients in the dynamical pressure, p' , plus the globally referenced Archimedean buoyancy, b .

25.7.4 Effective buoyancy and motional vertical acceleration

We decompose the vertical acceleration into two conceptually distinct processes. The first is the *effective buoyancy*, which is the vertical acceleration arising from the mass/density field. Operationally, we deduce the effective buoyancy by setting velocity to zero everywhere in the expression for the vertical acceleration

$$b_{\text{eff}} \equiv \left. \frac{Dw}{Dt} \right|_{\mathbf{v}=0}. \quad (25.68)$$

As viscous friction only arises when there is fluid motion, it does not contribute to the effective buoyancy (Section 19.7.5). Hence, equations (25.67b) and (25.67c) show that the effective buoyancy is given by the two equivalent expressions

$$\rho_0 b_{\text{eff}} = \rho_0 b - \partial_z p' \Big|_{\mathbf{v}=0} = -\partial_z p_{nh} \Big|_{\mathbf{v}=0}. \quad (25.69)$$

The first expression identifies the difference between the effective buoyancy and the Archimedean buoyancy. Namely, the difference arises from the nonzero $(\partial_z p')_{\mathbf{v}=0}$. This vertical pressure gradient means that the Archimedean buoyancy is an incomplete description of the vertical acceleration. The second expression in equation (25.69) can be thought of as a generalization of the local hydrostatic relation, $\partial_z p_h = \rho_0 b$, thus defining the effective buoyancy as minus the vertical gradient of the non-hydrostatic pressure. As a test of the formalism, note that for the special case when p' is

approximated by its hydrostatic component, so that $\rho_0 b = \partial_z p'$, then there is no effective buoyancy, which also corresponds to $\partial_z p_{\text{nh}} = 0$.

If one introduces an Archimedean buoyancy anomaly in a static fluid, then the initial vertical acceleration acting on the anomaly is given by its effective buoyancy. As the anomaly evolves, motion is generated, at which point b_{eff} is an incomplete measure of vertical acceleration. In the presence of motion we must also consider the *motional vertical acceleration*. Operationally, we deduce this acceleration by setting the density to a constant within the expression for the vertical acceleration

$$\rho_0 a_{\text{motion}} \equiv \left. \frac{Dw}{Dt} \right|_{\rho=\rho_0}. \quad (25.70)$$

Since the Archimedean buoyancy vanishes when density is uniform, we know that

$$a_{\text{motion}} = -\partial_z p'|_{\rho=\rho_0} + \rho_0 \mathbf{F} \cdot \hat{\mathbf{z}} = -\partial_z p_{\text{nh}}|_{\rho=\rho_0} + \rho_0 \mathbf{F} \cdot \hat{\mathbf{z}}, \quad (25.71)$$

so that the vertical pressure gradients of p' and p_{nh} are the same when the density field is uniform.

To completely specify the decomposition of vertical acceleration requires boundary conditions for the effective buoyancy and the motional acceleration. Mathematical details are simplified by restricting concern to rigid and flat upper and lower boundaries and no lateral boundaries. At the rigid and flat boundaries we have $Dw/Dt = 0$, and this boundary condition holds whether $\mathbf{v} = 0$ or $\rho = \rho_0$, so that $b_{\text{eff}} = 0$ and $a_{\text{motion}} = 0$ also hold at these boundaries. In this case we are ensured that the net vertical acceleration is the sum

$$\rho_0 \frac{Dw}{Dt} = \rho_0 (a_{\text{motion}} + b_{\text{eff}}). \quad (25.72)$$

Hence, the conceptual decomposition of the vertical acceleration also serves as an operational decomposition.

25.7.5 Poisson equations for the pressures and accelerations

Archimedean buoyancy plus perturbation pressure

Consider the momentum equation (25.66b) written using the decomposition of pressure and gravity that introduces the Archimedean buoyancy

$$\rho_0 (\partial_t + \mathbf{v} \cdot \nabla) \mathbf{v} + f \hat{\mathbf{z}} \wedge \rho_0 \mathbf{v} = -\nabla p' + \rho_0 b \hat{\mathbf{z}} + \rho_0 \mathbf{F}. \quad (25.73)$$

As in Section 24.3, we can eliminate the time tendency by taking a divergence to render a Poisson equation for the perturbation pressure

$$-\nabla^2 p' = \rho_0 \nabla \cdot [(\mathbf{v} \cdot \nabla) \mathbf{v} + f \hat{\mathbf{z}} \wedge \mathbf{v} - b \hat{\mathbf{z}} - \mathbf{F}]. \quad (25.74)$$

The Poisson equation is linear so that we can decompose the source terms and corresponding pressures according to

$$-\nabla^2 p'_{\text{buoy}} \equiv -\rho_0 \partial_z b \quad (25.75a)$$

$$-\nabla^2 p'_{\text{motion}} \equiv \rho_0 \nabla \cdot [(\mathbf{v} \cdot \nabla) \mathbf{v} + f \hat{\mathbf{z}} \wedge \mathbf{v} - \mathbf{F}], \quad (25.75b)$$

with the buoyancy pressure perturbation, p'_{buoy} , sourced by vertical derivatives of the Archimedean buoyancy, and the motional pressure perturbation, p'_{motion} , sourced by fluid motion that gives rise to an acceleration through self-advection, Coriolis, and friction.

From the definition (25.69) of the effective buoyancy, the decompositions (25.75a) and (25.75b) allow us to identify the Poisson equation for the effective buoyancy

$$\rho_0 b_{\text{eff}} = \rho_0 b - \partial_z p'_{\text{buoy}} \implies -\nabla^2 b_{\text{eff}} = -\nabla_z^2 b \implies -\nabla^2(b_{\text{eff}} - b) = \partial_{zz} b. \quad (25.76)$$

Hence, the source for the effective buoyancy is the horizontal Laplacian of the Archimedean buoyancy. Correspondingly, the source for the difference, $b_{\text{eff}} - b$, is the vertical curvature of the Archimedean buoyancy, $\partial_{zz} b$. The decompositions (25.75a) and (25.75b) also allow us to identify the Poisson equation for the motional acceleration

$$-\nabla^2 a_{\text{motion}} = -\partial_z \nabla \cdot [(\mathbf{v} \cdot \nabla) \mathbf{v} + f \hat{\mathbf{z}} \wedge \mathbf{v} - \mathbf{F}] - \nabla^2(\mathbf{F} \cdot \hat{\mathbf{z}}). \quad (25.77)$$

Local non-hydrostatic plus local hydrostatic pressures

We can also arrive at the above results by considering the momentum equation (25.66c) written using the decomposition of pressure into is local non-hydrostatic and local hydrostatic components

$$\rho_0 (\partial_t + \mathbf{v} \cdot \nabla) \mathbf{v} + f \hat{\mathbf{z}} \wedge \rho_0 \mathbf{v} = -\nabla p_{\text{nh}} - \nabla_z p_{\text{h}} + \rho_0 \mathbf{F}. \quad (25.78)$$

A divergence of this equation leads to the Poisson equation for the non-hydrostatic pressure

$$-\nabla^2 p_{\text{nh}} = \nabla_z^2 p_{\text{h}} + \rho_0 \nabla \cdot [(\mathbf{v} \cdot \nabla) \mathbf{v} + f \hat{\mathbf{z}} \wedge \mathbf{v} - \mathbf{F}]. \quad (25.79)$$

Taking a vertical derivative and use of the hydrostatic relation leads to the Poisson equation for the vertical derivative of the non-hydrostatic pressure

$$-\nabla^2(\partial_z p_{\text{nh}}) = -g \nabla_z^2 \rho + \rho_0 \partial_z \nabla \cdot [(\mathbf{v} \cdot \nabla) \mathbf{v} + f \hat{\mathbf{z}} \wedge \mathbf{v} - \mathbf{F}]. \quad (25.80)$$

With the motional acceleration given by $(\partial_z p_{\text{nh}})_{\rho=\rho_0}$, the general expression (25.71) for a_{motion} , along with the Poisson equation (25.80) for $\partial_z p_{\text{nh}}$, renders the Poisson equation (25.77).

Setting $\mathbf{v} = 0$ in equation (25.79) renders the Poisson equation

$$-(\nabla^2 p_{\text{nh}})_{\mathbf{v}=0} = \nabla_z^2 p_{\text{h}}. \quad (25.81)$$

This equation says that the static portion of the non-hydrostatic pressure is sourced by the horizontal Laplacian of the hydrostatic pressure. Equivalently, the convergence of $(\nabla p_{\text{nh}})_{\mathbf{v}=0}$ is balanced by the divergence of the horizontal hydrostatic pressure gradient, $\nabla_z p_{\text{h}}$. Furthermore, setting $\mathbf{v} = 0$ in equation (25.80) yields the Poisson equation for the effective buoyancy

$$-\rho_0 \nabla^2 b_{\text{eff}} = g \nabla_z^2 \rho, \quad (25.82)$$

which accords with equation (25.76).

Scales for effective buoyancy and Archimedean buoyancy

One way to emphasize the distinction between the effective buoyancy equation (25.82) and that for the Archimedean buoyancy is to compare their two elliptic equations

$$-\rho_0 \nabla^2 b = g \nabla^2 \rho \longleftrightarrow -\rho_0 \nabla^2 b_{\text{eff}} = g \nabla_z^2 \rho, \quad (25.83)$$

with the first equality following trivially by definition of Archimedean buoyancy, $b = -(g/\rho_0)(\rho - \rho_0)$. The different Laplacian operators acting in equation (25.82) on b_{eff} and b means that they

scale differently. To see how, consider an Archimedean buoyancy anomaly with a horizontal scale D and vertical scale H so that the Laplacian operators scale according to $\nabla^2 \sim D^{-2} + H^{-2}$ and $\nabla_z^2 \sim D^{-2}$. Plugging into the effective buoyancy equation (25.82) leads to

$$B_{\text{eff}} = \frac{B}{1 + D^2/H^2}, \quad (25.84)$$

where B_{eff} is the scale for the effective buoyancy and B is the scale for the Archimedean buoyancy. We see that the effective buoyancy is smaller in scale than the Archimedean buoyancy, with this scale reduction arising from the pressure contribution in the effective buoyancy that is missing from the Archimedean buoyancy.

As a buoyant fluid element moves vertically, it must displace the surrounding environmental fluid. The pressure contribution to the effective buoyancy accounts for the back-reaction of the environmental fluid on the buoyant fluid element. Since the Archimedean buoyancy ignores the back-reaction, it generally over estimates the magnitude of the vertical acceleration. By accounting for the pressure forces acting on the element from the surrounding fluid, the effective buoyancy offers a more accurate measure of the static vertical forces arising from density inhomogeneities.

25.7.6 Thought experiments for effective buoyancy

We here present some thought experiments for the purpose of developing an understanding of effective buoyancy. The thought experiments are somewhat trivial physically and yet they require us to confront basic assumptions, which is generally a useful exercise.

Horizontally unstratified density

Consider a horizontally unstratified density, $\rho = \rho(z)$, on a horizontally periodic domain with the fluid in hydrostatic equilibrium. By construction, the Archimedean buoyancy exactly balances the vertical pressure gradient. In the absence of horizontal density gradients, the effective buoyancy is everywhere a harmonic function since $\nabla^2 b_{\text{eff}} = 0$. With vanishing Dirichlet boundary conditions at the rigid bottom and top of the domain, $z = 0, H$, then $b_{\text{eff}} = 0$ everywhere, signaling the absence of any vertical acceleration.

It is notable that this result holds for an arbitrary vertical profile of density, even if the density is gravitationally unstable (Section 25.5). The assumed horizontal symmetry is the key point. This assumption precludes any vertical motion since no fluid element at a single horizontal position can be vertically displaced without breaking horizontal symmetry. The only way to maintain volume conservation ($\nabla \cdot \mathbf{v} = 0$) with vertical motion is for some fluid to move up while other fluid moves down, and for that to happen requires breaking horizontal symmetry. Once a tiny seed of horizontal asymmetry is presented to the fluid, effective buoyancy is sourced by $\nabla_z^2 \rho \neq 0$, which in turn allows the gravitational instability to grow.

Vertically unstratified density

Now consider a vertically unstratified density field, $\rho = \rho(x, y)$, so that the Archimedean buoyancy has no depth dependence. This vertically “neutrally buoyant” case commonly means that a fluid element can move vertically without feeling any buoyancy forces. Indeed, such is the case when referring to Archimedean buoyancy. What about effective buoyancy?

If density is a linear function of horizontal position then the effective buoyancy is a harmonic function so that $b_{\text{eff}} = 0$. If density is a nonlinear function of the horizontal then the effective buoyancy is nonzero. Yet is there vertical motion? Again we must confront the boundary conditions to

answer this question. Here, the rigid top and bottom boundaries preclude movement of fluid across these boundaries by imparting a boundary pressure form stress (Chapter 20) acting throughout the fluid to counteract the effective buoyancy. As a result, the fluid remains static. If instead we allow for a free surface, then the effective buoyancy would cause vertical motion that then leads to horizontal convergences, thus leading to further motion. Alternatively, if we allow for depth dependence of the density, then fluid can move laterally as well as vertically.

Localized source of Archimedean buoyancy

Although useful to garner some understanding of effective buoyancy, the previous examples are not realistic. More realistic applications are concerned with Archimedean buoyancy sources localized in both the horizontal and vertical directions. In this case the buoyancy source, $\nabla_z^2 \rho \neq 0$, and boundary conditions lead to a nontrivial structure for the effective buoyancy through solving the Poisson equation. Studies listed below offer examples, both analytical and numerical, to further an understanding of how effective buoyancy offers a more complete description of vertical acceleration than Archimedean buoyancy.

25.7.7 Comments and further study

Studies from [Davies-Jones \(2003b\)](#), [Doswell and Markowski \(2004\)](#), [Jeevanjee and Romps \(2015a\)](#), [Jeevanjee and Romps \(2015b\)](#), and [Tarshish et al. \(2018\)](#), point to the use of the effective buoyancy and the limitations of Archimedean buoyancy when studying buoyancy dominated motion, such as the early stages of a buoyant thermal. Much of the material in this section was gleaned from these papers, particularly from [Jeevanjee and Romps \(2015a\)](#) and [Jeevanjee and Romps \(2015b\)](#). Chapter 2 of [Markowski and Richardson \(2010\)](#) provides a pedagogical foundation for understanding pressure forces leading to vertical motion.

The structure of b_{eff} is distinct from the Archimedean buoyancy, b , with [Jeevanjee and Romps \(2015a\)](#) and [Tarshish et al. \(2018\)](#) providing examples where b and b_{eff} can even have opposite signs. Furthermore, [Tarshish et al. \(2018\)](#) made use of an elegant analogy between the Poisson equation for b_{eff} and the Poisson equation q for certain magnetostatics problems. The analogy allowed them to derive analytical expressions for b_{eff} from spherical and elliptical Archimedean buoyancy sources.



25.8 Exercises

EXERCISE 25.1: EXAMPLES OF BUOYANCY PERIOD

Using approximate but realistic values for the observed stratification, determine the buoyancy period ($T_b = 2\pi/N$) for

- mid-latitude troposphere
- stratosphere
- ocean thermocline
- ocean abyss.

Provide references for where you obtained the observed stratification. Hint: for both the atmosphere and ocean, it is sufficient to assume stratification is dominated by potential temperature.

EXERCISE 25.2: BUOYANCY FREQUENCY FOR AN IDEAL GAS

Write the expression for the squared buoyancy frequency of an ideal gas. Hint: first derive the expression for the potential density and then take its vertical derivative as per equation (25.36).

EXERCISE 25.3: VERTICAL INTEGRAL OF N^2

The expression for squared buoyancy frequency

$$N^2 = -g \left[\frac{\partial \ln \rho_{\text{pot}}}{\partial z} \right]_{p_R=p} \quad (25.85)$$

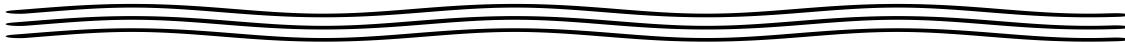
makes it tempting to consider its vertical integral according to

$$-g^{-1} \int_{-H}^{\eta} N^2 dz \stackrel{?}{=} [\ln \rho_{\text{pot}}]_{\eta} - [\ln \rho_{\text{pot}}]_{-H}. \quad (25.86)$$

Discuss what is wrong with this equation. Under what conditions is it correct?

EXERCISE 25.4: WATER LEVEL OF A BOAT WITH AND WITHOUT A STONE

Consider a boat floating in constant density water, ρ_w , contained in a tank with vertical sidewalls and cross-sectional area A . Place a stone of mass M_s and density $\rho_s > \rho_w$ in the boat and measure the water level on the tank wall, h_1 . Then throw the stone into the water. What is the new water level, h_2 , as a function of h_1 and the other properties listed above? Does the water level rise or fall along the sides of the tank as a result of throwing the stone over the side? Hint: Watch [this Physics Girl video](#).



Rotating flows

Large-scale and low frequency flows in the atmosphere and ocean are strongly affected by planetary rotation. The inviscid balance for such flows is termed *geostrophic*, in which the planetary Coriolis acceleration balances the pressure gradient acceleration. Geostrophically balanced flows exhibit some rather unfamiliar behaviors, such as motion parallel to rather than perpendicular to pressure surfaces. In this chapter, we introduce salient features of geostrophically balanced flow and the associated *thermal wind balance*. These two diagnostic relations involve no time derivatives, and so cannot be used to predict the fluid flow evolution. However, their power for diagnostic insights is unquestioned by providing a framework for interpreting large-scale and low frequency flow in the atmosphere and ocean.

After studying the basic elements of rotating flow, we examine two case studies. The first concerns *inertial waves*, which are linear fluctuations that vertically mediate information about rotation. In particular, low frequency inertial waves impart a vertically stiffened nature to rotating flows, thus providing the physical mechanism for the *Taylor-Proudman effect*. The second case study concerns the distinctive nature of isopycnal form stresses associated with geostrophically balanced eddy motions. Such form stresses are a key feature of the earth's planetary energy balance, whereby positive buoyancy in the tropics is, in part, transported meridionally through the action of geostrophic eddies.

READER'S GUIDE TO THIS CHAPTER

This chapter assumes an understanding of the primitive equations from Chapter 23 and the Coriolis acceleration from Chapters 10 and 11. The material in this chapter is fundamental to understanding the nature and mechanisms of large-scale flow in the atmosphere and ocean, so that we make great use of this chapter in the remainder of this book. Throughout the discussion we are not explicitly concerned with sphericity, thus enabling the use of Cartesian vector calculus.

Although the discussion of inertial waves in Section 26.6 is self-contained, the reader will benefit from having a basic understanding of kinematic features of linear waves, most of which will have been seen in an undergraduate physics course. Chapter 6 of *Vallis (2017)* provides a pedagogical review of wave kinematics for geophysical fluids.

26.1	Loose threads	578
26.2	Primitive equations	578
26.3	The Rossby number	579
26.3.1	Scaling for the Rossby number	579
26.3.2	Ratio of material acceleration to Coriolis acceleration	580
26.3.3	Ratio of local time tendency to Coriolis acceleration	580

26.3.4	Rossby number for a kitchen sink	580
26.3.5	Rossby number for a Gulf Stream ring	580
26.4	Geostrophic balance	581
26.4.1	Geostrophic balance is distinctly fluid mechanical	582
26.4.2	Geostrophic relation in geopotential coordinates	582
26.4.3	Cyclonic and anti-cyclonic orientation	583
26.4.4	Density gradients and thermal wind balance	583
26.4.5	Geostrophic relation in pressure coordinates	584
26.4.6	Further study	584
26.5	Planetary geostrophic mechanics	584
26.5.1	Planetary geostrophic equations	584
26.5.2	Planetary geostrophic vorticity equation	585
26.5.3	Taylor-Proudman and vertical stiffening	586
26.5.4	Meridional motion in response to vortex stretching and wind curls .	587
26.5.5	Characteristics of thermal wind balanced flows	588
26.5.6	Thermal wind balance for the atmosphere	590
26.5.7	Further study	591
26.6	Inertial waves	592
26.6.1	Inertial oscillations of a fluid ring	592
26.6.2	Formulation of the inertial wave equation	592
26.6.3	Dispersion relation for inertial waves	594
26.6.4	Group velocity	595
26.6.5	Relating the velocity components	595
26.6.6	“Near” inertial waves with $\varpi \approx \pm f $	596
26.6.7	Near zero frequency inertial waves	596
26.6.8	Connecting to Taylor-Proudman	597
26.6.9	Concerning other waves depending on Coriolis	598
26.6.10	Further study	598
26.7	Isopycnal form stress from geostrophic eddies	598
26.7.1	Zonal mean zonal form stress on an isopycnal surface	599
26.7.2	Zonal mean zonal form stress acting on an isopycnal layer	601
26.7.3	Comments and further study	602
26.8	Exercises	603

26.1 Loose threads

- Connect Section 26.6.1 to the rest of the inertial waves discussion.

26.2 Primitive equations

Throughout this chapter we make use of the hydrostatic primitive equations derived in Section 23.1

$$\partial \mathbf{u} / \partial t + (\mathbf{v} \cdot \nabla) \mathbf{u} + f \hat{\mathbf{z}} \wedge \mathbf{u} = -\rho^{-1} \nabla_z p \quad (26.1a)$$

$$\partial p / \partial z = -g \rho \quad (26.1b)$$

$$D\rho/Dt = -\rho \nabla \cdot \mathbf{v}, \quad (26.1c)$$

where the velocity vector is written using Cartesian coordinates

$$\mathbf{v} = \mathbf{u} + \hat{\mathbf{z}} w = \hat{\mathbf{x}} u + \hat{\mathbf{y}} v + \hat{\mathbf{z}} w, \quad (26.2)$$

and the horizontal gradient operator is

$$\nabla_z = \hat{\mathbf{x}} \partial_x + \hat{\mathbf{y}} \partial_y. \quad (26.3)$$

For some of the scale analysis in this chapter we assume an oceanic Boussinesq fluid (Section 24.1), in which case the mass continuity equation (26.1c) becomes the non-divergent condition on the velocity

$$\nabla \cdot \mathbf{v} = 0. \quad (26.4)$$

Furthermore, ρ in the Boussinesq horizontal momentum equation (26.1a) is converted to a constant reference density, ρ_0 , and yet it retains its full form when appearing in the hydrostatic equation since it is there multiplied by the gravitational acceleration.

26.3 The Rossby number

Large-scale geophysical fluid flows are strongly influenced by the earth's rotation. Indeed, the earth can be considered a rapidly rotating planet for much of the observed large-scale motion of the ocean and atmosphere. There are two points to emphasize in this regard. First, much of the ocean and atmosphere motion is close to solid-body rotation, in which weather patterns and ocean circulation are best viewed relative to the rotating earth rather than from an inertial frame. Second, human scale horizontal length scales are generally far too small to take a direct notice of the planetary rotation. This point is quantified by considering the *Rossby number*, which includes a horizontal length scale, a velocity scale, and angular rotation speed.

26.3.1 Scaling for the Rossby number

The Rossby number measures the ratio of the Coriolis acceleration to the material acceleration (acceleration of a fluid particle). The material acceleration has two contributions: one from local time tendencies and one from advection. We expose typical characteristic scales for the acceleration by writing

$$\frac{\partial \mathbf{u}}{\partial t} + (\mathbf{v} \cdot \nabla) \mathbf{u} \sim \frac{U}{T} + \frac{U^2}{L} + \frac{WU}{H}, \quad (26.5)$$

where U, W are typical horizontal and vertical velocity scales, L, H are typical horizontal and vertical length scales, and T is a typical time scale (recall a similar scale analysis for the hydrostatic balance in Section 23.3). Likewise, the Coriolis acceleration scales as

$$f \hat{\mathbf{z}} \wedge \mathbf{u} \sim f_0 U, \quad (26.6)$$

where f_0 is the scale for the Coriolis parameter. From the continuity equation for non-divergent flow ($\nabla \cdot \mathbf{v} = 0$) we see that the vertical and horizontal velocity scales are related by¹

$$W/H \sim U/L \implies W \sim U(H/L). \quad (26.7)$$

In either non-divergent or divergent flows, we assume the vertical to horizontal grid aspect ratio is small

$$\alpha_{\text{aspect}} = H/L \ll 1, \quad (26.8)$$

as per the hydrostatic approximation discussed in Section 23.3. Consequently, the vertical velocity scale is much less than the horizontal

$$W \ll U. \quad (26.9)$$

¹For divergent flows we can replace W with the scale for motion across pressure surfaces.

26.3.2 Ratio of material acceleration to Coriolis acceleration

Taking the ratio of the advection scale to the Coriolis scale leads to our first expression for the Rossby number

$$\text{Ro} = \frac{U^2/L}{f_0 U} = \frac{U}{f_0 L}. \quad (26.10)$$

Due to the latitudinal variation of the Coriolis parameter, the Rossby number is generally small in magnitude near the poles and large in the tropics. This functional dependence is modified by noting that the length and velocity scales are not constant.

26.3.3 Ratio of local time tendency to Coriolis acceleration

A complementary way to understand the Rossby number is to consider it as the ratio of the local time tendency for the horizontal velocity to the Coriolis acceleration

$$\text{Ro} = \frac{U/T}{U f_0} = \frac{1/T}{f_0}. \quad (26.11)$$

Thus, for motions that have a low frequency T^{-1} compared to the *inertial frequency* f_0 , the Rossby number is small. In both ways of writing the Rossby number, we associate small Ro with regimes of flow where the earth's rotation plays a crucial role in the dynamics. With small Rossby number, both the local time derivative and the advective acceleration are smaller than the Coriolis acceleration.

26.3.4 Rossby number for a kitchen sink

Consider flow in a kitchen sink (left panel of Figure 26.1). Here, the length scale is $L = 1$ m (sink size) and the velocity scale is $U = 0.01 - 0.1$ m s $^{-1}$, thus giving a typical time scale for sink motion of $L/U \approx 10$ s – 100 s. Hence, at 30° latitude, where $f = 2\Omega \sin \phi = \Omega$, the Rossby number for fluid motion in a sink is

$$\text{Ro}_{\text{sink}} \approx 10^2 - 10^3. \quad (26.12)$$

The effects from planetary rotation are tiny on these length scales, so that the Coriolis force is negligible for kitchen sink fluid dynamics. Hence, it is extremely difficult to experimentally determine a correlation between the hemisphere (northern or southern) to the rotational direction of water leaving a sink drain.

26.3.5 Rossby number for a Gulf Stream ring

For a Gulf Stream ring (right panel of Figure 26.1), the typical length scale is $L = 10^5$ m and velocity scale is $U = 0.1 - 1.0$ m s $^{-1}$, thus leading to a time scale $L/U \approx 10^5 - 10^6$ s. At 30° latitude the Rossby number is

$$\text{Ro}_{\text{ring}} \approx 10^{-2} - 10^{-1}. \quad (26.13)$$

Flow features of such large length scales can feel the planetary rotation so that the Coriolis acceleration is central to dynamics of Gulf Stream rings, as reflected in the small magnitude of the associated Rossby number.

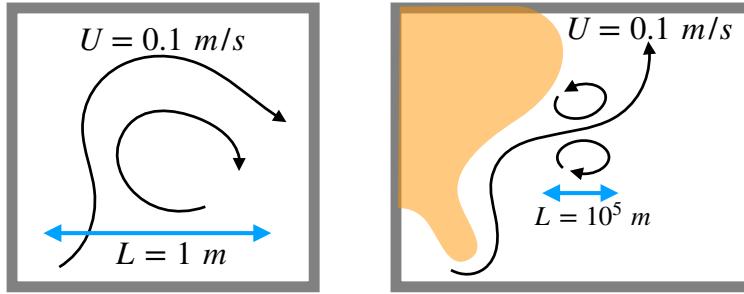


FIGURE 26.1: Estimating the Rossby number for flow in a kitchen sink (left panel) and rings spawned from the Gulf Stream (right panel). The kitchen sink has velocity scales on the order of $U \sim 0.01 - 0.1 \text{ m s}^{-1}$ whereas Gulf Stream rings have velocity scales on the order $U \sim 0.1 - 1.0 \text{ m s}^{-1}$. Their length scales are much more distinct, with the scale for a sink $L \sim 1 \text{ m}$ whereas for the Gulf Stream rings $L \sim 10^5 \text{ m}$. Taking the Coriolis parameter at 30° leads to $\text{Ro}_{\text{sink}} \sim 10^2 - 10^3$ and $\text{Ro}_{\text{ring}} \sim 10^{-2} - 10^{-1}$. The planetary Coriolis acceleration is central to Gulf Stream ring dynamics whereas it is utterly negligible for the kitchen sink.

26.4 Geostrophic balance

Under the influence of horizontal pressure forces, a fluid accelerates down the pressure gradient (movement from high pressure to low pressure). In the presence of rotation, a nonzero horizontal velocity couples to the Coriolis parameter, f , thus giving rise to a nonzero horizontally oriented Coriolis acceleration $-f \hat{z} \wedge \mathbf{u}$. The Coriolis acceleration acts perpendicular to the fluid motion

$$\mathbf{u} \cdot (\hat{z} \wedge \mathbf{u}) = 0, \quad (26.14)$$

and as such it affects the fluid motion but does not alter kinetic energy; i.e., it does zero work on the fluid.² In the northern hemisphere where $f > 0$, the Coriolis acceleration acts to the right of the parcel motion, thus causing counter-clockwise motion around low pressure centers and clockwise motion around high pressure centers (Figure 26.3). In the southern hemisphere, where $f < 0$, the Coriolis acceleration acts in the opposite direction. So in summary, the geostrophic balance is a diagnostic relation where the pressure gradient acceleration balances the Coriolis acceleration so that the net acceleration acting on a fluid element is zero. In Figure 26.2 we offer a heuristic model to help understand the geostrophic balance and remember its orientation.

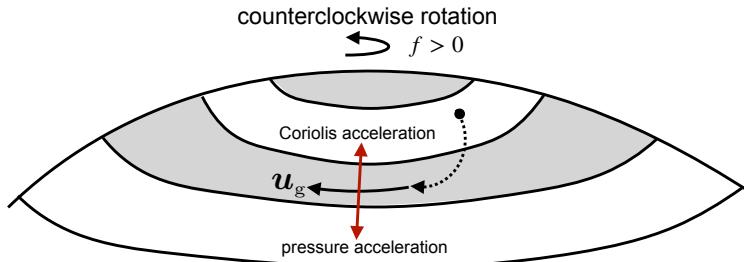


FIGURE 26.2: A particle on a rotating and frictionless hill as a model to help understand geostrophic balance. As the particle moves downhill it picks up a rightward component to its trajectory as a result of the Coriolis acceleration. Equilibrium arises when the downhill pressure gradient acceleration balances the uphill Coriolis acceleration. The downhill gravitational acceleration acting on the particle is the analog of an equatorward pressure acceleration acting on a fluid element.

²These characteristics of the Coriolis acceleration are directly analogous to the Lorentz force in electrodynamics ([Jackson, 1975](#)).

26.4.1 Geostrophic balance is distinctly fluid mechanical

When pressure and Coriolis forces balance, parcel motion is said to be in *geostrophic* balance, whereby large-scale atmospheric winds and ocean currents generally follow isobars (lines of constant pressure). Recall from Chapter 10 that point particles also experience a Coriolis acceleration when viewed in a rotating reference frame. However, geostrophic balance is not afforded to particles since particles do not experience a pressure force that can balance the Coriolis force. Hence, the geostrophic balance is a distinctly fluid mechanical phenomena.

26.4.2 Geostrophic relation in geopotential coordinates

Mathematically, the geostrophic balance becomes important when the Rossby number is small. When the Rossby number is small and friction is negligible, the leading order dynamical balance in the horizontal momentum equation (26.1a) is between the Coriolis acceleration and horizontal pressure gradient acceleration

$$f \hat{z} \wedge \mathbf{u}_g = -\rho^{-1} \nabla_z p, \quad (26.15)$$

or equivalently, we have the expression for the geostrophic velocity³

$$\mathbf{u}_g = \frac{\hat{z} \wedge \nabla p}{f \rho} \implies u_g = -\frac{1}{f \rho} \frac{\partial p}{\partial y} \quad \text{and} \quad v_g = \frac{1}{f \rho} \frac{\partial p}{\partial x}. \quad (26.16)$$

Note that the equator is special since the Coriolis parameter, $f = 2\Omega \sin \phi$, vanishes, thus precluding the relevance of geostrophy near the equator. Instead, at the equator flow generally moves down the horizontal pressure gradient since there is no Coriolis acceleration to turn the flow.

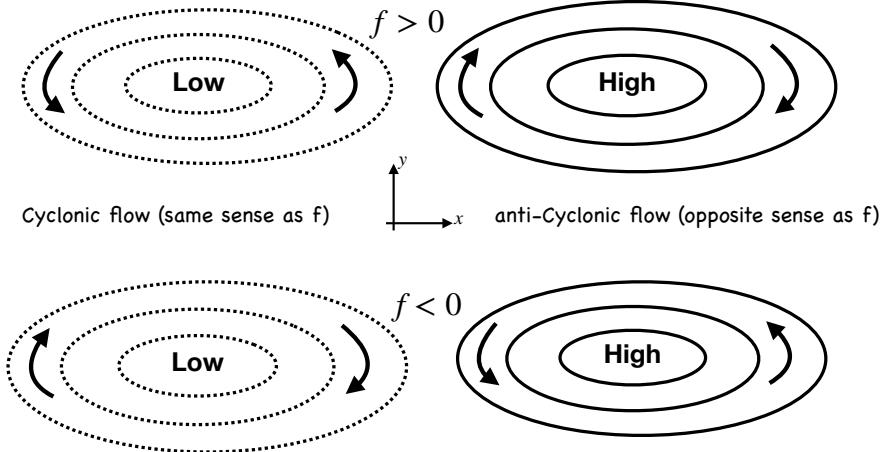


FIGURE 26.3: Geostrophic motion around low and high pressure centers in the northern hemisphere and southern hemisphere ($f = 2\Omega \sin \phi > 0$ in the north and $f < 0$ in the south). Upper panel: the counter-clockwise motion around the low pressure center in the northern hemisphere is in the same sense as the planetary rotation, and is thus termed cyclonic. Cyclonic motion in the southern hemisphere occurs in a clockwise direction, again corresponding to the planetary rotation direction as viewed from the south. Geostrophic motion around a high pressure center is counter to the planetary rotation in both hemispheres, and is thus termed anti-cyclonic.

³We can write either ∇ or ∇_z in equation (26.16). The reason is that the $\hat{z} \wedge$ operator selects only the horizontal portion of the gradient.

26.4.3 Cyclonic and anti-cyclonic orientation

When oriented in the same sense as the earth's rotation (i.e., same sign of the Coriolis parameter) rotational motion is said to be in a *cyclonic* sense. Oppositely oriented motion is *anti-cyclonic*. For example, geostrophic motion around a low pressure center in the northern hemisphere is counter-clockwise (Figure 26.3). Using the right hand rule, this motion represents a positively oriented rotation. Hence, with $f > 0$ in the north, counter-clockwise motion is cyclonic. In the southern hemisphere, geostrophic motion around a low pressure center is clockwise, which is a negatively oriented rotational motion (again, recall the right hand rule). In the south where $f < 0$, clockwise motion around a low pressure center represents cyclonic motion (Figure 26.3).

26.4.4 Density gradients and thermal wind balance

The horizontal momentum is affected by horizontal pressure gradient forces. Furthermore, the hydrostatic balance says that the vertical derivative of the horizontal pressure gradient is determined by horizontal density gradients

$$\frac{\partial(\nabla_z p)}{\partial z} = -g \nabla_z \rho. \quad (26.17)$$

Hence, in the presence of horizontal density gradients, the horizontal pressure gradient forces are depth dependent. Correspondingly, the horizontal velocity field experiences a depth dependent pressure force.

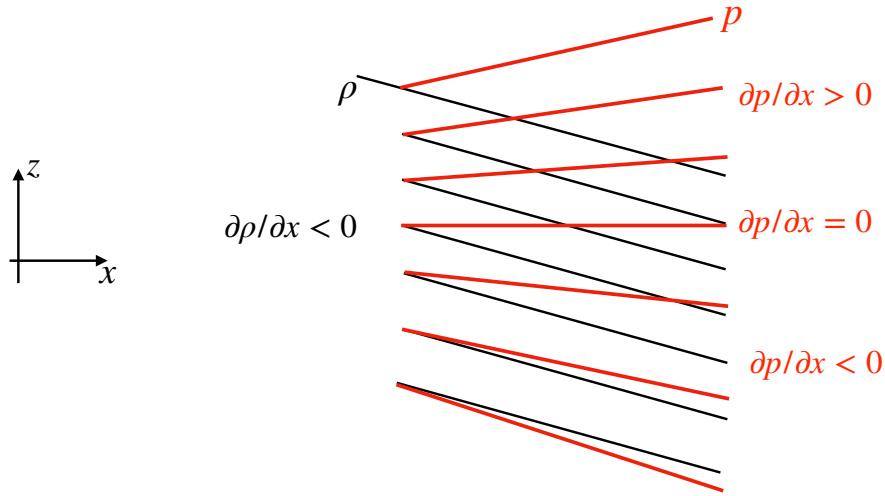


FIGURE 26.4: Horizontal density gradients support a vertical dependence to the horizontal gradient of the hydrostatic pressure via $\partial(\nabla_z p)/\partial z = -g \nabla_z \rho$. This figure depicts a constant horizontal density gradient with $\partial \rho / \partial x < 0$, thus leading to an increase in the zonal pressure gradient with height, $\partial(\partial p / \partial x) / \partial z > 0$. Depending on the thickness of the fluid column, the horizontal pressure gradient can change sign when moving up in the column, as shown here. Compare this figure to Figure 23.2, which discusses how to compute horizontal pressure gradients in a hydrostatic fluid.

We illustrate this depth dependence in Figure 26.4 with a depth independent horizontal density gradient, $\partial \rho / \partial x = \text{constant} < 0$ (see also Figure 23.2), thus leading to a depth dependent horizontal gradient in the hydrostatic pressure. This figure also illustrates how the sign of the horizontal pressure gradient can change when moving in the vertical, depending on the value of the gradient at depth. It also illustrates how horizontal density gradients lead to a nonzero baroclinicity vector

$$\mathbf{B} = \frac{\nabla \rho \wedge \nabla p}{\rho^2}. \quad (26.18)$$

As shown in Section 36.5, a nonzero baroclinicity imparts a torque on fluid elements that acts as a source for vorticity.

The depth dependence to the horizontal pressure gradient leads to a vertical shear in the horizontal geostrophic velocity

$$\frac{\partial(\rho f \mathbf{u}_g)}{\partial z} = \hat{\mathbf{z}} \wedge \nabla(\partial p / \partial z) = -g \hat{\mathbf{z}} \wedge \nabla \rho. \quad (26.19)$$

This connection between horizontal density gradients and vertical shears in the geostrophic velocity is known as the *thermal wind balance*, which we return to in Section 26.5.5.

26.4.5 Geostrophic relation in pressure coordinates

The hydrostatic balance

$$\partial p / \partial z = -\rho g \quad (26.20)$$

can be used to eliminate density from the geostrophic balance (26.15) to render

$$f \hat{\mathbf{z}} \wedge \mathbf{u}_g = \frac{g \nabla_z p}{\partial p / \partial z}. \quad (26.21)$$

The right hand side is minus the lateral gradient of the geopotential, $\Phi = g z$, as computed along surfaces of constant pressure⁴

$$f \hat{\mathbf{z}} \wedge \mathbf{u}_g = -\nabla_p \Phi \implies f \mathbf{u}_g = \hat{\mathbf{z}} \wedge \nabla_p \Phi. \quad (26.22)$$

This is a useful expression of geostrophy for the compressible atmosphere.

26.4.6 Further study

Visualizations from rotating tank experiments provide very useful illustrations of the Coriolis acceleration and geostrophic balance, such as the experiments shown near the 10 minute mark in [this video from Prof. Fultz](#).

26.5 Planetary geostrophic mechanics

We here introduce the planetary geostrophic (PG) equations, which have found great use in describing elements of the large-scale laminar ocean circulation. We state the equations and discuss their physical implications, deferring a systematic derivation for later. In particular, the shallow water PG equations are derived in Section 40.4 and the continuously stratified PG equations are derived in Section 41.2.

26.5.1 Planetary geostrophic equations

The governing equations for planetary geostrophy are based on the hydrostatic Boussinesq equations stated in Section 24.1.7, with the assumption of a steady state linear and frictional/geostrophic balance for the horizontal momentum

$$\rho_0 f (\hat{\mathbf{z}} \wedge \mathbf{u}) = -\nabla p - \rho g \hat{\mathbf{z}} + \partial \boldsymbol{\tau} / \partial z \quad (26.23a)$$

$$\nabla_z \cdot \mathbf{u} + \partial w / \partial z = 0 \quad (26.23b)$$

$$Db/Dt = \dot{b}. \quad (26.23c)$$

⁴See Section 43.11.2 for details of computing derivatives using generalized vertical coordinates.

The stress, $\boldsymbol{\tau}$, acts just in the horizontal directions so that the vertical component of the momentum equation (26.23a) is the hydrostatic balance

$$\partial p / \partial z = -\rho g. \quad (26.24)$$

We now list some of the key physical attributes of these equations.

Velocity is slaved to buoyancy

The only time derivative appearing in the PG equations appears in the buoyancy equation (26.23c). All other equations are diagnostic. As the buoyancy evolves, the hydrostatic pressure changes and so too does the geostrophic velocity. In effect, the velocity is a slave to the buoyancy field.

Planetary geostrophy admits no turbulence

The momentum equation is linear since PG drops the nonlinear advection of momentum. Hence, there is no turbulence phenomena in the planetary geostrophic fluid since turbulence relies on the nonlinear momentum advection. Instead, planetary geostrophy is used to describe laminar ocean circulation features at the large-scales.

Vertical transfer of horizontal stress and subgrid scale parameterizations

We introduced a horizontal stress (dimensions of force per area) into the momentum equation

$$\boldsymbol{\tau} = (\tau^x, \tau^y, 0). \quad (26.25)$$

This stress is associated with vertical transfer of momentum in the ocean interior through vertical viscosity, as well as vertical transport of momentum from the atmosphere to the ocean. The vertical stress transport is enhanced by waves and turbulent features such as mesoscale eddies. However, such transient processes are not represented by planetary geostrophy. Hence, they must be parameterized, which generally leads to an enhanced vertical viscosity relative to that from molecular viscosity. In general, all models (analytic or numerical) of planetary scale circulations are too coarse to resolve the scales over which molecular viscosity dominates the frictional dissipation. Consequently, it is necessary to provide rational *subgrid-scale (SGS) parameterizations* of the variety of physical processes that are unresolved by the model. We have more to say about the parameterization of vertical transfer of horizontal momentum in Section 50.2.4.

26.5.2 Planetary geostrophic vorticity equation

The vertical component of relative vorticity is given by

$$\zeta = \partial v / \partial x - \partial u / \partial y, \quad (26.26)$$

with a thorough discussion of vorticity given in Chapter 36. Here, we form the relative vorticity budget for the planetary geostrophic system by taking the curl of the momentum equation.

Curl of the PG momentum equation

Taking the curl of the momentum equation (26.23a), and rearranging terms, leads to the planetary geostrophic vorticity equation

$$-\rho_o f \frac{\partial \mathbf{u}}{\partial z} + \hat{z} \rho_o \nabla \cdot (f \mathbf{u}) = -g \nabla \wedge (\hat{z} \rho) + \frac{\partial (\nabla \wedge \boldsymbol{\tau})}{\partial z}. \quad (26.27)$$

Note that $\nabla \cdot (f \mathbf{u}) = \nabla_z \cdot (f \mathbf{u})$ since \mathbf{u} is the horizontal velocity vector. Introducing the Archimedean buoyancy (Section 24.1.2)

$$b = -g(\rho - \rho_o)/\rho_o \quad (26.28)$$

leads to

$$-f \frac{\partial \mathbf{u}}{\partial z} + \hat{z} \nabla_z \cdot (f \mathbf{u}) = \nabla \wedge (\hat{z} b) + \frac{1}{\rho_o} \frac{\partial (\nabla \wedge \boldsymbol{\tau})}{\partial z}. \quad (26.29)$$

Pressure gradients are removed from the vorticity equation

One of the key reasons to study vorticity is that its evolution equation is not explicitly affected by the pressure gradient since the curl of the pressure gradient vanishes. Eliminating the pressure force affords a simpler evolution equation for vorticity, and we pursue that evolution more thoroughly in Chapter 36. Pressure perturbations are rapidly transferred in a fluid, either by acoustic waves in divergent flows (Section 18.5) or instantaneously in non-divergent flows (Section 24.3). Hence, in the absence of pressure directly affecting the vorticity equation, contributions to vorticity evolution are more localized in space than contributions to velocity evolution.

Relative vorticity is absent from the PG vorticity equation

It is notable that there is no appearance of the relative vorticity, $\zeta = \partial_x v - \partial_y u$, in the planetary geostrophic vorticity equation (26.23a). The reason is that we dropped the material time derivative of velocity when forming the planetary geostrophic momentum equation (26.23a). By doing so, we drop all expressions of ζ in the vorticity equation. Planetary geostrophy is valid for those cases where

$$|\zeta| \ll |f|, \quad (26.30)$$

which means the absolute vorticity (sum of planetary vorticity plus relative vorticity) is dominated by the planetary vorticity. We encounter more complete versions of the vorticity equation in Chapter 36 where we do not make the planetary geostrophic assumption.

Rather than taking the curl of the planetary geostrophic momentum equation, we could have also derived the vorticity equation (26.29) by taking planetary geostrophic scaling in the full vorticity equation. We choose here the path through the planetary geostrophic momentum equation since we have yet to discuss the full vorticity equation (see Chapter 36).

26.5.3 Taylor-Proudman and vertical stiffening

Consider the vorticity equation (26.29) on an f -plane with zero friction, in which the horizontal geostrophic motion is horizontally non-divergent

$$\nabla_z \cdot \mathbf{u} = 0 \quad f\text{-plane geostrophy.} \quad (26.31)$$

Use of continuity (equation (26.23b)) means there is no vertical stretching of a vertical material line element (Section 14.2.6)

$$\partial w / \partial z = 0. \quad (26.32)$$

As shown in Chapter 36, a vortex tube exhibits the same kinematics as a material line element described in Section 14.2. Hence, $\partial w / \partial z = 0$ means there is no vertical stretching of a vortex tube in the planetary geostrophic fluid. This result has important implications that we now describe.

Flat bottom boundary and columnar motion

If there is a solid flat bottom to the domain, then the vertical velocity vanishes at that surface. With $\partial_z w = 0$ in the interior as well, w vanishes throughout the domain. Hence, the fluid has zero vertical velocity, and motion occurs on horizontal planes perpendicular to the rotation axis; i.e., the flow is two-dimensional. We furthermore assume zero horizontal buoyancy gradients, so that the vorticity equation (26.29) implies that the horizontal velocity has zero vertical shear

$$\partial \mathbf{u} / \partial z = 0 \quad f\text{-plane and homogeneous density.} \quad (26.33)$$

This result is known as the *Taylor-Proudman effect*.

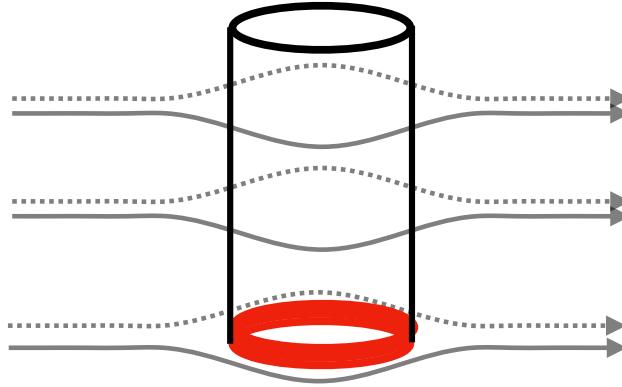


FIGURE 26.5: The Taylor-Proudman effect (26.33) says that horizontal flow in a homogeneous rapidly rotating fluid over a flat bottom (where $w = 0$) is depth-independent. Hence, when flow encounters an obstacle anywhere in the column, such as the red ring shown here at the bottom, then flow throughout the full depth coherently moves around the obstacle. The result is a vertically stiffened motion known as *Taylor columns*.

Relevance to the ocean and atmosphere

In the ocean and atmosphere, the assumptions leading to the Taylor-Proudman effect are rarely satisfied due to the presence of stratification (i.e., vertical density variations). Nonetheless, there is a tendency for vertical velocities to be quite small due to the effects of rotation; even smaller than the non-divergent flow scaling $W/H \sim U/L$ would indicate (see Section 26.3.1). Additionally, for unstratified or linearly stratified fluids, there is a tendency for geostrophic turbulence to cascade energy into the *gravest vertical mode*; i.e., the largest vertical scale. This largest vertical scale mode is termed the *barotropic mode*, and motion of this mode is predominantly horizontal and depth independent. Smaller vertical scales of variation are captured by an infinite hierarchy of *baroclinic modes*. The process of moving energy to the barotropic mode is referred to as *barotropization*.

26.5.4 Meridional motion in response to vortex stretching and windcurls

The vertical component to the vorticity balance (26.29) leads to

$$\nabla_z \cdot (f \mathbf{u}) = \frac{1}{\rho_0} \frac{\hat{z} \cdot \partial (\nabla \wedge \boldsymbol{\tau})}{\partial z} \quad (26.34)$$

which can be written

$$\beta v = -f \nabla_z \cdot \mathbf{u} + \frac{1}{\rho_0} \frac{\hat{z} \cdot \partial (\nabla \wedge \boldsymbol{\tau})}{\partial z} \quad (26.35)$$

where

$$\beta = \partial f / \partial y \quad (26.36)$$

is the gradient of planetary vorticity. The continuity equation (26.23b) can be used to remove the horizontal divergence, which yields the vorticity balance

$$\beta v = f \frac{\partial w}{\partial z} + \frac{1}{\rho_0} \hat{z} \cdot \partial (\nabla \wedge \boldsymbol{\tau}). \quad (26.37)$$

The left hand side is the meridional advection of planetary vorticity. The first term on the right hand side arises from vortex stretching by planetary vorticity; i.e., planetary induction or the β -effect discussed in Section 36.7.2. The second term is the vertical divergence of the curl of the stress.

Reading the vorticity equation (26.37) from right to left indicates that any process leading to a vorticity source via vortex stretching or wind stress curls must be balanced by meridional motion. That is, the fluid responds to vortex stretching and wind curls by moving meridionally through the planet's vorticity field. In a fluid that maintains planetary geostrophic balance, meridional movement is the only means for it to respond to vorticity input since its vorticity is set by planetary vorticity. Conversely, reading the equality (26.37) from the left to the right reveals that any meridional motion itself must be balanced by vortex stretching plus wind stress curls. We study the depth integrated form of this vorticity equation in Section 41.4, where we see how the depth integrated meridional flow is affected by stretching created by boundary torques.

26.5.5 Characteristics of thermal wind balanced flows

Horizontal components to the inviscid vorticity equation (equation (26.29) with $\boldsymbol{\tau}$ set to zero) form the thermal wind balance

$$f \partial \mathbf{u} / \partial z = -\nabla \wedge (\hat{z} b) = \hat{z} \wedge \nabla b, \quad (26.38)$$

which takes on the component form

$$f \partial u / \partial z = -\partial b / \partial y \quad \text{and} \quad f \partial v / \partial z = \partial b / \partial x. \quad (26.39)$$

As seen already in Section 26.4.4, these relations can also be derived directly from taking the vertical derivative of the horizontal momentum equation (26.23a) and then using the horizontal gradient of the hydrostatic balance (26.24). In either case, the thermal wind balance (26.38) says that the horizontal geostrophic velocity possesses a vertical shear where the buoyancy field has a horizontal gradient. Buoyancy with a horizontal gradient is termed *baroclinic* since it leads to a nonzero baroclinicity vector that provides a source for vorticity (see Section 36.8.2). Correspondingly, it is only the baroclinic (depth dependent) piece of geostrophic velocity that is related to horizontal buoyancy gradients. The depth-independent flow is not constrained by horizontal buoyancy gradients.

Thermal wind, the atmospheric jet stream and the Antarctic Circumpolar Current

Due to the increased solar radiation reaching the equator relative to the poles, the zonal averaged temperature generally reduces poleward. This poleward reduction in temperature corresponds to a poleward reduction in buoyancy. Also, for a stably stratified fluid, density increases with depth so that buoyancy decreases. Figure 26.6 illustrates this situation.

A zonal average around a zonally symmetric globe, such as in the Antarctic Circumpolar Current, removes all zonal variations, thus putting $\partial_x \rho = 0$ and so rendering the zonally averaged thermal wind relation

$$f \frac{\partial \bar{u}}{\partial z} = \frac{g}{\rho_0} \frac{\partial \bar{\rho}}{\partial y} = -\frac{\partial \bar{b}}{\partial y} > 0, \quad (26.40)$$

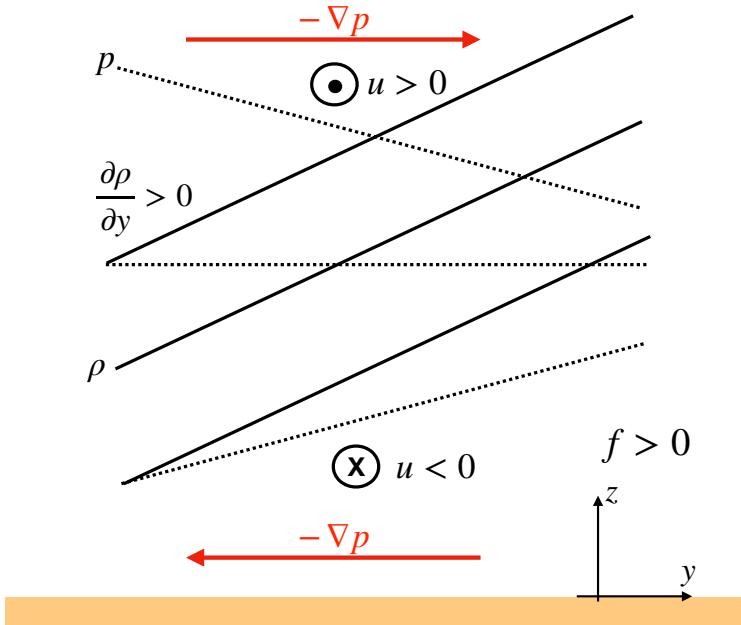


FIGURE 26.6: Schematic of the density and hydrostatic pressure fields and the associated thermal wind balanced flow in the northern hemisphere ($f > 0$) with north to the right and east out of the page. We show surfaces of constant density (solid lines) and constant pressure (isobars; dashed lines). Density increases poleward ($\partial\rho/\partial y > 0$) so that, according to the discussion surrounding Figure 26.4, the meridional pressure gradient decreases when moving upward, $\partial(\partial p/\partial y)/\partial z = -g \partial_y \rho < 0$. We illustrate isobars with an equatorward directed down gradient pressure force at lower elevations ($-\partial p/\partial y < 0$) and poleward directed pressure force at higher elevations ($-\partial p/\partial y > 0$). The zonal geostrophic wind is in geostrophic balance with these pressure gradients, with a westward zonal flow at lower elevations (easterly winds) and eastward flow at higher elevations (westerly winds). This flow configuration creates an eastward vertical shear of the zonal geostrophic winds, $\partial u_g/\partial z > 0$.

where $\overline{(\)}$ is the zonal mean operator. In the northern hemisphere, $\partial_y \bar{b} < 0$, so that the zonal averaged thermal wind shear is positive, $\partial_z \bar{u} > 0$. For example, a westerly zonal wind (blowing to the east) strengthens with height (easterly thermal wind shear). In the Southern Hemisphere, $f < 0$ with poleward decreasing buoyancy, $\partial_y \bar{b} > 0$, means there is also an eastward thermal wind shear. Note that movement towards the poles, where $|f|$ increases, leads to a smaller thermal wind shear given the same buoyancy gradient.

Diagnosing geostrophic velocity from the buoyancy field

Vertical integration of the thermal wind relation (26.38) between two constant depth surfaces leads to

$$\mathbf{u}(z) = \mathbf{u}(z_{\text{ref}}) - f^{-1} \nabla \wedge \hat{\mathbf{z}} \int_{z_{\text{ref}}}^z b \, dz. \quad (26.41)$$

Hence, knowledge of the buoyancy field (e.g., through hydrographic measurements of temperature and salinity in the ocean), along with knowledge of the geostrophic velocity at a single depth, allows for determination of the full geostrophic velocity in terms of density. Unfortunately, specification of the unknown reference velocity is unavailable just from hydrographic measurements. This is the origin of the *depth of no motion* problem in diagnostic oceanography.

26.5.6 Thermal wind balance for the atmosphere

The large-scale atmosphere is compressible and approximately in hydrostatic balance. The expression for geostrophic balance (26.22) in pressure coordinates is a suitable starting point to derive thermal wind for the atmosphere. For this purpose, we take the pressure derivative, $\partial/\partial p$, of (26.22) to render

$$f \frac{\partial \mathbf{u}}{\partial p} = \hat{\mathbf{z}} \wedge \nabla_p \left[\frac{\partial \Phi}{\partial p} \right]. \quad (26.42)$$

The hydrostatic relation $\partial p/\partial z = -\rho g$ takes the form

$$\frac{\partial p}{\partial \Phi} = -\rho \Rightarrow \frac{\partial \Phi}{\partial p} = -1/\rho \quad (26.43)$$

in which case

$$f \frac{\partial \mathbf{u}}{\partial p} = -\hat{\mathbf{z}} \wedge \nabla_p (1/\rho). \quad (26.44)$$

Ideal gas atmosphere

The specific volume takes the following form for an ideal gas atmosphere (see Section 21.9.1)

$$\rho^{-1} = R^M T/p. \quad (26.45)$$

Since the horizontal derivative in the thermal wind relation (26.44) is along pressure surfaces, we have

$$f \frac{\partial \mathbf{u}}{\partial p} = -\frac{R^M}{p} [\hat{\mathbf{z}} \wedge \nabla_p T]. \quad (26.46)$$

This expression gives rise to the name “thermal wind”, with vertical shears of the horizontal velocity generated by horizontal temperature gradients along isobars.

As for the ocean in equation (26.41), we vertically integrate the thermal wind expression (26.46), here between two pressure levels

$$\mathbf{u}(p_A) - \mathbf{u}(p_B) = f^{-1} R^M \hat{\mathbf{z}} \wedge \nabla_p \left[\int_{p_A}^{p_B} \frac{T dp}{p} \right], \quad (26.47)$$

where $p_A < p_B$, so that p_A is at a higher altitude than p_B . We define the thermal wind shear as the difference between the wind aloft (higher altitude and lower pressure) from that at a lower altitude (greater pressure)

$$\mathbf{u}_T = \mathbf{u}(p_A) - \mathbf{u}(p_B) \quad \text{with } p_A < p_B \quad (26.48)$$

so that

$$\mathbf{u}_T = \frac{R^M}{f} \hat{\mathbf{z}} \wedge \nabla_p \bar{T}^{\ln p}, \quad (26.49)$$

where we introduced the log-pressure weighted temperature between the two pressure surfaces

$$\bar{T}^{\ln p} = \int_{p_A}^{p_B} \frac{T dp}{p}. \quad (26.50)$$

The relation (26.49) means that on the f -plane, R^M/f times the log-pressure weighted temperature serves as a streamfunction for the thermal wind shear. Reconsider the previous example where the polar regions are colder than tropics, so that in the northern hemisphere on pressure surfaces, $\partial \bar{T}^{\ln p}/\partial y < 0$. Hence, the zonal westerly winds increase in magnitude with height. We depict this situation in Figure 26.7. Furthermore, the thermal wind shear points to the east. For a more general flow in the northern hemisphere, cold air sits on the left side of the thermal wind shear and warm air on the right. The opposite orientation holds for the southern hemisphere where the Coriolis parameter is negative, $f < 0$

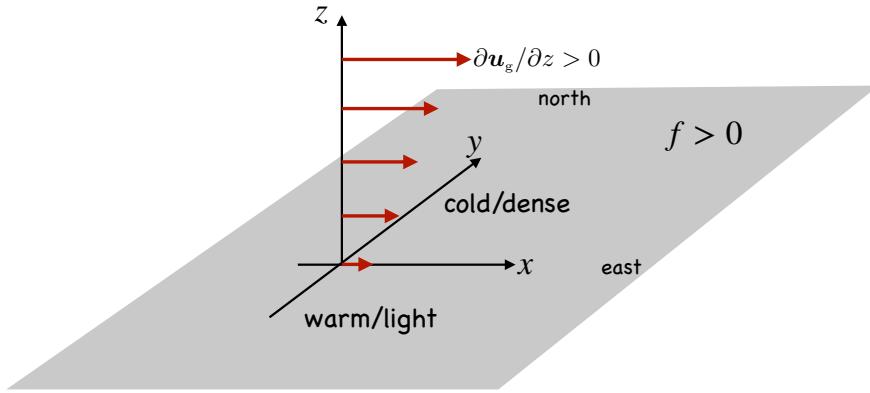


FIGURE 26.7: Thermal wind shear in the northern hemisphere ($f > 0$) middle latitude atmosphere, whereby cold/dense air sits to the north and warm/light air to the south. The zonal geostrophic winds increase to the east when rising in elevation, $\partial u_g / \partial z > 0$. We say that the zonal winds have an eastward thermal wind shear. In general, a geostrophic wind in the northern hemisphere atmosphere has cold/dense air to the left when facing downwind, whereas the opposite orientation holds for the southern hemisphere where $f < 0$.

Barotropic flow

Return to the thermal wind equation (26.44)

$$f \frac{\partial \mathbf{u}}{\partial p} = -\hat{\mathbf{z}} \wedge \nabla_p (1/\rho) = \frac{\hat{\mathbf{z}} \wedge \nabla_p \rho}{\rho^2}. \quad (26.51)$$

For the special case of density a function just of the pressure, $\rho = \rho(p)$, then $\nabla_p \rho = 0$. This situation defines a *barotropic* flow, which is characterized here by a horizontal geostrophic velocity with zero vertical variations. Note that we are here only concerned with the geostrophic flow. A density related to pressure through $\rho = \rho(p)$ can still support vertical variations of the ageostrophic flow.

We further discuss barotropic flow in Section 36.3 as part of our study of vorticity. As detailed in that discussion, the general definition of a barotropic flow is one whereby the *baroclinicity* vector vanishes, $\mathbf{B} = \nabla p \wedge \nabla(1/\rho) = 0$. The functional relation $\rho = \rho(p)$ (equivalently $p = p(\rho)$) is a sufficient condition for vanishing baroclinicity. As seen in Section 36.5, for a barotropic flow there is no generation of vorticity through the torques created when density isosurfaces deviate from pressure isosurfaces (isobars).

26.5.7 Further study

Much of the material in this section forms the basis for laminar theories of the large-scale ocean circulation. Many of the concepts are detailed in Chapter 7 of [Marshall and Plumb \(2008\)](#). Chapters 19-22 of [Vallis \(2017\)](#) as well as [Samelson \(2011\)](#) present ocean circulation theory making use of fundamental concepts of geophysical fluid dynamics. A compelling discussion of the cascade of energy from the baroclinic modes to barotropic mode is offered by [Smith and Vallis \(2001\)](#). [Gill \(1982\)](#) provides a discussion of the depth of no motion problem in dynamic oceanography.

Rotating tank laboratory experiments offer a powerful means to explore the variety of rotating fluid mechanics relevant to the atmosphere and oceans. The following resources exemplify the Taylor-Proudman effect (26.33) and the associated vertical stiffening of rotating fluids.

- One means to test Taylor-Proudman is to insert a dye into a rapidly rotating tank of unstratified water. After a few rotation periods the dye forms vertical sheets known as *Taylor*

curtains whose center is along the rotation axis. The fluid is said to have a *vertical stiffness* due to the effects of rotation. Vertical stiffening in turn means that flow over a small obstacle is deflected throughout the column rather than just near the bump. This situation is depicted in Figure 26.5 and more vividly illustrated in [this video from the UCLA SpinLab](#).

- Near the 20 minute mark of [this video, also from UCLA](#), we see how vortices in a rotating fluid maintain their coherency much more than in a non-rotating fluid.
- Another laboratory test shown in [this video from Prof. Fultz](#) shows that a buoyant object (a ping pong ball) placed into a rotating tank rises much slower than in a non-rotating tank. The reason for the slower rise is that the ball must push up against the vertically stiffened fluid column when rotating, thus slowing its rise relative to when in a non-rotating fluid. Later in the same video, near the 16 minute mark, we see Taylor curtains in rotating fluids.

26.6 Inertial waves

The vertical stiffness exhibited by the Taylor-Proudman effect (Section 26.5.3) has a somewhat magical appearance. As with any apparent action at a distance in classical physics, there is a local causal mechanism to be found in waves. *Inertial waves* provide the mechanism for mediating vertical stiffening. We here introduce the basics of inertial waves in a homogeneous and constantly rotating fluid without boundaries. We are somewhat descriptive since a more deductive presentation requires a bit more wave theory than offered in this book.

26.6.1 Inertial oscillations of a fluid ring

In Section 11.5 we studied the free motion of a point particle viewed from a rotating reference frame, whereby the particle exhibited anti-cyclonic *inertial oscillations* with period $2\pi/f$. Here we extend that study to the case of a fluid. To start the investigation, consider a coherent circular fluid ring within the horizontal plane. Let the plane be rotating about the vertical axis, $\Omega = \Omega \hat{z}$, that extends through the center of the ring (Figure 26.8). Now perturb the ring by expanding its radius outward, which then gives the ring a larger moment of inertia as computed relative to the rotational axis. As we saw when studying axial angular momentum in Sections 11.7, 11.8, and 18.4, the constraint imposed by angular momentum computed relative to the \hat{z} axis requires the ring to rotate anti-cyclonically (clockwise if $\Omega > 0$). As it rotates, each fluid particle within the ring experiences a Coriolis acceleration pointing towards the center of the ring (to the right of the particle motion). This Coriolis acceleration halts the outward perturbation and returns the ring towards its smaller radius, with the inward motion leading to a further Coriolis acceleration that causes the ring to rotate cyclonically (again, to the right of the inward particle motion). The whole process oscillates between radially outward and anti-cyclonic rotation, versus radially inward and cyclonic rotation. These inertial oscillations of a fluid ring provide the basic mechanism for inertial waves.

26.6.2 Formulation of the inertial wave equation

Consider a homogeneous inviscid fluid on a rotating plane governed by the momentum equation

$$[\partial_t + (\mathbf{v} \cdot \nabla)] \mathbf{v} + 2\boldsymbol{\Omega} \wedge \mathbf{v} = -\nabla p/\rho - \rho g \hat{z}, \quad (26.52)$$

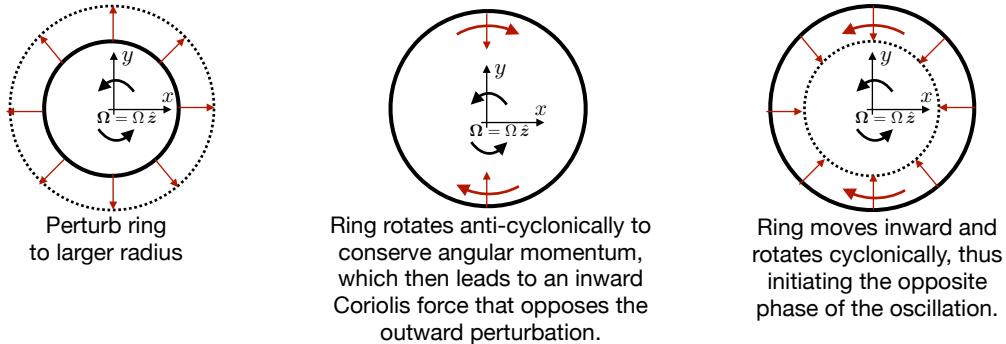


FIGURE 26.8: Schematic of inertial oscillations of a ring of fluid in the horizontal plane in the presence of rotation, $\Omega = \Omega \hat{z}$ with $\Omega > 0$, where the rotation axis extends through the center of the ring. The left panel shows the ring perturbed outward, with this perturbation increasing the ring's moment of inertia about the vertical axis running through the center of the ring. To conserve angular momentum the ring must turn opposite to the sense of the rotating reference frame; that is, it rotates anti-cyclonically, as shown in the middle panel. As it turns anti-cyclonically the ring generates a Coriolis acceleration that points inward (to the right of the motion), thus causing the ring to oscillate back to a smaller radius (right panel), where the oscillation turns around.

where g is the effective gravitational acceleration that arises from central gravity and planetary centrifugal (Section 10.12.4). For the most part, we have in mind the f -plane case with $\Omega = \Omega \hat{z} = (f/2) \hat{z}$. However, allowing for an arbitrary orientation of Ω assists in geometric understanding.

As when formulating the Boussinesq approximation in Section 24.1.1, it is useful to decompose pressure according to

$$p = \rho \varphi + p_0 \quad \text{where} \quad dp_0/dz = -\rho g, \quad (26.53)$$

with ρ the constant density and $p_0(z)$ the static background hydrostatic pressure that exactly balances the fluid weight. This decomposition brings the momentum equation 26.52 to the form

$$[\partial_t + (\mathbf{v} \cdot \nabla)] \mathbf{v} + 2\Omega \wedge \mathbf{v} = -\nabla \varphi. \quad (26.54)$$

Focusing attention on small amplitude fluctuations motivates us to linearize the momentum equation (26.54) by dropping the self-advection term, $(\mathbf{v} \cdot \nabla) \mathbf{v}$, thus leading to

$$\partial_t \mathbf{v} + 2\Omega \wedge \mathbf{v} = -\nabla \varphi. \quad (26.55)$$

With Ω independent of time, we see that the velocity projected onto Ω satisfies

$$\partial_t (\mathbf{v} \cdot \Omega) = -\Omega \cdot \nabla \varphi. \quad (26.56)$$

Hence, if there is no pressure gradient along the direction of the rotation axis, then the velocity of the linear flow in that direction remains constant in time. For example, if the rotation is around the vertical axis, $\Omega = \Omega \hat{z}$, and there is no vertical dynamic pressure gradient, $\partial_z \varphi = 0$, then the vertical velocity is static, $\partial_t w = 0$.

Taking the curl of the linear velocity equation (26.55) eliminates the pressure gradient and yields the linearized vorticity equation⁵

$$\partial_t \boldsymbol{\omega} = 2(\Omega \cdot \nabla) \mathbf{v} \quad \text{with} \quad \boldsymbol{\omega} = \nabla \wedge \mathbf{v}. \quad (26.57)$$

To reach this result requires the identity (2.37h) for the curl of a cross product. We also assumed Ω has no spatial dependence so that all of its derivatives vanish. To develop a wave equation, we take

⁵We study vorticity in Part VI of the book. For now, we only need to know that vorticity is the curl of the velocity.

a time derivative of the linear vorticity equation (26.57), and make use of the linearized momentum equation (26.55), thus leading to

$$\partial_t \boldsymbol{\omega} = -2(\boldsymbol{\Omega} \cdot \nabla)(2\boldsymbol{\Omega} \wedge \mathbf{v} + \nabla \varphi). \quad (26.58)$$

To eliminate the pressure gradient we take another curl and again make use of the identity (2.37h) to write

$$\nabla \wedge \boldsymbol{\omega} = \nabla \wedge (\nabla \wedge \mathbf{v}) = \nabla(\nabla \cdot \mathbf{v}) - \nabla^2 \mathbf{v} = -\nabla^2 \mathbf{v}, \quad (26.59)$$

where $\nabla \cdot \mathbf{v} = 0$ since the fluid density is constant. We also make use of Cartesian tensor calculus from Chapter 2 to derive the identity

$$(\nabla \wedge [(\boldsymbol{\Omega} \cdot \nabla)(\boldsymbol{\Omega} \wedge \mathbf{v})])_m = \epsilon_{mnp} \partial_n [\Omega_s \partial_s (\boldsymbol{\Omega} \wedge \mathbf{v})_p] \quad (26.60a)$$

$$= \epsilon_{mnp} \partial_n [\Omega_s \partial_s \epsilon_{pqr} \Omega_q v_r] \quad (26.60b)$$

$$= \epsilon_{mnp} \epsilon_{qrp} \Omega_s \Omega_q \partial_n \partial_s v_r \quad (26.60c)$$

$$= (\delta_{mq} \delta_{nr} - \delta_{mr} \delta_{nq}) \Omega_s \Omega_q \partial_n \partial_s v_r \quad (26.60d)$$

$$= \Omega_s (\Omega_m \partial_r \partial_s v_r - \Omega_n \partial_n \partial_s v_m) \quad (26.60e)$$

$$= -(\boldsymbol{\Omega} \cdot \nabla)^2 v_m, \quad (26.60f)$$

where we made use of the ϵ -tensor identity (1.38) as well as $\nabla \cdot \mathbf{v} = 0$. We are thus led to the wave equation for inertial waves

$$[\partial_{tt} \nabla^2 + (2\boldsymbol{\Omega} \cdot \nabla)^2] \mathbf{v} = 0. \quad (26.61)$$

26.6.3 Dispersion relation for inertial waves

The inertial wave equation (26.61) provides the starting point for developing the kinematics of inertial waves. Since we are interested in wave fluctuations, we consider a linear wave solution of the form

$$\mathbf{v} = \tilde{\mathbf{v}} \exp[i(\mathbf{k} \cdot \mathbf{x} - \varpi t)] \quad (26.62a)$$

$$\varphi = \tilde{\varphi} \exp[i(\mathbf{k} \cdot \mathbf{x} - \varpi t)], \quad (26.62b)$$

with

$$\mathbf{k} = \hat{\mathbf{x}} k_x + \hat{\mathbf{y}} k_y + \hat{\mathbf{z}} k_z \quad (26.63)$$

the wave vector, $\tilde{\mathbf{v}}$ and $\tilde{\varphi}$ the (generally complex) wave amplitudes for the velocity and pressure, and ϖ the radial frequency.⁶ As we are considering free space waves (no boundaries), there is no preferred length scale for inertial waves. Furthermore, since the flow is non-divergent, the constraint $\nabla \cdot \mathbf{v} = 0$ means that the velocity of fluid particles is perpendicular to the wavevector

$$\mathbf{v} \cdot \mathbf{k} = 0. \quad (26.64)$$

This orientation of wavevector and fluid velocity characterizes *transverse waves*, in which lines of constant wave phase (e.g., wave crests and troughs) are perpendicular to \mathbf{k} .

Plugging the wave ansatz (26.62a) into the inertial wave equation (26.61) leads to the *dispersion relation*

$$\varpi^2 = (2\boldsymbol{\Omega} \cdot \mathbf{k})^2 / |\mathbf{k}|^2. \quad (26.65)$$

⁶We use ϖ for the radial frequency rather than the more common ω in order to distinguish radial frequency from the vorticity, $\boldsymbol{\omega}$.

The radial frequency of inertial waves is thus directly proportional to the orientation of the wave vector relative to the rotation vector. Furthermore, it is independent of the magnitude of the wave vector. To more clearly exhibit these properties we find it useful to write

$$\boldsymbol{\Omega} \cdot \mathbf{k} = |\boldsymbol{\Omega}| |\mathbf{k}| \cos \Lambda, \quad (26.66)$$

where Λ is the angle between $\boldsymbol{\Omega}$ and \mathbf{k} . The dispersion relation (26.65) thus takes on the particularly compact form

$$\varpi = \pm 2 |\boldsymbol{\Omega}| \cos \Lambda \quad \text{with} \quad -\pi \leq \Lambda \leq \pi, \quad (26.67)$$

which serves to explicitly remove the wavevector magnitude. Sweeping through the possible orientation angles, Λ , reveals that the radial frequency for free space inertial waves spans the continuum range

$$-2 |\boldsymbol{\Omega}| \leq \varpi \leq 2 |\boldsymbol{\Omega}|. \quad (26.68)$$

26.6.4 Group velocity

The group velocity determines the speed and direction of wave energy propagation and it is given by

$$\mathbf{c}_g = \nabla_{\mathbf{k}} \varpi = \hat{\mathbf{x}} \frac{\partial \varpi}{\partial k_x} + \hat{\mathbf{y}} \frac{\partial \varpi}{\partial k_y} + \hat{\mathbf{z}} \frac{\partial \varpi}{\partial k_z}. \quad (26.69)$$

Making use of the dispersion relation (26.65) renders

$$\mathbf{c}_g = \frac{\pm \mathbf{k} \wedge (2 \boldsymbol{\Omega} \wedge \mathbf{k})}{|\mathbf{k}|^3}, \quad (26.70)$$

where we made use of the vector identity (1.40b). This expression leads to a particularly striking feature of the group velocity for inertial waves

$$\mathbf{c}_g \cdot \mathbf{k} = 0. \quad (26.71)$$

We are familiar with the Coriolis force acting perpendicular to the direction of a moving fluid particle. That orientation manifests for inertial waves via $\mathbf{k} \cdot \mathbf{c}_g = 0$ so that inertial waves carry energy (via the group velocity) in a direction parallel to wave crests (perpendicular to \mathbf{k}). A second property of the group velocity is found by projecting it onto the direction of the rotational axis

$$\mathbf{c}_g \cdot \boldsymbol{\Omega} = \pm 2 [\Omega^2 \mathbf{k} \cdot \mathbf{k} - (\boldsymbol{\Omega} \cdot \mathbf{k})^2] / |\mathbf{k}|^3 = \pm 2 [\Omega^2 - (\varpi/2)^2] / |\mathbf{k}| = \pm 2 (\Omega^2 / |\mathbf{k}|) \sin^2 \Lambda, \quad (26.72)$$

where we made use of the expressions (26.65) and (26.67) for the dispersion relation.

26.6.5 Relating the velocity components

The velocity amplitude, $\tilde{\mathbf{v}}$, is generally a complex number, which allows for there to be a variety of phase shifts between the velocity components. We here determine some general relations between these amplitudes by returning to the linear velocity equation (26.55) and inserting the wave ansatz (26.62a) and (26.62b) to render

$$-\mathrm{i} \varpi \tilde{u} - f \tilde{v} = -\mathrm{i} k_x \tilde{\varphi} \quad (26.73a)$$

$$-\mathrm{i} \varpi \tilde{v} + f \tilde{u} = -\mathrm{i} k_y \tilde{\varphi} \quad (26.73b)$$

$$-\mathrm{i} \varpi \tilde{w} = -\mathrm{i} k_z \tilde{\varphi}. \quad (26.73c)$$

For simplicity we assumed the rotation vector is aligned with the vertical direction so that $2\Omega = f \hat{z}$. Manipulations with the horizontal velocity equations lead to the relations between the horizontal velocity amplitudes

$$\tilde{u}(f^2 - \varpi^2) = -\tilde{\varphi}(k_x \varpi + ik_y f) \quad (26.74a)$$

$$\tilde{v}(f^2 - \varpi^2) = \tilde{\varphi}(-k_y \varpi + ik_x f) \quad (26.74b)$$

$$\frac{\tilde{u}}{\tilde{v}} = \frac{-k_x k_y (f^2 - \varpi^2) + i \varpi f (k_x^2 + k_y^2)}{(k_y \varpi)^2 + (k_x f)^2}. \quad (26.74c)$$

We make use of these relations in the following.

26.6.6 “Near” inertial waves with $\varpi \approx \pm|f|$

The inertial wave frequency is at its maximum magnitude when the wavevector is aligned parallel or anti-parallel to the rotation axis,

$$\mathbf{k} \wedge \boldsymbol{\Omega} = 0 \implies \varpi = \pm 2|\Omega| = \pm|f|. \quad (26.75)$$

Recall that inertial waves are transverse so that $\mathbf{v} \cdot \mathbf{k} = 0$. This property, in combination with $\mathbf{k} \wedge \boldsymbol{\Omega} = 0$, means that high frequency inertial waves have fluid particle motion in a plane perpendicular to the rotation axis: $\mathbf{v} \cdot \boldsymbol{\Omega} = 0$. For example, if the rotation vector is vertical, then the wavevector only has a vertical component, $\mathbf{k} = k_z \hat{z}$. In oceanography, inertial waves with $\varpi \approx \pm|f|$ are referred to as *near inertial waves*, which refers to their radial frequency being close to the Coriolis frequency. Since near inertial waves have their wavevector oriented close to the rotation axis, equation (26.70) indicates that they have a vanishingly small group velocity.

For inertial waves with $\varpi = f$ and $\mathbf{k} = k_z \hat{z}$, the fluid velocity amplitude relation (26.74c) indicates that $\tilde{u}/\tilde{v} = i$ so that the horizontal velocity of the wave is given by

$$\mathbf{u}/\tilde{u} = \hat{x} e^{i(k_z z - f t)} + \hat{y} e^{i(k_z z - f t - \pi/2)}. \quad (26.76)$$

Taking the real part renders

$$\mathbf{u}/\tilde{u} = \hat{x} \cos(k_z z - f t) + \hat{y} \sin(k_z z - f t). \quad (26.77)$$

With $k_x = k_y = 0$ in the velocity equations (26.73a) and (26.73b), we see that there is no coupling between the horizontal velocity components and pressure. Hence, the motion of fluid particles reduces to inertial oscillations in the horizontal plane just as we studied for point particle motion in Section 11.5. As a check, we see that for a fixed vertical position, say $z = 0$, the velocity relation (26.77) for near inertial waves is identical the velocity relation (11.15b) for particles undergoing inertial oscillations

$$\mathbf{u}/\tilde{u} = \hat{x} \cos(f t) - \hat{y} \sin(f t) \quad \text{for } z = 0. \quad (26.78)$$

26.6.7 Near zero frequency inertial waves

Low frequency inertial waves occur when the wavevector is nearly perpendicular to the rotation axis

$$\boldsymbol{\Omega} \cdot \mathbf{k} \approx 0 \implies \varpi/\Omega \approx 0. \quad (26.79)$$

Hence, the wave number parallel to the rotation axis is vanishingly small. For example, if the rotation axis is vertical, then low frequency inertial waves have a vanishingly small vertical wave number,

$$k_z^2 \ll k_x^2 + k_y^2. \quad (26.80)$$

Correspondingly, for the velocity vector in the form (26.62a), $\boldsymbol{\Omega} \cdot \mathbf{k} \approx 0$ means that

$$(\boldsymbol{\Omega} \cdot \nabla) \mathbf{v} = i(\boldsymbol{\Omega} \cdot \mathbf{k}) \mathbf{v} \approx 0. \quad (26.81)$$

That is, the velocity vector for low frequency inertial waves is coherent in the direction aligned with the rotation axis. Furthermore, when $\boldsymbol{\Omega} \cdot \mathbf{k} \approx 0$, the group velocity (26.72) has a magnitude

$$|\mathbf{c}_g| \approx 2|\boldsymbol{\Omega}|/|\mathbf{k}|, \quad (26.82)$$

so that low frequency inertial waves with long wavelength quickly transmit their energy along the rotation axis direction.

26.6.8 Connecting to Taylor-Proudman

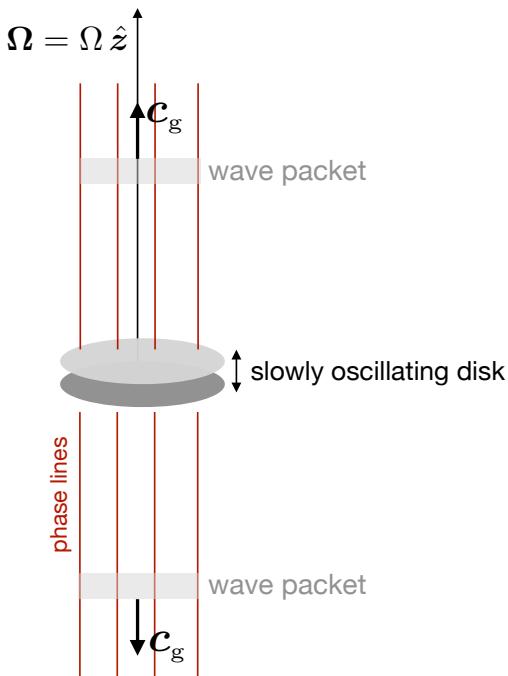


FIGURE 26.9: Schematic of inertial waves generated by a slowly oscillating disk in a rotating homogeneous fluid. The disk moves along the rotation axis in small amplitude oscillations whose radial frequency is much smaller than the rotational frequency, $|\boldsymbol{\Omega}|$. The oscillations generate low frequency inertial waves whose phase lines and group velocity are parallel to the rotation axis, and whose wavevector is perpendicular to the rotation axis. We depict two wave packets that send energy vertically away from the disk, with the long wave and low frequency waves having the highest magnitude for the group velocity. Since $\boldsymbol{\Omega} \cdot \mathbf{k} = 0$, the fluid particle velocity associated with the inertial waves is constant in the direction along the rotation axis: $(\boldsymbol{\Omega} \cdot \nabla) \mathbf{v} = 0$.

We bring the previous notions together to describe a realization of low frequency inertial waves transmitting information about vertical stiffening, with reference to Figure 26.9. For this purpose, imagine a slowly oscillating disk that moves up and down in a direction aligned with the rotation axis. If the oscillation frequency is much slower than the rotation frequency, then the disk generates low frequency inertial waves with frequency of the oscillating disk. The low frequency inertial waves have a wavevector oriented perpendicular to the rotation axis, $\boldsymbol{\Omega} \cdot \mathbf{k} = 0$. Since inertial waves have a group velocity that is itself perpendicular to the wavevector, the low frequency inertial waves have a group velocity parallel to the rotation axis: $\boldsymbol{\Omega} \wedge \mathbf{c}_g = 0$. Information about the oscillating disk

moves most rapidly transmitted by long wavelength low frequency inertial waves. Furthermore, the low frequency inertial waves generate fluid particle motion that is independent of the position along the rotational axis since $(\boldsymbol{\Omega} \cdot \nabla) \mathbf{v} = 0$. Hence, the particle motion occurs in the plane perpendicular to the rotational axis, and this motion is coherent in the direction along the rotation axis. That is, inertial waves transmit information about stiffening along the rotation axis.

26.6.9 Concerning other waves depending on Coriolis

As we saw in this section, inertial waves rely on the Coriolis force for their existence. There are other waves that share this reliance, such as inertia-gravity waves, Kelvin waves, and Rossby waves. Inertial waves are the most parsimonious since they rely only on the Coriolis force from a constantly rotating homogeneous fluid without boundaries. In contrast, inertia-gravity waves also require continuous stratification or a stratification jump at an interface (as in the shallow water inertia-gravity waves in Chapter 32 or the surface gravity waves in Section 30.1). Likewise, Kelvin waves require a boundary either from a solid wall or an effective boundary due to the sign change of the Coriolis parameter at the equator (Exercise 32.13). And Rossby waves require a spatial gradient in the rotation rate as per the planetary beta effect (Section 34.7).

For much of the ocean and atmosphere, the buoyancy frequency, N (Section 25.5.4), is larger than $|f|$, in which case the radial frequency of inertia-gravity waves is bounded by $|f| \leq \varpi \leq N$, such as the shallow water inertia-gravity waves discussed in Section 32.8. These waves are said to be *super-inertial* since their frequency is larger than $|f|$. However, when stratification is reduced below $|f|$, then the inertia-gravity wave frequency becomes sub-inertial, $N \leq \varpi \leq |f|$. In the limit that $N \rightarrow 0$, inertia-gravity waves reduce to the inertial waves discussed in this section.

26.6.10 Further study

Our discussion of inertial waves was inspired by Section 9.2 of [Davidson \(2015\)](#), who emphasizes the role of inertial waves in forming vertically stiff structures in turbulent rapidly rotating flows. A further pedagogical analysis of inertial waves is provided by [Mory \(1992\)](#). For a visualization of inertial waves, refer to the 18 minute mark from the [rotating tank experiments of Prof. Fultz](#). He illustrates inertial oscillations within a bounded rotating homogeneous fluid, where the presence of boundaries quantizes the frequency of the excited inertial waves.

26.7 Isopycnal form stress from geostrophic eddies

As introduced in Section 20.1, form stress is the horizontal stress arising from pressure acting on a sloped surface. The mathematical expression for the form stress acting on the top side of a surface is given by equation (20.6)

$$\boldsymbol{\Sigma}^{\text{form}} = p \nabla \eta, \quad (26.83)$$

with the opposite sign for the form stress on the bottom side of the surface. Here, $z = \eta(x, y, t)$ is the depth of the surface (see Figure 20.3 or Figure 26.10). The net horizontal force from form stress is the area integral over the surface.

In this section we examine the zonal mean zonal form stress acting on an isopycnal surface (Section 26.7.1) and on an isopycnal layer (Section 26.7.2), each for an adiabatic, Boussinesq, hydrostatic fluid in geostrophic balance and within a zonally periodic channel of length L . As we show, the zonal mean zonal form stress arising from geostrophically balanced fluctuations provide an eastward acceleration to the fluid. At the same time, these *geostrophic eddies* transport buoyancy and thickness/volume meridionally. Although the channel geometry is rather simple, it has

applications to the middle latitude atmospheric circulation as well as for ocean circulation, particularly in the Southern Ocean where there is circumpolar channel-like flow within the Antarctic Circumpolar Current. Furthermore, the discussion exposes key elements of eddy-mean flow interactions, sharing points with the leading order generalized Lagrangian mean of Section 49.2 and the quasi-Stokes transport discussed in Section 50.2.

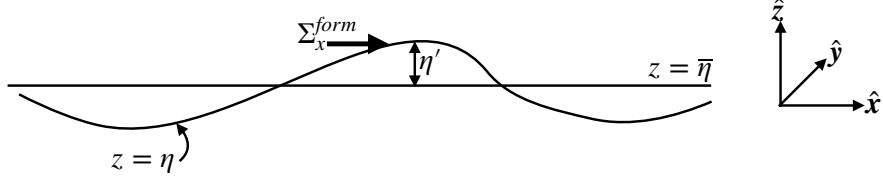


FIGURE 26.10: Schematic of the zonal form stress, Σ_x^{form} , acting on a surface whose zonal mean vertical position is $z = \bar{\eta}(y, t)$ and whose vertical position relative to the zonal mean is $z = \bar{\eta}(y, t) + \eta'(x, y, t)$.

26.7.1 Zonal mean zonal form stress on an isopycnal surface

We are here interested in the form stress acting on an isopycnal surface. Before specializing to an isopycnal, we decompose the form stress according to the zonal mean depth and its deviation from zonal mean (see Figure 26.10). Thereafter, specialization to an isopycnal surface in an adiabatic fluid connects the zonal mean form stress to the meridional eddy flux of buoyancy.

Zonal form stress on an arbitrary surface in a channel

The zonal mean vertical position of the surface is written

$$\bar{\eta} = \frac{1}{L} \int_0^L \eta \, dx \quad (26.84)$$

and its corresponding zonal fluctuation is

$$\eta' = \eta - \bar{\eta}. \quad (26.85)$$

The zonal component of the form stress is thus given by

$$p \partial_x \eta = p(x, \bar{\eta} + \eta') \partial_x (\bar{\eta} + \eta') \quad (26.86a)$$

$$= p(x, \bar{\eta} + \eta') \partial_x \eta' \quad (26.86b)$$

$$\approx [p(x, \bar{\eta}) + \partial_z p(x, \bar{\eta}) \eta'] \partial_x \eta' \quad (26.86c)$$

$$= p(x, \bar{\eta}) \partial_x \eta' + \mathcal{O}(\eta')^2. \quad (26.86d)$$

Hence, to second order in fluctuations, η' , the zonal form stress acting on the surface equals to $p(x, \bar{\eta}) \partial_x \eta'$, where it is important to note that pressure is evaluated at the zonal mean depth, $z = \bar{\eta}$.

To within the same accuracy, the zonal integral of the zonal form stress is given by

$$\int_0^L \Sigma_x^{\text{form}} \, dx \approx \int_0^L p(\bar{\eta}) (\partial \eta' / \partial x) \, dx = - \int_0^L \eta' [\partial p(\bar{\eta}) / \partial x] \, dx, \quad (26.87)$$

where the final equality follows from zonal periodicity. Now assume the zonal pressure gradient at $\bar{\eta}$ is balanced by a meridional geostrophic velocity

$$\partial p(\bar{\eta}) / \partial x = f \rho_0 v(\bar{\eta}), \quad (26.88)$$

where

$$v(\bar{\eta}) = \bar{v}(\bar{\eta}) + v'(\bar{\eta}) \quad (26.89)$$

so that

$$\int_0^L \Sigma_x^{\text{form}} dx = -\rho_0 f \int_0^L \eta' v' dx, \quad (26.90)$$

where we noted that the Coriolis parameter is independent of zonal position. Hence, there is a nonzero zonal mean zonal form stress when there is a nonzero zonal correlation between fluctuations in the meridional velocity and the depth of the surface

$$\bar{\Sigma}_x^{\text{form}} = -\rho_0 f \bar{v}' \bar{\eta}'. \quad (26.91)$$

Zonal mean zonal form stress acting on an isopycnal surface

To further unpack the correlation appearing in equation (26.91), specialize to the case of an isopycnal surface in an adiabatic fluid. As shown in our discussion of generalized Lagrangian mean averaging in Sections 49.2.6 and 49.4.7, vertical fluctuations in the position of the isopycnal surfaces, relative to the zonal mean $\bar{\eta}$, are related to zonal fluctuations in the density

$$\eta' \approx -\frac{\rho'}{\partial \bar{\rho}/\partial z} = -\frac{b'}{N^2}, \quad (26.92)$$

where we introduced the squared buoyancy frequency of the zonal mean state as well as the fluctuating buoyancy

$$N^2 = -\frac{g}{\rho_0} \frac{\partial \bar{\rho}}{\partial z} \quad \text{and} \quad b' = -\frac{g \rho'}{\rho_0}. \quad (26.93)$$

The zonally averaged zonal form stress thus takes the form

$$\bar{\Sigma}_x^{\text{form}} = \frac{\rho_0 f}{N^2} \bar{v}' \bar{b}'. \quad (26.94)$$

Again, the assumptions rendering the result (26.94) are (i) zonal periodicity, (ii) adiabatic and Boussinesq fluid, (iii) geostrophically balanced flow. Under these assumptions, the zonal mean zonal form stress acting on an isopycnal surface is proportional to the zonal correlation between fluctuations in the meridional velocity and the buoyancy. It is a general statistical property of geostrophic eddies in the atmosphere and ocean to transport positive buoyancy (e.g., warm air/water) poleward and negative buoyancy (e.g., cold air/water) equatorward, thus ameliorating the equator-to-pole buoyancy difference setup by solar radiation that preferentially warms the tropics. In turn, this property of geostrophic eddies leads to a positive zonal mean zonal form stress

$$\bar{\Sigma}_x^{\text{form}} > 0. \quad (26.95)$$

Hence, in addition to transporting buoyancy poleward, geostrophic eddies provide a positive zonal mean force through zonal integrated form stress that accelerates the fluid in the eastward direction. These two properties of geostrophic eddies (poleward flux of positive buoyancy anomalies along with an eastward acceleration from form stress) are fundamental to the middle latitude atmospheric circulation as well as for ocean circulation, particularly within the channel-like Antarctic Circumpolar Current.

26.7.2 Zonal mean zonal form stress acting on an isopycnal layer

We offer yet another means to understand the zonal mean zonal form stress by here examining the form stress acting on a constant density layer of adiabatic Boussinesq fluid such as shown in Figure 26.11. This layered/isopycnal analysis anticipates some of the development considered for the stacked shallow water model in Chapters 31 and 32 as well as for isopycnal models in Section 46.1.

The net form stress acting on the upper and lower layer interfaces in Figure 26.11 is given by

$$\Sigma^{\text{layer form}} = p_1 \nabla \eta_1 - p_2 \nabla \eta_2 \quad (26.96\text{a})$$

$$= p(\eta + h/2) \nabla(\eta + h/2) - p(\eta - h/2) \nabla(\eta - h/2) \quad (26.96\text{b})$$

$$\approx [p(\eta) - \rho g h/2] \nabla(\eta + h/2) - [p(\eta) + \rho g h/2] \nabla(\eta - h/2) \quad (26.96\text{c})$$

$$= p \nabla h - \rho g h \nabla \eta \quad (26.96\text{d})$$

$$= \nabla(p h) - h \nabla(p + \rho g \eta) \quad (26.96\text{e})$$

$$= \nabla(p h) - \rho_0 h \nabla M. \quad (26.96\text{f})$$

In this relation we set $z = \eta$ for the vertical position at the center of the layer, introduced the Montgomery potential from Section 46.1.1

$$M \rho_0 = p + \rho g \eta, \quad (26.97)$$

and noted that ρ is a uniform constant layer density so that it commutes with the horizontal gradient operator computed along ρ surfaces. We also made use of the hydrostatic balance to approximate the interface pressures as

$$p(\eta + h/2) \approx p(\eta) + \frac{\partial p}{\partial z} \frac{h}{2} = p(\eta) - \rho g h/2 \quad (26.98\text{a})$$

$$p(\eta - h/2) \approx p(\eta) - \frac{\partial p}{\partial z} \frac{h}{2} = p(\eta) + \rho g h/2. \quad (26.98\text{b})$$

The zonal mean of the zonal layer form stress is thus given by the correlation between the layer thickness fluctuations and fluctuations in the zonal derivative of the Montgomery potential

$$\bar{\Sigma}_x^{\text{layer form}} = -\rho_0 \overline{h' \partial M'/\partial x}, \quad (26.99)$$

where we set $\overline{\partial M/\partial x} = 0$ due to zonal periodicity. As seen in Section 46.1.1, the Montgomery potential is the geostrophic streamfunction in isopycnal coordinates, so that the fluctuating meridional geostrophic velocity is given by

$$f v' = \partial_x M'. \quad (26.100)$$

Consequently, the zonal mean zonal form stress acting on the layer equals to the correlation between the thickness fluctuations and fluctuations in the meridional geostrophic velocity

$$\bar{\Sigma}_x^{\text{layer form}} = -\rho_0 f \overline{v' h'}. \quad (26.101)$$

Hence, as the geostrophic eddies provide a net eastward acceleration to the layer (equation (26.95)), they also move volume around within isopycnal layers meridionally, moving positive thickness fluctuations equatorward.

To further understand the physics of the form stress in equation (26.101), parameterize the velocity-thickness eddy correlation, $\overline{v' h'}$, by downgradient diffusion of thickness

$$\overline{v' h'} = -\kappa \partial_y \bar{h}, \quad (26.102)$$

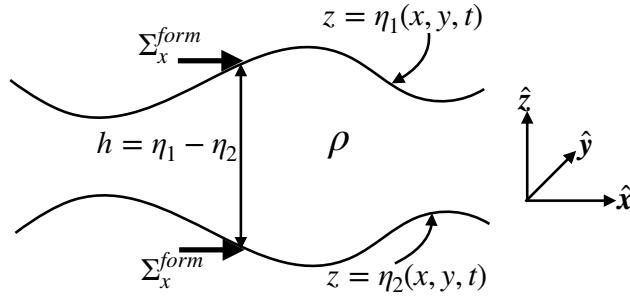


FIGURE 26.11: Schematic of a constant density layer of an adiabatic, hydrostatic, Boussinesq fluid with thickness $h(x, y, t) = \eta_1(x, y, t) - \eta_2(x, y, t) = (\eta + h/2) - (\eta - h/2)$, and uniform density $\rho = \text{constant}$. East points to the right and north is oriented into the page. The zonal form stress, Σ_x^{form} , acting on the upper and lower interfaces at a horizontal position (x, y) are shown by the thick horizontal vectors. The zonal form stress is the horizontal component of the compressive pressure force per area acting on the layer interfaces, with the sign of the form stress determined by the slope of the layer interface. For a zonally periodic fluid layer, the net zonal pressure force acting on the layer arises from the zonal form stress integrated over the layer interfaces.

where $\kappa > 0$ is a nonzero kinematic diffusivity (dimensions of squared length per time). This parameterization is suggested by the work of [Gent and McWilliams \(1990\)](#) as discussed in Section 50.2.5. As noted there, thickness diffusion as a parameterization reflects the general tendency of geostrophic eddies to reduce horizontal gradients in layer thickness as they reduce the available potential energy of the flow. In this case the zonal mean zonal form stress is

$$\bar{\Sigma}_x^{\text{layer form}} = \rho_0 f \kappa \partial_y \bar{h}. \quad (26.103)$$

So in the northern hemisphere in regions where the zonal mean layer thickness increases to the north, $\partial_y \bar{h} > 0$, there is a corresponding eastward zonal mean zonal form stress arising from parameterized geostrophic eddies acting on layer thickness. This situation corresponds to the case in Section 26.7.1, where we saw that geostrophic eddies preferentially transport positive buoyancy anomalies poleward and negative buoyancy anomalies equatorward. In the present analysis, meridional changes to the layer thickness correspond to a nonzero thermal wind shear. If layer thickness increases poleward, as for the case of weaker vertical stratification in the high latitudes, then the zonal velocity has a positive vertical shear, thus contributing an eastward zonal mean form stress.

26.7.3 Comments and further study

A key feature of geostrophic eddies exposed by this discussion concerns the connection between zonal form stress (providing an eastward force on the zonally periodic channel flow) and meridional eddy transport of buoyancy (positive buoyancy anomalies are transported poleward) and thickness (positive thickness anomalies are transported equatorward). The periodic channel domain is highly idealized. Nonetheless, the basic ideas form the roots for much of how we think about geostrophic eddies in the middle latitude atmosphere and the Southern Ocean. Further generalizations lead to the generalized Lagrangian mean, whose kinematic rudiments are discussed in Section 49.2, as well as the thickness weighted average, discussed in Chapter 47.

The fundamental role of form stress in geostrophic turbulent flows is pedagogically treated by [Vallis \(2017\)](#). See, in particular, his Chapter 21 for a thorough and insightful discussion of circulation in the Southern Ocean. We also return to form stress within the shallow water fluid in Section 32.4. That discussion complements the presentation given here, with a focus on a layer of shallow water fluid. We also touch on the notions of form stress when discussing the [Gent and McWilliams \(1990\)](#) mesoscale eddy parameterization in Section 50.2.

26.8 Exercises

EXERCISE 26.1: SMALL ROSSBY NUMBER AT HUMAN SCALES

Consider motion of a car at a speed $U \sim 10^5$ m hour⁻¹ and a length scale of $L \sim 10$ m.

- (a) What is the rotation period required render a unit Rossby number for the given scales? Give result in units of seconds.
- (b) If the earth rotated at the angular speed Ω_{human} , what would be the solid-body speed for a point at rest on the earth's surface? Give result in units of meter per second.
- (c) How does the solid-body speed compare to the speed of sound at standard atmospheric conditions? What about the root-mean-square speed for air molecules? Hint: read Section 12.3.4.
- (d) Discuss one or two astronomical objects that have very large rotational speeds.

EXERCISE 26.2: THE BETA SPIRAL

Consider a steady state Boussinesq planetary geostrophic fluid in the absence of mixing. Write the geostrophic velocity as

$$u = |\mathbf{u}| \cos \Delta \quad v = |\mathbf{u}| \sin \Delta, \quad (26.104)$$

where Δ is the angle measured counter-clockwise from east. Use thermal wind and the steady state perfect fluid buoyancy equation to determine an expression for $\partial \Delta / \partial z$. Show that for $f > 0$ (northern hemisphere) and $\partial b / \partial z = N^2 > 0$ (gravitationally stable fluid column), then $\partial \Delta / \partial z$ has opposite sign from the vertical velocity, w . This spiralling of the geostrophic velocity is known as the *beta spiral* in oceanography.

EXERCISE 26.3: ALTERNATIVE FORM OF THERMAL WIND

Consider a fluid with density a function of pressure and potential temperature

$$\rho = \rho(p, \theta). \quad (26.105)$$

A physical realization of this equation of state is a lake. Show that the thermal wind shear for a hydrostatic and *compressible* fluid with this equation of state can be written in the form

$$\frac{\partial \mathbf{u}}{\partial z} = \left[\frac{N^2}{f \rho g} \right] (\hat{\mathbf{z}} \wedge \nabla_\theta p), \quad (26.106)$$

where

$$N^2 = -\frac{g}{\rho} \frac{\partial \rho}{\partial \theta} \frac{\partial \theta}{\partial z} = g \beta_\theta \frac{\partial \theta}{\partial z} > 0 \quad (26.107)$$

is the squared buoyancy frequency, assumed positive so that the fluid is gravitationally stable in the vertical (see Section 25.5.4). The term β_θ is the thermal expansion coefficient written in terms of potential temperature (Section 25.3.4). Finally, the horizontal gradient projected onto constant θ surfaces is given by (see Section 43.11.2)

$$\nabla_\theta = \hat{\mathbf{x}} \left[\frac{\partial}{\partial x} \right]_{y,\theta} + \hat{\mathbf{y}} \left[\frac{\partial}{\partial y} \right]_{x,\theta} \quad (26.108a)$$

$$= \nabla_z - \left[\frac{\nabla_z \theta}{\partial \theta / \partial z} \right] \frac{\partial}{\partial z}. \quad (26.108b)$$

Hint: This exercise requires careful use of the chain rule and the hydrostatic relation, along with the equations given in the problem statement. Furthermore, assume the fluid is fully compressible.

Hint: Some may wish to “warm-up” by showing that the result holds for the simpler equation of state $\rho = \rho(\theta)$. Some of the steps used for the simpler case are relevant for the case with $\rho = \rho(\theta, p)$.



Horizontal flow balances

We here consider a variety of inviscid horizontal (tangent plane) flow regimes characterized by a balance between a subset of terms appearing in the horizontal momentum equation. This discussion allows us to directly compare the geostrophically balanced flow of Chapter 26 to a variety of other balanced flows such as gradient wind, inertial motion, and cyclostrophic balance. We provide a categorization of the flow following *natural coordinates*, where natural coordinates offer a concise means to compare the relative magnitudes of the Coriolis, pressure, and centrifugal accelerations acting on a fluid element moving horizontally.

READER'S GUIDE TO THIS CHAPTER

We make use of the Boussinesq momentum equation from Section 24.1 and assume an understanding of the geostrophic balance from Chapter 26. We also use of some geometric notions discussed in Chapter 5, though most of the salient points are revisited here so that Chapter 5 is an option rather than a requirement. Throughout this chapter we assume a tangent plane geometry and associated equations from Section 23.2, thus allowing for Cartesian coordinates. Also, we ignore all vertical motion. Some of this material is used in subsequent chapters, in particular Chapter 28 on Ekman dynamics as well as Chapters 40 and 42 on quasi-geostrophy. Much of this material can also be found in Section 3.2 of Holton (1992) and Section 2.9 of Vallis (2017).

27.1	Loose ends	606
27.2	Horizontal flow described by natural coordinates	606
27.2.1	Natural coordinates	606
27.2.2	Material acceleration	607
27.2.3	Centripetal and centrifugal accelerations	608
27.2.4	Coriolis and pressure gradient	608
27.2.5	Horizontal momentum equation and local Rossby number	609
27.2.6	Eulerian decomposition of the acceleration	610
27.2.7	Further study	611
27.3	Exact geostrophic balance	611
27.3.1	Steady f -plane flow	612
27.3.2	Steady β -plane flow	612
27.3.3	What about geostrophic balance with curved motion?	613
27.4	Inertial motion	613
27.4.1	Anti-cyclonic circular motion on f -plane	614
27.4.2	Period for inertial motion	614
27.4.3	Observing inertial motion	615

27.4.4	Inertial motion is Lagrangian	615
27.4.5	“Inertial” motion does not refer to an inertial frame	615
27.4.6	Inertial oscillations on a sphere	615
27.5	Cyclostrophic balance	616
27.6	Gradient wind balance	617
27.6.1	Constraints on gradient wind flow	618
27.6.2	The variety of gradient wind flows	619
27.6.3	Comments	620

27.1 Loose ends

- How to determine whether the velocity field is divergent or non-divergent for the cyclostrophic and gradient wind motions?
- Exercises needed for this chapter.

27.2 Horizontal flow described by natural coordinates

In this section we decompose the horizontal Boussinesq momentum equation into motion parallel to and motion perpendicular to the instantaneous trajectory of a fluid element. That is, we characterize the velocity and acceleration according to the local flow direction. Furthermore, we are only concerned with horizontal motion (on a constant geopotential). Using the *natural coordinates* arising from this description exposes the centripetal/centrifugal acceleration that arises from curvature in the trajectory as measured by the radius of curvature. This non-inertial acceleration is distinct from the centrifugal acceleration that arises from planetary rotation, with planetary centrifugal acceleration already contained within the effective gravitational acceleration that acts in the local vertical direction (see Section 10.10.1). We also decompose the accelerations from pressure, friction, and Coriolis into their natural coordinate components.

27.2.1 Natural coordinates

Natural coordinates for horizontal motion are defined by a locally orthogonal set of unit vectors (see Figure 27.1)

$$\hat{z} = \hat{u} \wedge \hat{n} = \text{vertical direction} \quad (27.1a)$$

$$\hat{u} = \hat{n} \wedge \hat{z} = \text{tangent to horizontal velocity} \quad (27.1b)$$

$$\hat{n} = \hat{z} \wedge \hat{u} = \text{normal direction to the left of motion.} \quad (27.1c)$$

The unit vector \hat{u} is tangent to the velocity vector (which is horizontal), so that

$$\mathbf{u} = |\mathbf{u}| \hat{u} = \frac{Ds}{Dt} \hat{u}, \quad (27.2)$$

where s is the arc-length measured along the trajectory as introduced in Section 2.4. The unit vector \hat{n} is perpendicular to the velocity and points to the left of the trajectory facing downstream.

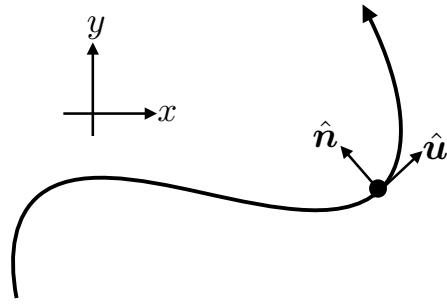


FIGURE 27.1: Illustrating the decomposition of horizontal motion of a fluid element into natural coordinate directions. These directions are defined by a unit tangent vector, \hat{u} , pointing in the direction of the fluid element motion, and a unit normal vector, \hat{n} , pointing to the left of the motion facing downstream.

27.2.2 Material acceleration

When writing the velocity according to equation (27.2), we decompose the acceleration into the change in speed and change in direction

$$\frac{D\mathbf{u}}{Dt} = \frac{D|\mathbf{u}|}{Dt} \hat{\mathbf{u}} + |\mathbf{u}| \frac{D\hat{\mathbf{u}}}{Dt}. \quad (27.3)$$

Following our discussion of rotation in Section 10.4 (see Figure 10.4), the magnitude of the direction change can be written in terms of the infinitesimal angle swept out by the motion as the fluid element moves along a trajectory

$$|\delta\hat{\mathbf{u}}| = |\delta\vartheta|. \quad (27.4)$$

The infinitesimal angle swept out by the trajectory is related to the radius of curvature, R (Figure 27.2), and the arc-length increment, δs , traversed by the trajectory

$$\delta\vartheta = \frac{\delta s}{R}. \quad (27.5)$$

Finally, the infinitesimal change in tangent, $\delta\hat{\mathbf{u}}$, is directed normal to the motion, which we see by noting

$$\hat{\mathbf{u}} \cdot \hat{\mathbf{u}} = 1 \implies \delta\hat{\mathbf{u}} \cdot \hat{\mathbf{u}} = 0. \quad (27.6)$$

That is, $\delta\hat{\mathbf{u}}$ is orthogonal to $\hat{\mathbf{u}}$, so that it points either parallel or anti-parallel to $\hat{\mathbf{n}}$. We detailed this property of rotating unit vectors in Section 2.1.4 (see Figure 2.2). Our convention is that $\hat{\mathbf{n}}$ points to the left of $\hat{\mathbf{u}}$, so that if the trajectory turns to the left, then $\delta\hat{\mathbf{u}}$ points parallel to $\hat{\mathbf{n}}$, whereas if the trajectory turns to the right then $\delta\hat{\mathbf{u}}$ points anti-parallel to $\hat{\mathbf{n}}$. That is, $\delta\hat{\mathbf{u}}$ always points towards the center of the circle used to compute the radius of curvature as in Figure 27.2.

Bringing these results together leads to the expression for the infinitesimal unit vector change

$$\delta\hat{\mathbf{u}} = \hat{\mathbf{n}} \frac{\delta s}{R}. \quad (27.7)$$

Our sign convention takes $R > 0$ for a fluid element turning in the direction of $\hat{\mathbf{n}}$ (to the left facing downstream) and $R < 0$ when turning opposite to $\hat{\mathbf{n}}$ (to the right facing downstream). Hence, the material time change is

$$\frac{D\hat{\mathbf{u}}}{Dt} = \frac{D\hat{\mathbf{u}}}{Ds} \frac{Ds}{Dt} = \frac{\hat{\mathbf{n}}}{R} \frac{Ds}{Dt} = \frac{\hat{\mathbf{n}}}{R} |\mathbf{u}|, \quad (27.8)$$

where the speed is given by the time change of the arc-length along the trajectory

$$|\mathbf{u}| = \frac{Ds}{Dt}. \quad (27.9)$$

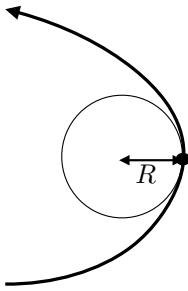


FIGURE 27.2: Illustrating the radius of curvature associated with turning motion of a fluid element. The radius of curvature equals to the radius of a tangent circle (the *curvature circle*) that approximates, to second order accuracy, the trajectory at a particular point. The radius is smaller in magnitude when the trajectory is highly curved, and $|R| = \infty$ when the trajectory is straight. The radius is positive when the trajectory turns into the normal direction as depicted here (to the left; concave as defined by $\hat{\mathbf{n}}$) and negative when turning in the opposite direction (to the right; convex as defined by $\hat{\mathbf{n}}$). See Section 5.2 for more details on curvature, with Figure 5.4 offering more details for how to determine the radius of curvature.

Combining these results renders the acceleration

$$\frac{D\mathbf{u}}{Dt} = \frac{D|\mathbf{u}|}{Dt} \hat{\mathbf{u}} + \frac{|\mathbf{u}|^2}{R} \hat{\mathbf{n}}. \quad (27.10)$$

The acceleration has thus been decomposed into the change in speed of the fluid element along the direction of the motion, plus the centripetal acceleration due to curvature of the trajectory. In Section 27.2.3 we justify referring to $(|\mathbf{u}|^2/R) \hat{\mathbf{n}}$ as the *centripetal* acceleration.

27.2.3 Centripetal and centrifugal accelerations

The centripetal acceleration points towards the concave side of a turning trajectory; “centripetal” means “towards the center.” Its opposing partner, the centrifugal (“away from center”) acceleration points to the convex side (see Figure 27.3). So how do we interpret $\hat{\mathbf{n}} |\mathbf{u}|^2/R$? For motion turning to the left, towards $\hat{\mathbf{n}}$, the radius of curvature is positive, $R > 0$, so that $\hat{\mathbf{n}} |\mathbf{u}|^2/R$ points to the concave side of the trajectory (left side). For a trajectory turning to the right then $R < 0$, which again means that $\hat{\mathbf{n}} |\mathbf{u}|^2/R$ points to the concave side (now on the right). We conclude that the acceleration $\hat{\mathbf{n}} |\mathbf{u}|^2/R$ indeed represents a centripetal acceleration and $-\hat{\mathbf{n}} |\mathbf{u}|^2/R$ is the centrifugal acceleration.

27.2.4 Coriolis and pressure gradient

The Coriolis acceleration takes the following form in natural coordinates

$$-f \hat{\mathbf{z}} \wedge \mathbf{u} = -(\hat{\mathbf{z}} \wedge \hat{\mathbf{u}}) f |\mathbf{u}| = -\hat{\mathbf{n}} f |\mathbf{u}|, \quad (27.11)$$

so that the Coriolis acceleration always points to the right of the flow direction. In contrast, the pressure gradient has two components

$$\nabla p = \hat{\mathbf{u}} (\hat{\mathbf{u}} \cdot \nabla p) + \hat{\mathbf{n}} (\hat{\mathbf{n}} \cdot \nabla p) = \hat{\mathbf{u}} \frac{\partial p}{\partial s} + \hat{\mathbf{n}} \frac{\partial p}{\partial n}, \quad (27.12)$$

one pointing along the flow direction and one normal to the direction.

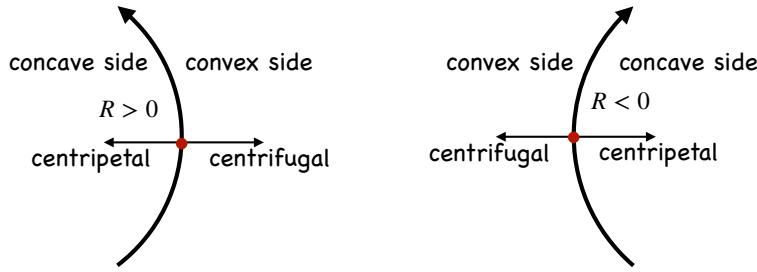


FIGURE 27.3: Centripetal acceleration of a turning fluid element, $\hat{\mathbf{n}} |\mathbf{u}|^2 / R$, points to the concave side of the curve (towards the center) whereas the centrifugal acceleration, $-\hat{\mathbf{n}} |\mathbf{u}|^2 / R$, points to the convex side (away from the center). The centripetal and centrifugal accelerations are paired action/reaction accelerations. The normal unit vector, $\hat{\mathbf{n}}$, always points to the left of the motion, whereas the radius of curvature, R , is positive or negative depending on the direction of the turning motion. For a left turning trajectory (in direction of $\hat{\mathbf{n}}$), the concave side is on the left and has positive radius of curvature, $R > 0$, whereas for the right turning trajectory (opposite direction of $\hat{\mathbf{n}}$) the concave side is to the right with $R < 0$. To help remember the signs, note that centrifugal means “away from the center” whereas centripetal means “towards the center”. It is the centrifugal acceleration that pulls one away from the center of a merry-go-round whereas one’s arms and hands provide the balancing centripetal acceleration.

27.2.5 Horizontal momentum equation and local Rossby number

Bringing the above results together leads to the horizontal momentum equation decomposed into natural coordinates

$$\frac{D|\mathbf{u}|}{Dt} = -\frac{1}{\rho_0} \frac{\partial p}{\partial s} + \mathbf{F} \cdot \hat{\mathbf{u}} \quad \text{motion parallel to } \hat{\mathbf{u}} \quad (27.13a)$$

$$\frac{|\mathbf{u}|^2}{R} + f |\mathbf{u}| = -\frac{1}{\rho_0} \frac{\partial p}{\partial n} + \mathbf{F} \cdot \hat{\mathbf{n}} \quad \text{motion perpendicular to } \hat{\mathbf{u}}, \quad (27.13b)$$

where \mathbf{F} is the frictional acceleration and ρ_0 is the reference density for the Boussinesq fluid. These equations decompose the accelerations into those acting parallel to and normal to the trajectory. It is notable that the normal component is purely diagnostic; there is no time derivative in equation (27.13b). Instead, it is a balance between centrifugal, Coriolis, normal pressure gradient, and normal component of friction. In the next few sections we consider certain limiting cases as revealed by the equations of motion (27.13a) and (27.13b). Friction remains zero in this chapter but is nonzero for the discussion of Ekman mechanics in Chapter 28.

Steady frictionless flow

The frictionless balanced motions considered in this chapter all occur with a fixed velocity magnitude for the fluid element. Hence, the along-trajectory component of the momentum equation (27.13a) for frictionless motion reduces to

$$\frac{D|\mathbf{u}|}{Dt} = -\frac{1}{\rho_0} \frac{\partial p}{\partial s} = 0. \quad (27.14)$$

Hence, there is no pressure gradient aligned with the fluid element motion, and correspondingly the frictionless fluid motion preserves its kinetic energy

$$\frac{1}{2} \frac{D(\mathbf{u} \cdot \mathbf{u})}{Dt} = |\mathbf{u}| \frac{D|\mathbf{u}|}{Dt} = 0. \quad (27.15)$$

Local Rossby number

As discussed in Section 26.3, the Rossby number is the ratio of the velocity advection term to the Coriolis term. In the normal component to the momentum equation (27.13b), we have the advection term manifest as the local centrifugal term. Following Chapter 1 of [van Heijst \(2010\)](#) we define the *local Rossby number* as the ratio of the centrifugal acceleration to the Coriolis acceleration

$$\text{Ro}_{\text{local}} = \frac{|\mathbf{u}|^2/R}{f/|\mathbf{u}|} = \frac{|\mathbf{u}|}{Rf}. \quad (27.16)$$

In this chapter we consider flow regimes as determined by values of the local Rossby number.

27.2.6 Eulerian decomposition of the acceleration

The horizontal equations of motion (27.13a) and (27.13b) offer a relatively simple and insightful Lagrangian description of the motion. In contrast, an Eulerian description requires us to expand the material time derivative. Although we focus in this chapter on the Lagrangian perspective, we here consider two decompositions available when using the Eulerian form.

Advective form

The standard form of the material time derivative is given by

$$\frac{\partial \mathbf{u}}{\partial t} + (\mathbf{u} \cdot \nabla) \mathbf{u} = \frac{\text{D}|\mathbf{u}|}{\text{Dt}} \hat{\mathbf{u}} + \frac{|\mathbf{u}|^2}{R} \hat{\mathbf{n}}, \quad (27.17)$$

so that

$$\hat{\mathbf{u}} \cdot \left[\frac{\partial \mathbf{u}}{\partial t} + (\mathbf{u} \cdot \nabla) \mathbf{u} \right] = \frac{\text{D}|\mathbf{u}|}{\text{Dt}} = -\frac{1}{\rho_0} \frac{\partial p}{\partial s} + \mathbf{F} \cdot \hat{\mathbf{u}} \quad (27.18a)$$

$$\hat{\mathbf{n}} \cdot \left[\frac{\partial \mathbf{u}}{\partial t} + (\mathbf{u} \cdot \nabla) \mathbf{u} \right] = \frac{|\mathbf{u}|^2}{R} = -f |\mathbf{u}| - \frac{1}{\rho_0} \frac{\partial p}{\partial n} + \mathbf{F} \cdot \hat{\mathbf{n}}. \quad (27.18b)$$

Depending on the information provided by a field measurement or numerical simulation, one might more readily diagnose the kinematic expressions on the left side of these equations or the force balances on the right side. For example, the inverse radius of curvature for the fluid flow can be diagnosed in either of the following ways

$$R^{-1} = \frac{\hat{\mathbf{n}} \cdot [\partial \mathbf{u} / \partial t + (\mathbf{u} \cdot \nabla) \mathbf{u}]}{|\mathbf{u}|^2} = \frac{-f |\mathbf{u}| - \rho_0^{-1} \partial p / \partial n + \mathbf{F} \cdot \hat{\mathbf{n}}}{|\mathbf{u}|^2}. \quad (27.19)$$

Vector invariant form

We can transform the self-advection term using the vector identity (see Section 2.3.4)

$$(\mathbf{u} \cdot \nabla) \mathbf{u} = \nabla K - \mathbf{u} \wedge \zeta \hat{\mathbf{z}}, \quad (27.20)$$

where $K = \mathbf{u} \cdot \mathbf{u}/2$ is the kinetic energy per mass in the horizontal flow, and $\zeta = \partial_x v - \partial_y u$ is the vertical component to the relative vorticity (see Chapter 33). The deflecting acceleration, $\mathbf{u} \wedge \zeta \hat{\mathbf{z}}$,

is known as the *Magnus acceleration* and was discussed in Section 18.2.3. Projecting into the $\hat{\mathbf{u}}$ and $\hat{\mathbf{n}}$ directions leads to

$$\hat{\mathbf{u}} \cdot \frac{\partial \mathbf{u}}{\partial t} = -\frac{\partial(p/\rho_0 + K)}{\partial s} + \mathbf{F} \cdot \hat{\mathbf{u}} \quad (27.21a)$$

$$\hat{\mathbf{n}} \cdot \frac{\partial \mathbf{u}}{\partial t} = -(f + \zeta) |\mathbf{u}| - \frac{\partial(p/\rho_0 + K)}{\partial n} + \mathbf{F} \cdot \hat{\mathbf{n}}. \quad (27.21b)$$

The gradient acceleration is affected by both pressure and the *dynamic pressure* afforded by the kinetic energy per mass. The Magnus acceleration appears only in the normal component equation, which is expected since it acts just as the Coriolis acceleration to deflect the fluid element perpendicular to its trajectory. Following the example in Figure 18.1, consider a positive relative vorticity, $\zeta > 0$. For this case, the Magnus acceleration deflects the fluid element to the right of its trajectory; i.e., opposite to the normal direction, which points to the left as per our convention in Figure 27.1.

27.2.7 Further study

Section 3.2 of [Holton \(1992\)](#) details the use of natural coordinates for geophysical flows, with a similar decomposition provided in Section 7.10 of [Gill \(1982\)](#) and Section 2.9 of [Vallis \(2017\)](#). Natural coordinates are also used in describing non-rotating flows as illustrated in [this video](#).

27.3 Exact geostrophic balance

Frictionless flow parallel to pressure contours experiences no pressure gradient ($\partial p / \partial s = 0$), so that the speed of a fluid element remains constant. Furthermore, if this motion occurs with an infinite radius of curvature (straight line motion parallel to pressure contours), then the force balance is between the normal pressure gradient and Coriolis. In this case the local Rossby number (27.16) vanishes

$$\text{Ro}_{\text{local}} = \frac{|\mathbf{u}|}{R f} = 0 \quad \text{if } |R| = \infty. \quad (27.22)$$

More precisely, exact geostrophic balance occurs under the following conditions:

- fluid element move on a straight line so that the radius of curvature is infinite, $|R| = \infty$, thus making the centripetal acceleration and local Rossby number both vanish;
- fluid elements move along lines of constant pressure so that $\partial p / \partial s = 0$;
- friction vanishes.

In this case the equations of motion (27.13a) and (27.13b) take the form

$$\frac{D|\mathbf{u}|}{Dt} = 0 \quad (27.23a)$$

$$f |\mathbf{u}| = -\frac{1}{\rho_0} \frac{\partial p}{\partial n}. \quad (27.23b)$$

Equation (27.23a) says that the speed of a fluid element is constant, so that the horizontal kinetic energy likewise is constant. Equation (27.23b) says that the pressure gradient normal to the motion balances the Coriolis acceleration. We refer to this flow, depicted in Figure 27.4, as *exact geostrophic balance* since it is an exact solution under the above assumptions.

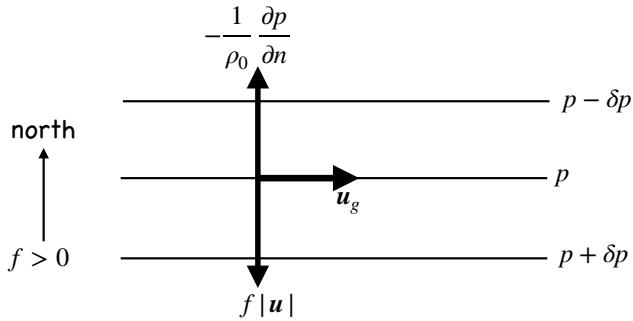


FIGURE 27.4: Exact geostrophic balance on a tangent plane occurs when the flow is horizontal, frictionless, straight, and follows contours of constant pressure. For this case the pressure gradient exactly balances the Coriolis acceleration so that the motion is perpendicular to both of these accelerations. We here depict motion assuming $f > 0$ as for the northern hemisphere. If flow is on an f -plane then the exact geostrophic balance is steady for any arbitrary flow direction. On a β -plane, steady exact geostrophic balance holds only for zonal flow.

Writing the horizontal advection of speed in the form

$$\mathbf{u} \cdot \nabla |\mathbf{u}| = |\mathbf{u}| \hat{\mathbf{u}} \cdot \nabla |\mathbf{u}| = |\mathbf{u}| \frac{\partial |\mathbf{u}|}{\partial s}, \quad (27.24)$$

allows us to write the material constancy of the flow speed as

$$\frac{\partial |\mathbf{u}|}{\partial t} + |\mathbf{u}| \frac{\partial |\mathbf{u}|}{\partial s} = 0. \quad (27.25)$$

Hence, a steady flow speed, with $\partial |\mathbf{u}| / \partial t = 0$, only holds for the exact geostrophic balance if the flow speed is fixed along each trajectory path

$$\frac{\partial |\mathbf{u}|}{\partial s} = 0 \implies \frac{\partial |\mathbf{u}|}{\partial t} = 0. \quad (27.26)$$

What is required for this condition to hold? We examine two cases, again restricted to a tangent plane.

27.3.1 Steady f -plane flow

Geostrophic motion on an f -plane is horizontally non-divergent (Section 26.4)

$$\nabla \cdot \mathbf{u} = \nabla \cdot (\hat{\mathbf{u}} |\mathbf{u}|) = 0. \quad (27.27)$$

Flow in a straight line, with each trajectory parallel to one another, has the trajectory direction independent of space. Hence, the non-divergent condition means that

$$\nabla \cdot (\hat{\mathbf{u}} |\mathbf{u}|) = (\hat{\mathbf{u}} \cdot \nabla) |\mathbf{u}| = \frac{\partial |\mathbf{u}|}{\partial s} = 0, \quad (27.28)$$

which then proves that exact geostrophic flow on an f -plane is steady.

27.3.2 Steady β -plane flow

The geostrophic velocity in the presence of a meridional gradient of the Coriolis parameter, $f = f(y)$, satisfies (Section 26.4)

$$\nabla \cdot (f \mathbf{u}) = 0. \quad (27.29)$$

Making use of $\nabla \cdot \hat{\mathbf{u}} = 0$ for straight line trajectories leads to

$$\nabla \cdot (f \mathbf{u}) = \frac{\partial(f |\mathbf{u}|)}{\partial s} = 0. \quad (27.30)$$

We conclude that $\partial|\mathbf{u}|/\partial s = 0$ holds only for trajectories that are parallel to latitude lines, in which case $\partial f/\partial s = \partial f/\partial x = 0$. Therefore, exact geostrophic motion on the β -plane is steady only for trajectories that follow constant latitude lines; i.e., zonal trajectories as depicted in Figure 27.4.

27.3.3 What about geostrophic balance with curved motion?

The geostrophically balanced flows discussed in Chapter 26 generally have curvature; i.e., the flows are not straight, such as for the geostrophic motion around a pressure center as shown in Figure 26.3. But as emphasized by the natural coordinate decomposition as per equations (27.13a) and (27.13b), curved motion has an associated centrifugal acceleration. So when speaking of geostrophic balance for flow that has a nonzero curvature, then the local Rossby number (27.16) is not precisely zero. Rather, its magnitude is small but nonzero

$$|\text{Ro}_{\text{local}}| \ll 1 \quad \text{approximate geostrophic flow.} \quad (27.31)$$

In this limit it is accurate to ignore the centrifugal acceleration, which is commonly the case for large-scale flows. Even so, it is an approximation, with the centrifugal acceleration identically zero only for straight line motion on a plane.

27.4 Inertial motion

Inertial motion occurs under the following conditions:

- vanishing pressure gradient
- vanishing friction,

so that the equations of motion (27.13a) and (27.13b) take the form

$$\frac{D|\mathbf{u}|}{Dt} = 0 \quad (27.32a)$$

$$\frac{|\mathbf{u}|^2}{R} + f |\mathbf{u}| = 0. \quad (27.32b)$$

Equation (27.32a) says that inertial motion occurs with constant speed, whereas equation (27.32b) says that the motion maintains the balance between Coriolis and centrifugal accelerations

$$f |\mathbf{u}| = -\frac{|\mathbf{u}|^2}{R}. \quad (27.33)$$

Hence, local Rossby number has a unit magnitude

$$|\text{Ro}_{\text{local}}| = \frac{|\mathbf{u}|}{|R| |f|} = 1. \quad (27.34)$$

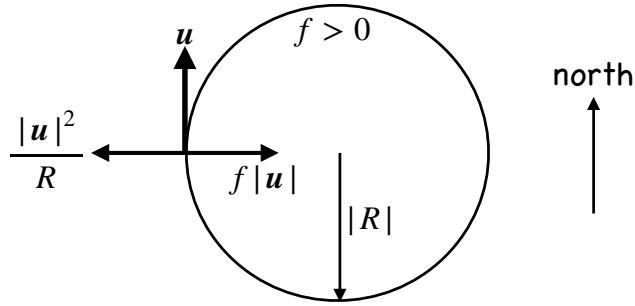


FIGURE 27.5: Inertial motion of a fluid on a plane occurs when the flow is horizontal, frictionless, and the centrifugal acceleration balances the Coriolis in the presence of zero pressure gradient. We here depict motion assuming $f > 0$ as for the northern hemisphere, revealing that inertial motion is an anti-cyclonic circular motion with radius $|R| = |\mathbf{u}|/|f|$. When turning to the right as in the northern hemisphere, the radius of curvature for the inertial circle is negative (see Figure 27.2 for sign convention), so that $R = -|\mathbf{u}|/f < 0$.

27.4.1 Anti-cyclonic circular motion on f -plane

To further understand inertial motion, rearrange equation (27.33) so that

$$f = -\frac{|\mathbf{u}|}{R}, \quad (27.35)$$

in which case the radius for the inertial circle is

$$R = -|\mathbf{u}|/f. \quad (27.36)$$

Equation (27.35) can be satisfied in the northern hemisphere ($f > 0$) only for motion turning to the right (in which $R < 0$). The opposite orientation occurs in the southern hemisphere, where inertial motion turns to the left so that the radius of curvature is positive, $R = -|\mathbf{u}|/f > 0$ (see Figure 27.2 for the sign convention on the radius of curvature). Hence, inertial motion is oriented anti-cyclonically (orientated opposite to the earth's rotation). If the Coriolis parameter is constant, then the motion is circular, as depicted in Figure 27.5.

To emphasize the balance, return to equation (27.33) and recall that the Coriolis acceleration in the northern hemisphere points to the right when facing downstream, as per equation (27.11). Hence, the balance (27.33) is between the Coriolis acceleration pointing to the right and the centrifugal acceleration pointing to the left. That is, the Coriolis acceleration provides the centripetal acceleration to balance the centrifugal acceleration.

27.4.2 Period for inertial motion

Equation (27.36) says that the speed of a fluid element is given by the radius of curvature times the magnitude of the Coriolis parameter

$$|\mathbf{u}| = R|f|. \quad (27.37)$$

The time for a fluid element to traverse an inertial circle is given by the circumference of the circle, $2\pi R$, divided by the constant speed, thus yielding the inertial period

$$T_{\text{inertial}} = \frac{2\pi R}{|\mathbf{u}|} = \frac{2\pi}{|f|}. \quad (27.38)$$

We encountered this inertial period in Section 11.5 when considering inertial oscillations for a point particle.

27.4.3 Observing inertial motion

Inertial motion is rarely observed in the atmosphere since fluid motion nearly always occurs in the presence of a pressure gradient. In contrast, surface ocean flow is commonly generated by wind stresses that setup motion even in the absence of pressure gradients. The moving fluid then engenders a Coriolis acceleration, in which case there can be a balance between centrifugal and Coriolis for the moving ocean fluid. As a result, the observed surface ocean currents have nontrivial power within the inertial frequency band, rivaling energy contained in frequencies associated with astronomical tides (e.g., see Figure 3.3 of [Holton \(1992\)](#)).

How large is an inertial circle? Consider a surface ocean current speed of $|u| \sim 0.1 \text{ m s}^{-1}$, which is not atypical of current speeds outside of strong boundary currents or mesoscale eddies, and assume the Coriolis parameter $f = 10^{-4} \text{ s}^{-1}$. In this case the inertial radius is

$$R_{\text{inertial}} \approx 10^3 \text{ m.} \quad (27.39)$$

Observations of inertial motion, such as that reproduced in Figure 8.3 of [Gill \(1982\)](#), confirm that the radii are indeed on the order of a few kilometers.

27.4.4 Inertial motion is Lagrangian

The analysis in this section concerns a fluid element moving without feeling the impacts from pressure forces. The fluid thus exhibits the same force balance as the point particle discussed in Section 11.5. So although we can measure inertial oscillations at a fixed point in space, the present considerations are Lagrangian in nature, focusing on motion of fluid elements. Furthermore, the inertial period refers to the time it takes for a fluid element to move around the inertial circle at its constant speed. It does not refer to the period of a wave, for example, and yet there are inertial waves (Section 26.6) with this period as well as inertia-gravity waves that have periods close to the inertial period (Section 32.8),

27.4.5 “Inertial” motion does not refer to an inertial frame

We make use of the term “inertial” when referring to inertial motion since both the Coriolis and centrifugal accelerations are nonzero only in the presence of motion; i.e., they require the inertia obtained by a moving massive body. Hence, as noted in Section 11.5.4, “inertial motion” in this context does *not* refer to motion as viewed in the inertial reference frames such as those discussed in Section 10.3.

27.4.6 Inertial oscillations on a sphere

In the analysis thus far, we have assumed an *f*-plane so that inertial oscillations are circular. Without solving the spherical equations for inertial oscillations we can anticipate what happens when such motion occurs on a sphere. As a fluid element moves to higher latitudes the magnitude of the Coriolis parameter increases, thus decreasing the radius of curvature. The opposite happens when moving equatorward. This effect of planetary sphericity leads to an egg-shaped pattern that does not close but instead drifts to the west. Now consider inertial motion that spans the equator. As the fluid crosses the equator, where $f = 0$, the radius of curvature is infinite so that the motion is straight. When moving away from the equator the Coriolis parameter increases, which causes a fluid element to turn and close its path, again with a drift to the west. Motion north of the equator turns to the right whereas motion to the south turns left, so that inertial oscillations that span the equator form a figure-8 path. We illustrate this motion in Figure 27.6.

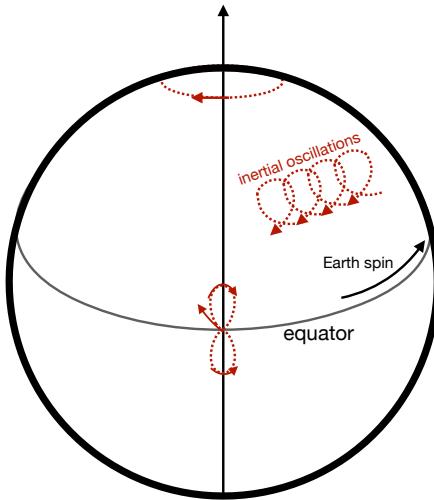


FIGURE 27.6: Depicting inertial oscillations on a sphere. As the fluid element moves poleward, f increases so that the radius of curvature decreases. Correspondingly, the fluid motion does not close, but instead the path drifts to the west. This westward drift holds for both hemispheres, where the sense of the oscillations is anti-cyclonic. The only closed and circular inertial oscillations are those that encircle the pole. Fluid elements that cross the equator exhibit a figure-8 pattern that also drifts to the west. This figure is taken after Figure 4-14 of [von Arx \(1962\)](#).

27.5 Cyclostrophic balance

Cyclostrophic balance occurs under the following conditions:

- fluid element move along lines of constant pressure so that $\partial p / \partial s = 0$;
- vanishing Coriolis acceleration;
- vanishing friction.

The resulting equations of motion (27.13a) and (27.13b) take the form

$$\frac{D|\mathbf{u}|}{Dt} = 0 \quad (27.40a)$$

$$\frac{|\mathbf{u}|^2}{R} = -\frac{1}{\rho_0} \frac{\partial p}{\partial n}. \quad (27.40b)$$

With a vanishing Coriolis acceleration we see that cyclostrophic balance corresponds to local Rossby number that has an infinite magnitude

$$|\text{Ro}_{\text{local}}| = \infty \quad \text{if } f = 0. \quad (27.41)$$

Approximate cyclostrophic balance holds when $|\text{Ro}_{\text{local}}| \gg 1$.

Again, equation (27.40a) says that the speed is constant following a material fluid element. Equation (27.40b) says that cyclostrophic flow occurs when the centrifugal acceleration balances the pressure gradient, with the squared speed given by

$$|\mathbf{u}|^2 = -\frac{R}{\rho_0} \frac{\partial p}{\partial n}. \quad (27.42)$$

This equation can be satisfied for either clockwise or counter-clockwise motion around a low pressure center, as shown in Figure 27.7. For clockwise flow, the radius of curvature is negative, $R < 0$,

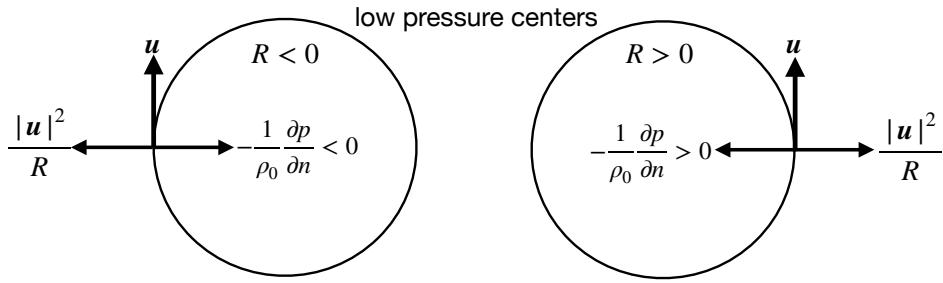


FIGURE 27.7: Cyclostrophic motion of a fluid element on a tangent plane occurs when the flow is horizontal, non-rotating, frictionless, with constant speed, and where the centrifugal acceleration balances the pressure gradient normal to the flow direction. We here depict motion for clockwise and counter-clockwise cyclostrophic flow, both around a low pressure center. Left panel: clockwise motion with radius of curvature, $R < 0$, and the pressure gradient pointing in the direction of the normal, $\partial p / \partial n > 0$. Right panel: counter-clockwise motion with radius of curvature, $R > 0$, and the pressure gradient pointing opposite to the direction of the normal, $\partial p / \partial n < 0$. Cyclostrophic balance does not occur for flow around a high pressure center. The reason is that if both the pressure and centrifugal accelerations point away from the center, then they are unable to balance one another.

whereas $\partial p / \partial n > 0$. The signs are swapped for counter-clockwise flow. Cyclostrophic balance cannot be maintained around a high pressure center. The reason is that if both the pressure and centrifugal accelerations point away from the circle's center, then they are unable to balance one another.

Cyclostrophic balance is relevant for scales on the order of a tornado, with a radius on the order of 300 m where tangential speeds are on the order of 30 m s^{-1} (see Section 3.2.4 of Holton (1992)). For this flow scale, the Rossby number is on the order of 1000 at middle latitudes, thus justifying neglect of the Coriolis acceleration. Although tornadoes in cyclostrophic balance can rotate either clockwise or counter-clockwise, they are more often observed rotating cyclonically given that they are generally embedded within cyclonic storm systems. In contrast, smaller motions such as dust devils and water spouts are quite often seen rotating in either direction.

27.6 Gradient wind balance

Gradient wind balance occurs under the following conditions:

- fluid elements move along lines of constant pressure so that $\partial p / \partial s = 0$;
- vanishing friction.

The resulting equations of motion (27.13a) and (27.13b) take the form

$$\frac{D|\mathbf{u}|}{Dt} = 0 \quad (27.43a)$$

$$\frac{|\mathbf{u}|^2}{R} + f |\mathbf{u}| = -\frac{1}{\rho_0} \frac{\partial p}{\partial n}, \quad (27.43b)$$

Again, equation (27.43a) says that the speed is constant following a material fluid element. Equation (27.43b) says that gradient wind balanced flow occurs when the centrifugal and Coriolis accelerations balance the pressure gradient acting normal to the motion.

The local Rossby number is order unity for the gradient wind balance

$$|\text{Ro}_{\text{local}}| = \frac{|\mathbf{u}|}{|R| |f|} \sim 1, \quad (27.44)$$

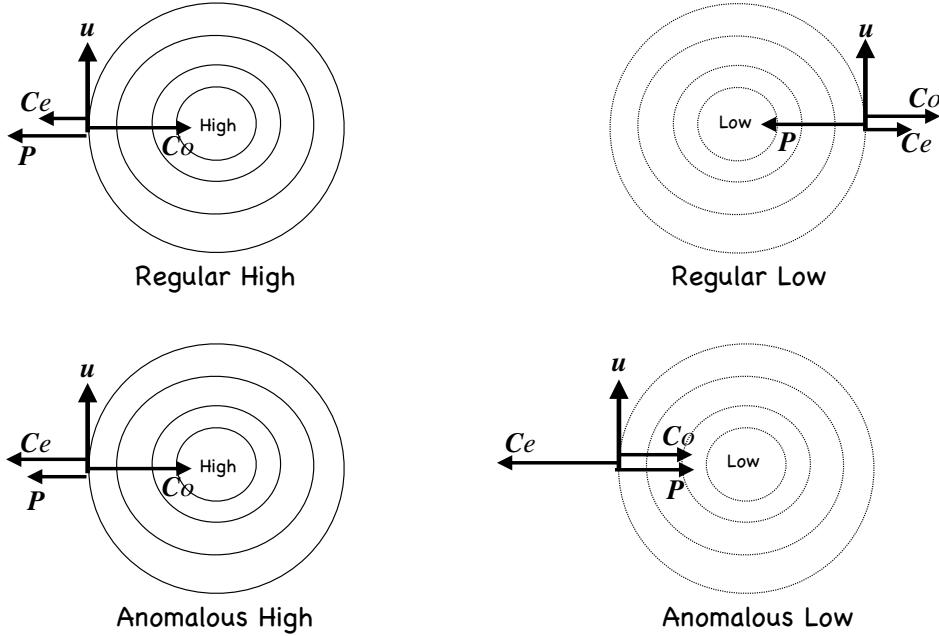


FIGURE 27.8: The variety of gradient wind balances available in the northern hemisphere ($f > 0$). Gradient wind balance occurs when the flow is horizontal, frictionless, with constant speed, and where the centrifugal, pressure, and Coriolis accelerations balance under a variety of magnitudes. To reduce clutter, we use the following shorthand for the accelerations: $P = -\rho_0^{-1} \partial p / \partial n$, $C_o = f |\mathbf{u}|$, and $C_e = |\mathbf{u}|^2 / R$. Upper left panel: motion around a regular high pressure center, whereby the centrifugal acceleration helps the pressure acceleration to balance the Coriolis acceleration. The pressure acceleration is larger in magnitude than the centrifugal. This flow is termed “regular” as it directly corresponds to geostrophic flow around a high pressure center. Lower left panel: motion around an anomalous high pressure center, whereby the centrifugal acceleration helps the pressure acceleration to balance the Coriolis acceleration, with the pressure acceleration smaller in magnitude than the centrifugal. This flow is termed “anomalous” as the pressure acceleration is subdominant to the centrifugal, in contrast to the case of geostrophic flow. Upper right panel: motion around a regular low pressure center, whereby the Coriolis and centrifugal accelerations balance the pressure acceleration. Lower right panel: motion around an anomalous low pressure center, whereby the Coriolis and pressure accelerations balance the centrifugal acceleration. Note the opposite flow orientation between the regular and anomalous lows, whereas the regular and anomalous highs have the same flow orientation.

meaning that both centrifugal and Coriolis accelerations are both important as they balance the pressure gradient. Recall that the inertial motion from Section 27.4 has $|\text{Ro}_{\text{local}}| = 1$, which arises when the pressure gradient vanishes so that the Coriolis and centrifugal terms have equal but opposite magnitudes. The nonzero pressure gradient makes gradient wind flow fundamentally distinct from inertial motion.

27.6.1 Constraints on gradient wind flow

The quadratic formula leads to the following expression for the speed of gradient wind flow

$$|\mathbf{u}| = \frac{R}{2} \left[f \pm \sqrt{f^2 - \frac{4}{R} \frac{1}{\rho_0} \frac{\partial p}{\partial n}} \right]. \quad (27.45)$$

The speed is a real number if the pressure gradient, Coriolis parameter, and radius of curvature satisfy

$$f^2 > \frac{4}{R} \frac{1}{\rho_0} \frac{\partial p}{\partial n} \implies \frac{1}{\rho_0} \frac{\partial p}{\partial n} \leq \frac{R f^2}{4}. \quad (27.46)$$

This relation has direct implications for the structure of the pressure field depending on the sign of the radius of curvature. In particular, as seen in the following, this constraint implies that the pressure gradient at the center of a high pressure region must go to zero as the radius of curvature vanishes, which renders the pressure field relatively flat near the center of highs. In contrast, there is no analogous limit for the magnitude of the pressure gradient approaching a low pressure center. This asymmetry between high and low pressures manifests in atmospheric flow with low pressure centers (cyclonic lows) having stronger magnitude than high pressure centers (anti-cyclonic highs).

27.6.2 The variety of gradient wind flows

We now identify the following force balances available with a gradient wind balance, with illustrations provided in Figure 27.8.

Regular high pressure center (right turn with high pressure on right)

A *regular high pressure* occurs with $R < 0$ and $\partial p/\partial n \leq 0$. This case occurs with the centrifugal and pressure accelerations pointing away from the center, and these balance the Coriolis acceleration pointing to the high pressure center (upper left panel of Figure 27.8).

The inequality (27.46) provides a bound to the size of the pressure gradient since

$$\frac{1}{\rho_0} \frac{\partial p}{\partial n} \leq \frac{R f^2}{4} \text{ with } R \leq 0 \text{ and } \frac{\partial p}{\partial n} \leq 0 \implies \frac{1}{\rho_0} \left| \frac{\partial p}{\partial n} \right|_{\max} = \frac{|R| f^2}{4}. \quad (27.47)$$

That is, the pressure gradient for a regular high cannot be larger than this bound in order for there to be a gradient wind solution. Since $R \rightarrow 0$ as the center is approached, the normal pressure gradient, $\partial p/\partial n$, in turn must vanish towards the center. Note that Holton (1992) identifies two subcases for this balance depending on the relative size of the pressure and centrifugal accelerations, with the anomalous high the case where the pressure gradient acceleration is weaker than the centrifugal (lower left panel of Figure 27.8).

Regular low (left turn with low pressure on left)

This flow occurs with $R \geq 0$ and $\partial p/\partial n \leq 0$, so that the inequality (27.46) is always satisfied

$$\frac{1}{\rho_0} \frac{\partial p}{\partial n} \leq \frac{R f^2}{4} \text{ with } R \geq 0 \text{ and } \frac{\partial p}{\partial n} \leq 0 \implies \text{arbitrary size to } \left| \frac{\partial p}{\partial n} \right|. \quad (27.48)$$

Hence, there is no constraint imposed by gradient wind balance on the size of the pressure gradient magnitude, $|\partial p/\partial n|$. So the low pressure center can be arbitrarily strong and still maintain a gradient wind balance. Furthermore, the Coriolis and centrifugal accelerations point away from the low pressure center, and these two accelerations balance the pressure acceleration that points toward the center (upper right panel of Figure 27.8).

Anomalous low (right turn with low pressure on right)

This flow occurs with $R < 0$ and $\partial p/\partial n \geq 0$. This case occurs with the Coriolis and pressure accelerations pointing toward the low pressure center, and these two accelerations balance the centrifugal acceleration pointing away from the center (lower right panel of Figure 27.8). As with the regular low, the inequality (27.46) provides no bound to the magnitude of the low pressure. Note the opposite orientation for the flow around an anomalous low relative to the regular low.

Left turn with high pressure on left

In this case $R > 0$ and $\partial p / \partial n > 0$. There is no solution for the northern hemisphere since all accelerations point to the right of the motion thus disallowing any balance.

27.6.3 Comments

As noted in Section 3.2 of [Holton \(1992\)](#), the difference between gradient wind speeds and geostrophic wind speeds is no more than 10% to 20% in middle latitude synoptic atmosphere flow. In the tropics, where geostrophy becomes less relevant, it is important to apply the gradient wind relation to capture the balanced flow states. Furthermore, [van Heijst \(2010\)](#) and Chapter 18 of [Cushman-Roisin and Beckers \(2011\)](#) make use of a gradient wind analysis for the study of ocean vortices. The deviations from geostrophy become important when considering relatively small ocean vortices and/or tropical vortices.



Ekman mechanics

Ekman mechanics is concerned with flow affected by accelerations from horizontal pressure gradients, vertical friction, and Coriolis, with particular attention given to regions near boundaries where friction is especially large and *Ekman boundary layers* form. The leading role for Coriolis acceleration causes Ekman boundary layers to exhibit behaviors quite distinct from their non-rotating cousins discussed in Section 19.9.7. In particular, Ekman boundary layer flows are horizontally divergent, thus leading to vertical exchange of mass, tracers, momentum, and vorticity between the boundary layer and the fluid region outside of the boundary layer (i.e., the fluid interior). In so doing, the Ekman layer flow imparts a stretching and squeezing of interior fluid columns that strongly couples the boundary layer to vorticity and circulation of the fluid interior. This role for Ekman layers is especially crucial for the ocean general circulation.

READER'S GUIDE TO THIS CHAPTER

To introduce the subject we exhibit the role of friction in producing a down pressure gradient component to the flow, making use of natural coordinates from Chapter 27 for this purpose. The remainder of the chapter focuses on the mechanics of Ekman boundary layers, with this study greatly extending our understanding of strongly rotating flows from Chapter 26. Indeed, the addition of friction to an otherwise geostrophically balanced flow provides a surprising level of richness that motivates the variety of perspectives presented in this chapter. Cartesian coordinates are sufficient for our purposes throughout this chapter.

Ekman boundary layers are a key element in the study of ocean circulation, particularly the wind-driven circulation. We here borrow liberally from the material in Section 9.2 of [Gill \(1982\)](#), Section 6.2 of [Apel \(1987\)](#), Section 7.4 of [Marshall and Plumb \(2008\)](#), Chapter 8 of [Cushman-Roisin and Beckers \(2011\)](#), and Section 5.7 of [Vallis \(2017\)](#). Since there are many nuances to Ekman theory, the interested reader will find it useful to study more than one presentation.

28.1 The dynamical balances	622
28.2 Horizontal balances in natural coordinates	623
28.2.1 Natural coordinates according to isobars	623
28.2.2 Geostrophic and Ekman balances	624
28.2.3 Rayleigh drag	624
28.2.4 Cross isobar flow driven by Rayleigh drag	625
28.2.5 Horizontal spiral plus vertical rising/sinking	626
28.2.6 Further study	626
28.3 Ekman number and Ekman layer thickness	627
28.3.1 Laplacian vertical friction	627

28.3.2	Non-dimensionalization	627
28.3.3	Defining the Ekman number and layer thickness	628
28.3.4	Estimates for the vertical eddy viscosity	629
28.4	Ocean surface Ekman layer	629
28.4.1	Horizontal mass transport within the Ekman layer	629
28.4.2	Example Ekman mass transports	630
28.4.3	Mass budget for the Ekman layer	632
28.4.4	Ekman layer coupled to the geostrophic interior	633
28.4.5	Further study	635
28.5	Bottom Ekman layer	636
28.5.1	Horizontal mass transport within the Ekman layer	636
28.5.2	Mass budget for the bottom Ekman layer	636
28.5.3	An analytic bottom Ekman spiral velocity	637
28.5.4	Comments	640
28.6	Arrested bottom Ekman flows	640
28.6.1	Description	640
28.6.2	Applications	640
28.7	Exercises	641

28.1 The dynamical balances

Throughout this chapter we make use of a steady, linear, Boussinesq, and hydrostatic fluid maintaining a balance between the horizontal accelerations from Coriolis, pressure, and friction

$$f \hat{z} \wedge \mathbf{u} = -\frac{1}{\rho_o} \nabla_z p + \mathbf{F}. \quad (28.1)$$

This balance is most relevant over large horizontal length scales as per the planetary geostrophic equations of Section 26.5.

The frictional acceleration, \mathbf{F} , of interest in this chapter arises from the vertical exchange of horizontal momentum between fluid layers. Turbulence induced viscous exchange is especially large in boundary regions such as the ocean surface and the ocean and atmosphere bottom. In such turbulent boundary layer regions we make use of the eddy viscosity, which is much larger than molecular values

$$\nu^{\text{eddy}} \gg \nu. \quad (28.2)$$

We have more to say regarding the frictional operator in Section 28.3.1.

For conceptual and mathematical convenience, we find it useful to separate the horizontal velocity into two components. The first is the geostrophic velocity defined by a balance between pressure gradient and Coriolis accelerations

$$f \hat{z} \wedge \mathbf{u}_g = -\frac{1}{\rho_o} \nabla_z p \implies \mathbf{u}_g = \frac{1}{f \rho_0} \hat{z} \wedge \nabla p. \quad (28.3)$$

In some treatments, \mathbf{u}_g is referred to as the *pressure driven velocity*. The second is an *ageostrophic* or Ekman component defined by a balance between the frictional and Coriolis accelerations

$$f \hat{z} \wedge \mathbf{u}_e = \mathbf{F} \implies \mathbf{u}_e = -f^{-1} \hat{z} \wedge \mathbf{F}, \quad (28.4)$$

so that \mathbf{u}_e is the *frictional driven velocity*.

This velocity decomposition has the appearance of superposing linearly independent flows, one geostrophic (pressure driven) and one ageostrophic (friction driven). However, the flows are coupled

and thus not linearly independent. Namely, ageostrophic motions alter the pressure field which in turn affects the geostrophic flow. So the presence of friction and the associated ageostrophic flows lead to geostrophic flows differing from the inviscid case. Conversely, geostrophic flows affect the level of friction. Hence, the above decomposition does not reflect a physical decoupling of geostrophic and ageostrophic flows. Rather, it is only meant to help in our description of the flow and their force balances.

28.2 Horizontal balances in natural coordinates

We here study balances occurring in horizontal flows maintaining a frictional geostrophic balance. As per the definition (28.3), geostrophic motion occurs along lines of constant pressure, with frictionally induced deviations crossing isobars and providing a downgradient component to the fluid trajectory. Motivated by the discussion in Chapter 27, we make use of natural coordinates for the kinematics.

28.2.1 Natural coordinates according to isobars

We represent the horizontal flow according to natural coordinates defined along an arbitrary geopotential surface. Instead of defining the natural coordinates according to the flow direction, as done in Section 27.2 for the frictionless case, we here decompose the motion according to pressure contours (isobars)

$$\hat{z} = \hat{s} \wedge \hat{n} = \text{vertical direction} \quad (28.5a)$$

$$\hat{n} = \hat{z} \wedge \hat{s} = \text{down pressure gradient direction} \quad (28.5b)$$

$$\hat{s} = \hat{n} \wedge \hat{z} = \text{tangent to isobar.} \quad (28.5c)$$

The unit vector \hat{s} is tangent to isobars in the horizontal plane, whereas \hat{n} is perpendicular to isobars and oriented down the horizontal pressure gradient

$$\hat{n} = -\frac{\nabla_z p}{|\nabla_z p|}. \quad (28.6)$$

These natural coordinates are illustrated in Figure 28.1. For the case with zero friction, the flow reduces to the geostrophic balance in which fluid moves along isobars, $\hat{s} = \hat{u}$.

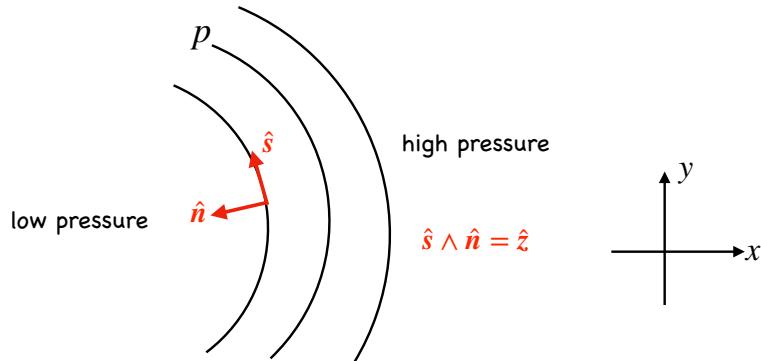


FIGURE 28.1: Natural coordinates defined along an arbitrary geopotential surface according to isobars in the horizontal plane. The normal direction, $\hat{n} = -\nabla_z p / |\nabla_z p|$, is oriented down the horizontal pressure gradient so that it points to the left of the geostrophic wind (facing downwind) in the northern hemisphere and to the right in the south. The tangent direction, \hat{s} , points along the isobar so that $\hat{s} \wedge \hat{n} = \hat{z}$.

28.2.2 Geostrophic and Ekman balances

As found when studying the geostrophic balance in Section 27.3, the geostrophic velocity flows along isobars and so only has a component in the \hat{s} direction

$$\hat{s} \cdot \mathbf{u}_g = -\frac{1}{f\rho_0} \frac{\partial p}{\partial n} \quad \text{and} \quad \hat{n} \cdot \mathbf{u}_g = 0. \quad (28.7)$$

In contrast, the Ekman velocity has a component both along and across isobars

$$\hat{s} \cdot \mathbf{u}_e = f^{-1} \hat{n} \cdot \mathbf{F} \quad \text{and} \quad \hat{n} \cdot \mathbf{u}_e = -f^{-1} \hat{s} \cdot \mathbf{F}. \quad (28.8)$$

As expected, the Ekman velocity vanishes when the frictional acceleration vanishes. These equations show that friction aligned along isobars drives Ekman velocities across isobars. Conversely, friction aligned across isobars drives Ekman velocities along isobars. If there is no friction component along the isobar, $\hat{s} \cdot \mathbf{F} = 0$, then the Ekman flow follows isobars just as the geostrophic flow. More commonly, friction has a nonzero component along isobars, which leads to Ekman flow crossing isobars.

We can arrive at a complementary perspective on the origin of cross-isobar flow through the following considerations. Without friction, the Coriolis and pressure gradient accelerations balance when the flow is geostrophic. In the presence of friction, the velocity is slowed so that the Coriolis acceleration weakens. If the pressure gradient acceleration is retained, as occurs if it is determined by large scale balances outside of the Ekman layer, then the Coriolis acceleration no longer balances the pressure gradient. Consequently, flow is diverted from isobars and develops a component down the pressure gradient.

28.2.3 Rayleigh drag

The relative simplicity of Rayleigh drag affords analytical expressions for the Ekman velocity using natural coordinates, written in terms of the geostrophic velocity (equation (28.7)). In doing so we are afforded an explicit illustration of how the Ekman flow provides a cross-isobar component to the flow in the direction down the pressure gradient. Before developing the Ekman flows we here summarize elements of Rayleigh drag.

In Section 22.3.2 we studied how Rayleigh drag affects the kinetic energy budget. As a reminder, consider a frictional acceleration in the form of a Rayleigh drag acting on the velocity field

$$\mathbf{F}^{\text{drag}} = -\frac{U_{\text{fric}} \mathbf{u}}{\delta} = -\gamma \mathbf{u}, \quad (28.9)$$

where δ is a depth scale and U_{fric} is a friction velocity scale with dimensions L/T . The ratio

$$\gamma = \frac{U_{\text{fric}}}{\delta} \quad (28.10)$$

has dimensions T^{-1} and is an inverse spin-down time. That is, if only Rayleigh drag affected changes to the horizontal momentum, $\partial_t \mathbf{u} = -\gamma \mathbf{u}$, then the flow would exponentially come to a halt with an e-folding time γ^{-1} . The drag is relatively large over rough surfaces, thus leading to a small e-folding time. In particular, drag on the lower atmospheric winds is larger over land than over the ocean. The reason is that trees, cities, and mountains dissipate more of the atmosphere's mechanical energy than interactions with the relatively smooth ocean surface (see Section 19.9.7 for a discussion of roughness length).

Rayleigh drag dissipates all flow features regardless of their spatial structure. That is, Rayleigh drag does not prefer any particular length scales in the fluid flow. This lack of scale selectivity contrasts to the Laplacian friction discussed in Section 28.3.1, with Laplacian friction dissipating small spatial scales more strongly than large scales. Correspondingly, Rayleigh drag does not generally provide the means to produce a boundary layer. Hence, when studying physics within the Ekman boundary layer in Section 28.3 we make use of Laplacian friction. But for now, Rayleigh drag provides a very useful means to analytically illustrate the role of friction in producing spiral flows with non-zero horizontal divergence.

28.2.4 Cross isobar flow driven by Rayleigh drag

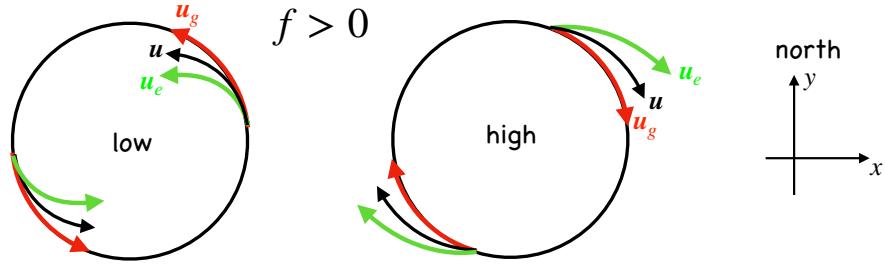


FIGURE 28.2: Illustrating frictional geostrophic flow in the northern hemisphere ($f > 0$). Left panel: geostrophic flow, \mathbf{u}_g , around a low pressure center is counter-clockwise and aligned with pressure isosurfaces. Friction aligned along the isobars drives flow down the pressures gradient, thus causing the fluid to spiral into the low pressure center. Right panel: the opposite oriented flow occurs around high pressure centers, where fluid spirals away from the high due to the cross-isobar flow driven by friction.

Making use of the Rayleigh drag (28.9) brings the expressions (28.8) for the Ekman velocity into the form

$$\hat{\mathbf{s}} \cdot \mathbf{u}_e = -(\gamma/f) \hat{\mathbf{n}} \cdot \mathbf{u}_e \quad (28.11a)$$

$$\hat{\mathbf{n}} \cdot \mathbf{u}_e = (\gamma/f) \hat{\mathbf{s}} \cdot (\mathbf{u}_e + \mathbf{u}_g). \quad (28.11b)$$

These equations allow us to express the Ekman velocity in terms of the geostrophic velocity (equation (28.7))

$$\hat{\mathbf{s}} \cdot \mathbf{u}_e = -(\gamma/f) \hat{\mathbf{n}} \cdot \mathbf{u}_e \quad (28.12a)$$

$$\hat{\mathbf{n}} \cdot \mathbf{u}_e = \hat{\mathbf{s}} \cdot \mathbf{u}_g \left[\frac{(\gamma/f)}{1 + (\gamma/f)^2} \right], \quad (28.12b)$$

so that the velocity components for the total velocity, $\mathbf{u} = \mathbf{u}_g + \mathbf{u}_e$, and its squared magnitude, are

$$\hat{\mathbf{s}} \cdot \mathbf{u}_g = -\frac{1}{f\rho_0} \frac{\partial p}{\partial n} \quad \text{geostrophic velocity (aligned along isobars)} \quad (28.13a)$$

$$\hat{\mathbf{s}} \cdot \mathbf{u} = \hat{\mathbf{s}} \cdot \mathbf{u}_g \left[\frac{1}{1 + (\gamma/f)^2} \right] \quad \text{isobaric velocity component} \quad (28.13b)$$

$$\hat{\mathbf{n}} \cdot \mathbf{u} = (\gamma/f) \hat{\mathbf{s}} \cdot \mathbf{u} \quad \text{normal (to left) velocity component} \quad (28.13c)$$

$$(\hat{\mathbf{s}} \cdot \mathbf{u})^2 + (\hat{\mathbf{n}} \cdot \mathbf{u})^2 = \frac{(\hat{\mathbf{s}} \cdot \mathbf{u}_g)^2}{1 + (\gamma/f)^2} \quad \text{horizontal kinetic energy per mass.} \quad (28.13d)$$

The cross-isobar flow (equation (28.13c)) is directly driven by the Rayleigh drag, and it is directed down the pressure gradient so long as the flow has a positive projection onto the tangent direction

$$\hat{s} \cdot \mathbf{u} > 0 \implies \hat{n} \cdot \mathbf{u} > 0. \quad (28.14)$$

When flow is moving counter-clockwise around a low pressure in the northern hemisphere, where \hat{n} points towards the low pressure center, then Rayleigh drag causes the fluid to spiral into the low pressure center. Conversely, when flow is moving clockwise around a high pressure, with \hat{n} pointing away from the high pressure center, then Rayleigh drag causes the fluid to spiral away from the high pressure center. We depict these cases in Figure 28.2. Furthermore, equation (28.13d) shows that when $\gamma \neq 0$ the magnitude of the total flow is reduced relative to the geostrophic flow, thus reflecting the dissipation of kinetic energy arising from Rayleigh drag.

28.2.5 Horizontal spiral plus vertical rising/sinking

Thus far, we have focused on the horizontal spiral motion as shown in Figure 28.2. Through continuity we infer a corresponding vertical motion induced by the convergence of mass into the low pressure center and the divergence of mass away from the high pressure center. Figure 28.3 illustrates the vertical motion in a bottom Ekman boundary layer of either the atmosphere or ocean whereby mass rises above a low pressure center in response to the horizontal convergence of mass in the Ekman layer. Conversely, mass diverges from the high pressure Ekman layer, with this divergence inducing a sinking motion over the high pressure to replace the diverging mass. In subsequent sections of this chapter we develop the formalism needed to compute the mass transport into and out of the Ekman boundary layer, with this mass transport playing a particularly important role in how mechanical forcing drives the ocean general wind-driven circulation.

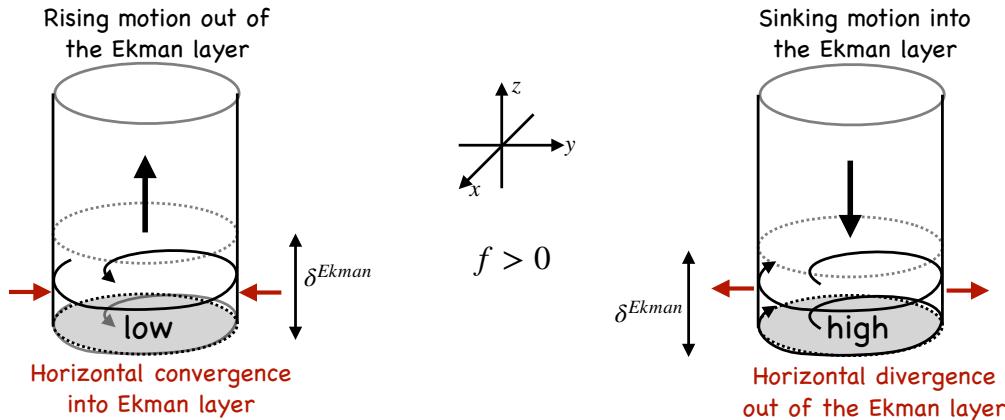


FIGURE 28.3: Illustrating the three-dimensional flow associated with Ekman layers in the northern hemisphere ($f > 0$) next to a bottom boundary. Left panel: flow spiralling into a low pressure center creates convergence of mass into the Ekman layer. Mass continuity means that flow must vertically leave the Ekman layer to thus enter the interior fluid above. Right panel: flow spiralling away from a high pressure center creates divergence of mass away from the Ekman layer. Mass continuity means that flow must vertically sink into the Ekman layer from above. The Ekman layer thickness is denoted by δ^{Ekman} (Section 28.3). An analogous picture holds for the surface Ekman layer in the ocean, yet with the Ekman layer at the top of the column rather than the bottom.

28.2.6 Further study

The discussion of Ekman velocity arising from Rayleigh drag follows a similar treatment in Section 7.4 of [Marshall and Plumb \(2008\)](#).

28.3 Ekman number and Ekman layer thickness

Friction that is most prominent near surface and bottom boundaries arises from vertical shears in the horizontal velocity, with shears leading to flow instabilities and the development of turbulence and associated turbulent friction. There is no deductive theory for turbulent friction so we must rely on empirical expressions. These expressions are typically based on a Laplacian viscous operator, partly motivated by the form arising from kinetic theory of gases briefly mentioned in Chapter 12 (equation (19.85)). Most notably, this operator supports the development of a boundary layer through maintenance of both the no-slip (homogeneous Dirichlet) and stress (Neumann) boundary conditions.

28.3.1 Laplacian vertical friction

We introduced the Laplacian friction operator in Section 19.7.5 when studying stress in fluids. The Laplacian friction operator considered here is given just in terms of the vertical shears of horizontal stresses

$$\mathbf{F}^{\text{viscous}} = \frac{1}{\rho_0} \frac{\partial \boldsymbol{\tau}}{\partial z} = \frac{\partial}{\partial z} \left[\nu^{\text{eddy}} \frac{\partial \mathbf{u}}{\partial z} \right], \quad (28.15)$$

with $\nu^{\text{eddy}} > 0$ a turbulent kinematic viscosity with dimensions $L^2 T^{-1}$. This form of the friction operator emulates the Laplacian operator representing molecular viscous friction (equation (19.85)). It is also the form most commonly used in theoretical and numerical models.

Expanding the derivative in equation (28.15) reveals that the Laplacian friction is nonzero where there is curvature in the vertical profile of the horizontal velocity, and where there is depth dependence to the viscosity and velocity

$$\mathbf{F}^{\text{viscous}} = \frac{\partial \nu^{\text{eddy}}}{\partial z} \frac{\partial \mathbf{u}}{\partial z} + \nu^{\text{eddy}} \frac{\partial^2 \mathbf{u}}{\partial z^2}. \quad (28.16)$$

The turbulent viscosity generally has a depth dependence, with larger values in the boundary layer where turbulence is most energetic. This form of the friction preferentially acts on velocity exhibiting nontrivial vertical structure, thus acting to smooth any depth dependent behaviour. In Section 48.4 we discuss further mathematical properties of Laplacian friction.

28.3.2 Non-dimensionalization

We non-dimensionalize the equations to isolate important non-dimensional numbers affecting the flow regime. In particular, we identify the Ekman number as a measure of the importance of friction relative to rotation, with friction important where the Ekman number is order unity or larger, and unimportant where the Ekman number is much smaller than unity.¹

We make use of the following scales and associated non-dimensional quantities

$$(x, y) = L(\hat{x}, \hat{y}) \quad z = H \hat{z} \quad (u, v) = U(\hat{u}, \hat{v}) \quad f = f_0 \hat{f} \quad p = P \hat{p} \quad (28.17)$$

where the hat terms are non-dimensional,² and we introduced typical scales for horizontal length (L), vertical height (H), velocity (U), Coriolis parameter (f_0), and pressure (P). For the pressure scale we assume it follows geostrophic scaling so that it can be written

$$P = f_0 \rho_0 U L. \quad (28.18)$$

¹For flows unaffected by rotation, the Reynolds number is the key non-dimensional number (Section 19.8) determining where viscous friction is important.

²In Section 27.2.1 we defined $\hat{\mathbf{u}}$ as the unit vector pointing along the trajectory of a fluid element. In contrast, we here let $\hat{\mathbf{u}}$ be the non-dimensional horizontal velocity.

Inserting the relations (28.17) into equation (28.1) leads to the non-dimensional frictional geostrophic equation

$$\hat{\mathbf{f}} \wedge \hat{\mathbf{u}} = -\hat{\nabla} \hat{p} + \frac{\mathbf{F}}{f_0 U}. \quad (28.19)$$

28.3.3 Defining the Ekman number and layer thickness

The *Ekman number* as a non-dimensional measure of the relative importance of the frictional acceleration due to vertical shears versus the Coriolis acceleration

$$Ek = \frac{\text{frictional acceleration from vertical shears}}{\text{Coriolis acceleration}}. \quad (28.20)$$

The Ekman number increases when there is more boundary layer turbulence, in which case the eddy viscosity, ν^{eddy} , is large relative to its small values in the interior region outside of the boundary layer. Additionally, the Ekman number increases when moving towards the equator, where the Coriolis parameter reduces.³

For the viscous stress form of Laplacian vertical frictional acceleration (equation (28.15))

$$\mathbf{F}^{\text{viscous}} = \frac{\nu^{\text{eddy}} U}{H^2} \frac{\partial^2 \hat{\mathbf{u}}}{\partial z^2}, \quad (28.21)$$

the Ekman number is given by

$$Ek = \frac{\nu^{\text{eddy}}}{f_0 H^2}. \quad (28.22)$$

If we take the vertical scale, H , equal to the depth scale over which interior flow processes occur, then the Ekman number will be very small, even if the eddy viscosity is relatively large. In this case we conclude that friction is negligible, as indeed it is for many purposes where the boundary layer is not of concern.

However, the Ekman number multiplies the highest derivative in equation (28.19). So setting the Ekman number to zero represents a *singular limit*, whose mathematical meaning is that we change the order of the differential equation when setting $Ek = 0$. Reducing the order of the differential equation then means we can only satisfy a reduced number of boundary conditions relative to the $Ek > 0$ case. In particular, with $Ek = 0$ we can no longer satisfy the no-slip condition at the solid-fluid boundary. In contrast, with any non-zero value of $Ek > 0$, viscosity will drag the flow to zero towards the boundary within a *boundary layer* where friction is leading order. We expect a boundary layer to form over a boundary layer thickness, $H = \delta^{\text{Ekman}}$, in which the Ekman number is order unity since friction is of leading order importance

$$H = \delta^{\text{Ekman}} \implies Ek = \frac{\nu^{\text{eddy}}}{f_0 (\delta^{\text{Ekman}})^2} = 1. \quad (28.23)$$

Turning this equation around we see that the vertical scale, δ^{Ekman} , defines the viscous Ekman boundary layer thickness as a function of the eddy viscosity and Coriolis parameter

$$Ek = 1 \implies \delta^{\text{Ekman}} = \delta^{\text{viscous}} = \sqrt{\nu^{\text{eddy}} / f_0}. \quad (28.24)$$

³When getting very close to the equator, our assumption of a frictional geostrophic balance breaks down so that other terms in the momentum equation, such as advection, become important.

28.3.4 Estimates for the vertical eddy viscosity

The eddy viscosity is not readily available from direct measurements or first principles theory. However, measuring the boundary layer thickness provides a means to infer a bulk viscosity for the boundary layer

$$\nu^{\text{eddy}} = f_0 (\delta^{\text{Ekman}})^2. \quad (28.25)$$

In the atmosphere, the boundary layer thickness is order 1000 m, so that at mid-latitudes, with $f_0 = 10^{-4} \text{ s}^{-1}$, we expect

$$\nu_{\text{atmos}}^{\text{eddy}} \sim 10^2 \text{ m}^2 \text{ s}^{-1}. \quad (28.26)$$

In the ocean, the upper ocean boundary layer depth, outside of the deep convection regions, is roughly 50 m, in which case

$$\nu_{\text{ocean}}^{\text{eddy}} \sim 0.25 \text{ m}^2 \text{ s}^{-1}. \quad (28.27)$$

28.4 Ocean surface Ekman layer

It is possible to establish some integrated mass transport properties of the Ekman layer even without specifying details of the friction (i.e., the viscosity) or the stratification. What is needed is the boundary stress. This stress is commonly estimated for the surface ocean given information about the wind speed and atmospheric stratification. Hence, Ekman theory has found much application to studies of the wind-driven ocean circulation, with the integrated properties the most consequential results from Ekman theory. We here focus on the ocean surface Ekman layer and then consider bottom Ekman boundaries for the atmosphere and ocean in Section 28.5.

28.4.1 Horizontal mass transport within the Ekman layer

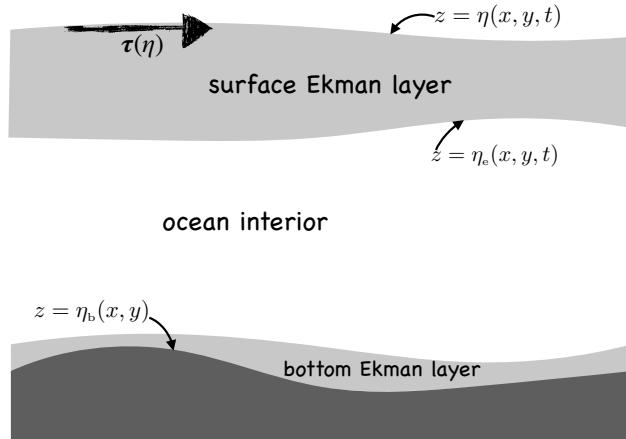


FIGURE 28.4: Ekman layer at the ocean surface, defined for vertical position $\eta_e(x, y, t) \leq z \leq \eta(x, y, t)$, with η_e specifying the Ekman layer bottom and η the free surface vertical position. Boundary stress from winds and/or sea ice imparts horizontal momentum to the upper ocean that is transmitted throughout the Ekman layer via the vertical divergence of turbulent horizontal shear stresses. As studied in Section 28.6, there is also a bottom Ekman layer created by stresses active next to the ocean bottom.

We are concerned with the mass budget for the ocean surface Ekman layer sitting between the Ekman layer bottom and the ocean free surface

$$\eta_e(x, y, t) \leq z \leq \eta(x, y, t), \quad (28.28)$$

as depicted in Figure 28.4. Knowledge of the mass budget has implications for how mechanical energy imparted to the boundary layer drives circulation well within the interior of the ocean. This transport and associated circulation are how Ekman mechanics, limited to the boundary layer, affect large-scale ocean circulation throughout the fluid column.

Integrating the horizontal Ekman balance (28.4) over the depth of the Ekman layer leads to

$$\mathbf{M}_e = \int_{\eta_e}^{\eta} \rho_o \mathbf{u}_e dz \implies f \hat{z} \wedge \mathbf{M}_e = \int_{\eta_e}^{\eta} \rho_o \mathbf{F} dz, \quad (28.29)$$

with \mathbf{M}_e the depth integrated ageostrophic horizontal mass transport within the Ekman boundary layer. Assuming friction to be in the form of a vertical stress divergence as in equation (28.15) leads to the horizontal Ekman mass transport

$$\mathbf{M}_e = -f^{-1} \hat{z} \wedge [\boldsymbol{\tau}(\eta) - \boldsymbol{\tau}(\eta_e)]. \quad (28.30)$$

Stress at the bottom of the Ekman layer, $\boldsymbol{\tau}(\eta_e)$, matches to the stress in the ocean interior. The stress in the fluid interior is generally much smaller than stress in the turbulent upper ocean surface, $\boldsymbol{\tau}(\eta)$, so that we can neglect $\boldsymbol{\tau}(\eta_e)$ when computing the mass transport. We are thus led to the expression for the surface stress induced ageostrophic horizontal mass transport within the upper ocean Ekman layer

$$\mathbf{M}_e = -f^{-1} \hat{z} \wedge \boldsymbol{\tau}(\eta). \quad (28.31)$$

The surface stress induced mass transport given by equation (28.31) is very useful in practice. Notably, we do not need to know the depth of the Ekman layer nor details about the boundary layer turbulence. Rather, the mass transport is determined solely by the surface boundary stress. Furthermore, the horizontal mass transport within the Ekman layer is directed at right angles to the surface stress, as depicted in Figure 28.5, with mass transport to the right of the surface stress in the northern hemisphere and to the left in the southern hemisphere.

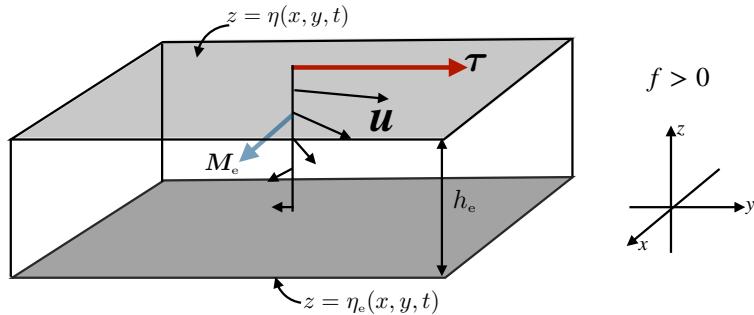


FIGURE 28.5: Horizontal transport integrated over the depth of the surface ocean Ekman layer in the region $\eta_e(x, y, t) \leq z \leq \eta(x, y, t)$. The net mass transport is directed at perpendicular to the boundary stress, oriented to the right in the northern hemisphere and to the left in the southern hemisphere. Here, the boundary stress, $\boldsymbol{\tau}$, is shown directed to the north so that in the northern hemisphere ($f > 0$), the depth integrated horizontal Ekman transport, \mathbf{M}_e , is to the east. This perpendicular mass transport is the result of the vertically spiralling Ekman induced flow. The Ekman mass transport resulting from boundary stress is independent of the assumptions made about ocean friction within the ocean boundary layer, and it is independent of the stratification assumed for the Ekman layer.

28.4.2 Example Ekman mass transports

A particularly relevant example is when zonal boundary stress is eastward, $\boldsymbol{\tau}(\eta) = |\tau| \hat{x}$, in which the Ekman transport is

$$\mathbf{M}_e = -(|\tau|/f) \hat{y}, \quad (28.32)$$

which points equatorward. Conversely, in the equatorial region winds are generally easterly so that $\tau(\eta) = -|\tau| \hat{x}$, in which case the horizontal Ekman mass transport causes waters to move poleward away from the equator. Mass continuity is then satisfied by upwelling waters along the equator within the Ekman layer. We sketch the elements of this flow in Figure 28.6. Exercise 28.1 considers an analogous situation for a channel in the southern hemisphere, thus illustrating a basic feature of the wind-driven overturning circulation in the Southern Ocean.

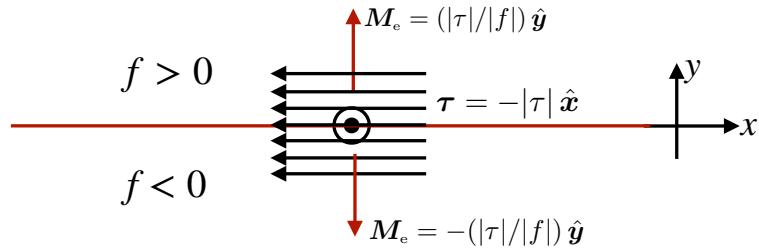


FIGURE 28.6: Easterly winds along the equator drive poleward horizontal mass transport within the upper ocean Ekman layer, as per equation (28.32). Furthermore, as discussed in Section 28.4.3, a steady state Ekman layer mass budget is typically realized by the upwelling of interior waters into the Ekman layer.

A second example concerns the case of a wind stress with a component aligned with a coastline as depicted in Figure 28.7. When the horizontal Ekman mass transport is directed away from coast, a steady mass balance in the Ekman layer will be associated with the upwelling of waters from beneath the Ekman layer. Conversely when the horizontal mass transport is directed toward the coast, steady mass balance is realized by coastal downwelling. This process is very important for coastal physical and biological oceanography.

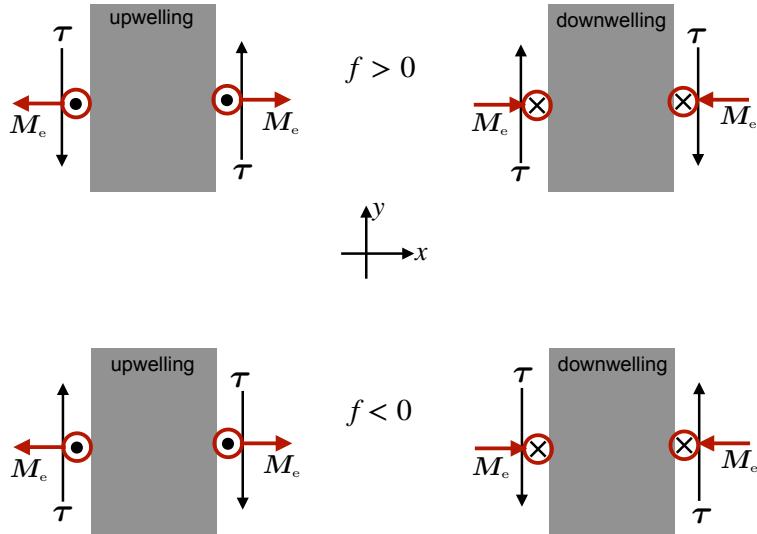


FIGURE 28.7: Wind stresses with a component that is parallel to coastlines lead to horizontal mass transport away from or towards the coast. Mass continuity then leads to coastal upwelling or downwelling of waters into or out of the Ekman layer. Here we depict the various scenarios for the northern hemisphere (top row) and southern hemisphere (bottom row), with the gray regions representing land, the black vectors the wind stresses, and the red arrows the horizontal wind induced mass transports in the Ekman layer. This figure is adapted from Figure 4-24 of [von Arx \(1962\)](#).

28.4.3 Mass budget for the Ekman layer

As seen in Figure 28.3, the horizontal transport of fluid within the Ekman layer induces a vertical transport into or out of the Ekman layer. To obtain a mathematical expression for the vertical transport, integrate the continuity equation $\nabla \cdot \mathbf{v} = 0$ over the vertical extent of the Ekman layer

$$\frac{\partial}{\partial x} \left[\int_{\eta_e}^{\eta} u \, dz \right] + \frac{\partial}{\partial y} \left[\int_{\eta_e}^{\eta} v \, dz \right] + [w(\eta) - \mathbf{u}(\eta) \cdot \nabla \eta] - [w(\eta_e) - \mathbf{u}(\eta_e) \cdot \nabla \eta_e] = 0. \quad (28.33)$$

For a Boussinesq fluid, the kinematic boundary condition at the ocean free surface is given by equation (17.5)

$$w + \frac{Q_m}{\rho_0} = \frac{\partial \eta}{\partial t} + \mathbf{u} \cdot \nabla \eta \quad \text{at } z = \eta(x, y, t). \quad (28.34)$$

Similarly, at the bottom of the Ekman layer we measure the volume transport through this layer by computing the dia-surface transport, $w^{(\dot{\eta}_e)}$, according to equation (44.38)

$$w^{(\dot{\eta}_e)} = w - (\partial_t z + \mathbf{u} \cdot \nabla z) \quad \text{at } z = \eta_e(x, y, t). \quad (28.35)$$

The sign convention is such that $w^{(\dot{\eta}_e)} > 0$ means that mass enters (entrains into) the surface Ekman layer through its base, whereas $w^{(\dot{\eta}_e)} < 0$ means that mass leaves (detains from) the surface Ekman layer base.

Using the kinematic boundary conditions (28.34) and (28.35) in the depth integrated volume budget (28.33), and rearranging, leads to the Ekman layer mass budget

$$\rho_0 w^{(\dot{\eta}_e)} + Q_m = \rho_0 \partial h_e / \partial t + \nabla \cdot \mathbf{M}. \quad (28.36)$$

In this equation we wrote

$$h_e = \eta - \eta_e \quad (28.37)$$

for the thickness of the Ekman layer and

$$\mathbf{M} = \rho_0 \int_{\eta_e}^{\eta} \mathbf{u} \, dz \quad (28.38)$$

for the Ekman layer integrated horizontal mass transport. As a check on the above manipulations, let the Ekman layer go to the ocean bottom (so that $w^{(\dot{\eta}_e)} = 0$ and $\eta_e = \eta_b$), in which case the mass budget equation (28.36) correctly reduces to the kinematic free surface equation for the full ocean column as given by equation (17.8)

$$\rho_0 \frac{\partial \eta}{\partial t} = Q_m - \rho_0 \nabla \cdot \mathbf{U}, \quad (28.39)$$

with $\mathbf{U} = \int_{\eta_b}^{\eta} \mathbf{u} \, dz$ the depth integrated horizontal velocity.

The horizontal mass transport given by equation (28.38) has a contribution from both the Ekman transport and remaining processes, such as geostrophic flow and ageostrophic flows not associated with Ekman. We write this mass transport in the form

$$\mathbf{M} = \mathbf{M}_e + \mathbf{M}_{\text{other}}. \quad (28.40)$$

The horizontal Ekman transport is determined by the boundary stress according to equation (28.31), with its divergence given by

$$\nabla \cdot \mathbf{M}_e = \hat{z} \cdot [\nabla \wedge (\boldsymbol{\tau}/f)]. \quad (28.41)$$

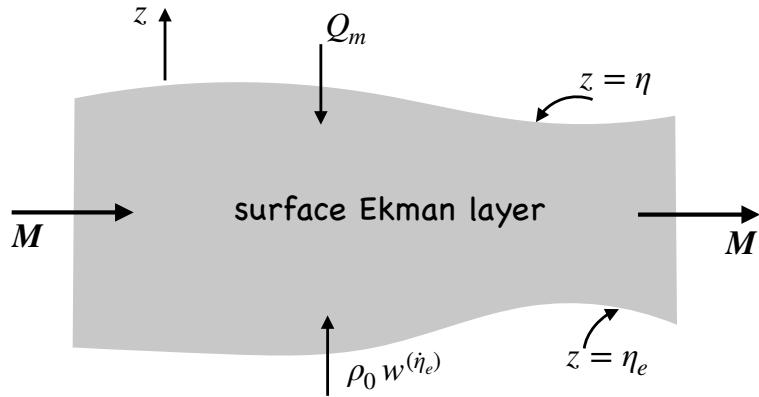


FIGURE 28.8: Mass budget over the surface Ekman layer of the ocean, with contributions from surface mass transport, Q_m , transport through the bottom of the layer, $w^{(\dot{\eta}_e)}$, and horizontal transport, \mathbf{M} , within the layer. If there are any imbalances then the layer thickness will have a nonzero time tendency, $\partial h_e / \partial t \neq 0$. Our sign convention is such as $w^{(\dot{\eta}_e)} > 0$ corresponds to water entering (entraining into) the Ekman layer (vertically upward motion) through the Ekman layer base at $z = \eta_e$, and likewise $Q_m > 0$ corresponds to water entering the Ekman layer through the free surface at $z = \eta$. Entrainment through the base of the Ekman layer is referred to as *Ekman suction* or *Ekman upwelling*. The opposite case is referred to as *Ekman pumping* or *Ekman downwelling* when water leaves the Ekman layer and enters the ocean interior.

This result brings the Ekman layer mass budget (28.36) into the form

$$\rho_0 w^{(\dot{\eta}_e)} + Q_m = \rho_0 \frac{\partial h_e}{\partial t} + \nabla \cdot \mathbf{M}_{\text{other}} + \hat{\mathbf{z}} \cdot [\nabla \wedge (\boldsymbol{\tau}/f)]. \quad (28.42)$$

The left hand side measures the mass transport crossing the bottom of the Ekman layer, $\rho_0 w^{(\dot{\eta}_e)}$, plus the transport crossing the free surface, Q_m . This transport balances a time change in the Ekman layer thickness (first right hand side term) plus the horizontal divergence of mass within the layer. A steady state Ekman layer thickness, $\partial h_e / \partial t = 0$, is realized if the horizontal divergence of mass within the Ekman layer is exactly balanced by mass entering the Ekman layer through the top and/or bottom of the layer. We illustrate this budget in Figure 28.8. For example, in the equatorial case of Figure 28.6, the diverging horizontal Ekman layer flow induced by easterly winds (poleward on both sides of the equator) is balanced by water upwelling into the Ekman layer through the base, $w^{(\dot{\eta}_e)} > 0$, along with a generally smaller effects from surface mass fluxes through Q_m and possible other contributions through $\nabla \cdot \mathbf{M}_{\text{other}}$.

28.4.4 Ekman layer coupled to the geostrophic interior

The effects from boundary stress in equation (28.42) warrant particular attention whereby

$$\rho_0 w_{\text{Ekman}}^{(\dot{\eta}_e)} \equiv \nabla \cdot \mathbf{M}_e = \hat{\mathbf{z}} \cdot [\nabla \wedge (\boldsymbol{\tau}/f)]. \quad (28.43)$$

The stress curl, as well as changes in f on the sphere, drive vertical motion through the base of the Ekman layer. The flow crossing the Ekman layer boundary acts to stretch or compress vertical fluid columns in the adjoining fluid interior. Interior fluid columns in a rotating fluid are stiffened through the effects of Taylor-Proudman (Section 26.5.3). From our understanding of vorticity (studied in Chapter 36), particularly the notions of vortex stretching, we see that the Ekman induced stretching/compression of interior fluid columns leads to a change in vorticity of the fluid interior, and can then lead to meridional motion due to the beta effect.

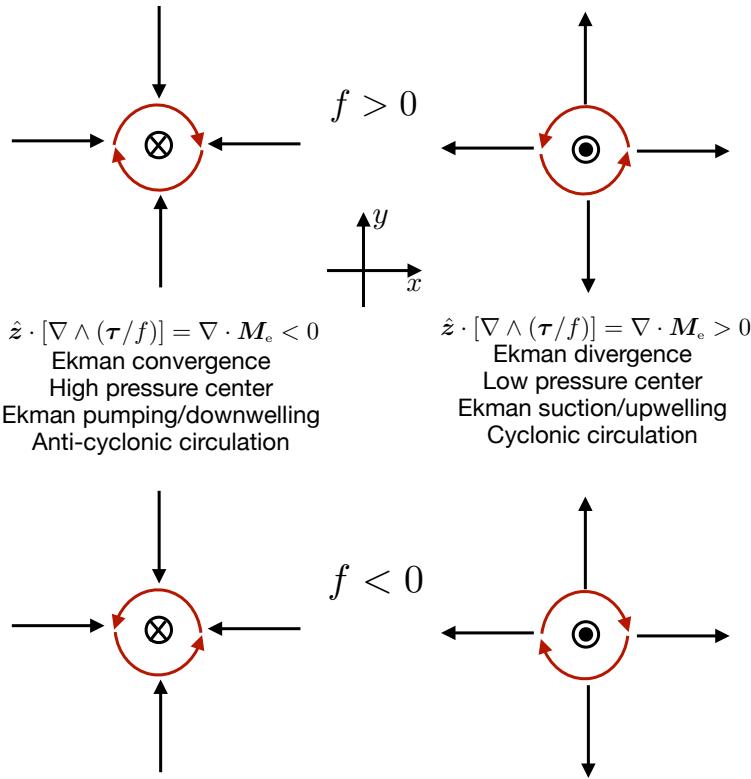


FIGURE 28.9: Plan view depicting the steady state horizontal and vertical transport in the surface Ekman layer (ignoring the possible transport across the ocean surface with Q_m). The left column shows horizontal wind-induced convergence in the Ekman layer that induces vertical pumping/downwelling, where $\nabla \cdot \mathbf{M}_e = \hat{z} \cdot [\nabla \wedge (\boldsymbol{\tau}/f)] < 0$. The right column shows the horizontal divergence within the Ekman layer that is balanced by suction/upwelling, where $\nabla \cdot \mathbf{M}_e = \hat{z} \cdot [\nabla \wedge (\boldsymbol{\tau}/f)] > 0$. The top row is for the northern hemisphere, with $f > 0$, and the bottom row is for the southern hemisphere. The red arrows depict the sense for the induced circulation. Ekman pumping is associated with anti-cyclonic circulation (clockwise in the northern hemisphere and anti-clockwise in southern hemisphere). In contrast, Ekman suction is associated with cyclonic circulation. The circulation is supported by pressure gradients, with high pressure in regions of Ekman convergence, $\nabla \cdot \mathbf{M}_e < 0$, due to the accumulation of mass towards the center, thus giving rise to anti-cyclonic geostrophic flow in the interior. The opposite holds for regions of Ekman divergence, $\nabla \cdot \mathbf{M}_e > 0$, where water leaves the region thus leaving a low pressure center and inducing a cyclonic interior geostrophic flow.

Consider an example with $\nabla \cdot \mathbf{M}_e = \hat{z} \cdot [\nabla \wedge (\boldsymbol{\tau}/f)] > 0$, so that winds induce a divergence within the Ekman layer. In a steady state, the interior flow accommodates this Ekman layer mass divergence by upwelling water through the Ekman layer base, $w^{(\dot{\eta}_e)} > 0$. This process of entraining interior water into the Ekman layer is known as *Ekman suction* or *Ekman upwelling*. For the opposite case with $\hat{z} \cdot [\nabla \wedge (\boldsymbol{\tau}/f)] < 0$, water leaves (detrains) from the Ekman layer and moves into the interior. Water detraining from the Ekman layer is known as *Ekman pumping* or *Ekman downwelling*. As water diverges it produces a local low pressure so that the induced flow in the geostrophic interior is cyclonic around a region of Ekman divergence/upwelling. Conversely, the induced interior geostrophic flow is anti-cyclonic around a region of Ekman convergence/downwelling. Figure 28.9 provides an illustration for the variety of cases found in the northern and southern hemispheres.

In the language of vorticity, developed in Part VI of this book, Ekman upwelling with $w^{(\dot{\eta}_e)} > 0$ leads to vortex stretching of interior fluid columns, whereas Ekman downwelling with $w^{(\dot{\eta}_e)} > 0$ squashes the interior fluid columns. Vertical stiffening through Taylor-Proudman within the

geostrophic interior, coupled to Ekman induced vortex stretching/squashing, makes what happens within the Ekman boundary layer of primary importance to the interior geostrophic flow. This boundary-interior coupling forms a key mechanism for how mechanical forcing from surface boundary stress creates the *wind driven* ocean circulation. It is notable that the coupling between boundary layer and interior flow is absent from non-rotating boundary layer flows. In Figure 28.10 we offer a highly idealized schematic of the circulation implied by Ekman dynamics in a homogeneous fluid on an *f*-plane, thus illustrating the coupling of the upper surface Ekman layer to the geostrophic interior and then to the bottom Ekman layer.

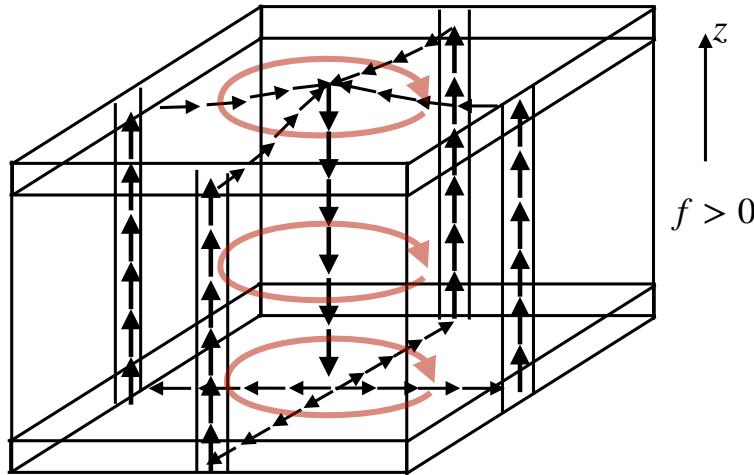


FIGURE 28.10: Schematic Ekman gyre circulation in a rectangular box in the northern hemisphere on an *f*-plane as forced by an anti-cyclonic wind stress curl. The wind stress curl causes fluid to pile up in the center of the gyre and develop an anti-cyclonic flow. The horizontal convergence of mass towards the center of the gyre leads to downwelling into the gyre interior below the Ekman layer (as per the upper left panel of Figure 28.9). This downwelling water enters the geostrophic interior and further into the bottom Ekman layer. As discussed in Section 28.5, the downwelling water in the bottom Ekman layer causes a horizontal divergence of mass. Assuming the domain has fixed vertical side walls, continuity requires an upwelling along the outer portion of the gyre. If the wind stress is symmetric about the center of the domain, then so is the flow. However, as studied in Section 35.9, the axial symmetry is broken on the *β*-plane, in which flow is stronger on the western side of the gyre even if the wind remains symmetric about the domain center. This figure is adapted from Figure 6-15 of [von Arx \(1962\)](#).

28.4.5 Further study

The following videos offer visuals to help develop further intuition for Ekman boundary flows.

- [This 4-minute video from Science Primer](#) provides an overview of how Ekman transport affects ocean circulation features near the coast and in open ocean gyres.
- [This video from MIT Earth, Atmospheric, and Planetary Sciences](#) illustrates the spiral flow found within an Ekman layer as realized in a rotating tank experiment.
- [This video from the UCLA SpinLab](#), near the 18 minute mark, shows how Ekman transport helps to explain the garbage patches found near the center of the ocean's sub-tropical gyres.
- [This video from the University of Chicago](#), starting near the 23 minute mark, provides examples of Ekman layers in a rotating tank. The other portions of this video exhibit many other novel aspects of rotating fluids and is highly recommended.

28.5 Bottom Ekman layer

In this section we study the mechanics of a bottom Ekman boundary layer. For the mass transport, we merely translate the results from the upper ocean Ekman layer considered in Section 28.4, whereby the mass transports are specified by the bottom boundary stress. We go further in this analysis by also providing an analytic expression for the velocity profile within the Ekman layer, with the velocity profile allowing us to diagnose the bottom boundary stress according to the Neumann boundary condition placed on the horizontal velocity (equation (28.56) below).

28.5.1 Horizontal mass transport within the Ekman layer

Turning the derivation from Section 28.4.1 upside-down leads to a bottom horizontal Ekman mass transport that is directed to the left of the bottom stress in the northern hemisphere and to the right in the southern hemisphere

$$\mathbf{M}_e = +f^{-1} \hat{\mathbf{z}} \wedge \boldsymbol{\tau}(\eta_b). \quad (28.44)$$

Consider the special case where this bottom Ekman layer is the bottom of the atmosphere sitting over the ocean. Now the stress at the bottom of the atmosphere is equal to that acting on the top of the ocean.⁴ Hence, the frictional stress induced mass transport (28.31) for the upper ocean is equal and opposite to the frictional stress induced mass transport in the atmosphere

$$\mathbf{M}_e^{\text{ocn}} = -f^{-1} \hat{\mathbf{z}} \wedge \boldsymbol{\tau}(\eta_b) = -\mathbf{M}_e^{\text{atm}} \implies \mathbf{M}_e^{\text{ocn}} + \mathbf{M}_e^{\text{atm}} = 0. \quad (28.45)$$

Since the density of the atmosphere and ocean are quite different, the equal mass transports correspond to very different volume transports.

28.5.2 Mass budget for the bottom Ekman layer

Following the derivation of the Ekman layer mass budget in Section 28.4.3, and assuming no mass enters through the solid earth, we are led to the mass budget for the bottom Ekman boundary layer⁵

$$\rho_0 w^{(\dot{\eta}_e)} = \rho_0 \frac{\partial h_e}{\partial t} - \nabla \cdot \mathbf{M} \quad (28.46)$$

where $z = \eta_e(x, y, t)$ is the vertical position for the top of the bottom Ekman layer, and

$$h_e = \eta_e - \eta_b \quad (28.47)$$

is the Ekman layer thickness. For a steady state, the budget equation (28.4.3) says that the horizontal convergence of mass into the bottom Ekman boundary layer leads to a detrainment of mass from the Ekman layer into the interior fluid above (upwelling). Conversely, when fluid horizontally diverges from the bottom Ekman layer there is a balance from an entrainment (downwelling) of fluid from the interior into the Ekman layer. This orientation for the mass transport is illustrated in Figure 28.3 as part of our earlier discussion.

⁴Recall our discussion of stress in Chapter 19, whereby stress on one side of an interface matches that on the other, which is a result following from Newton's third law.

⁵Note the sign swap in the budget (28.46) in front of the horizontal transport term as compared to the surface boundary layer mass budget (28.36).

28.5.3 An analytic bottom Ekman spiral velocity

When studying the bottom boundary layer for the atmosphere or the surface boundary layer for the ocean, we are generally afforded an estimate of the frictional boundary stress, $\tau(\eta)$. In turn, we can estimate the boundary stress induced mass transport and its divergence, and we can do so without making assumptions about the density profile or viscous stresses within the boundary layer. However, this estimate is less directly accessible for the ocean bottom Ekman layer, where we need $\tau(\eta_b)$ to determine the mass transport. We here take an alternative approach that produces an analytic profile for the velocity within the Ekman layer, so long as we know the viscosity within the boundary layer. Knowing the velocity profile then affords an estimate of the boundary stress. This approach requires a few assumptions that are not always met, in particular it requires the viscosity. Even so, it provides physical insights that further our understanding of Ekman mechanics thus motivating the analysis.

Physical configuration

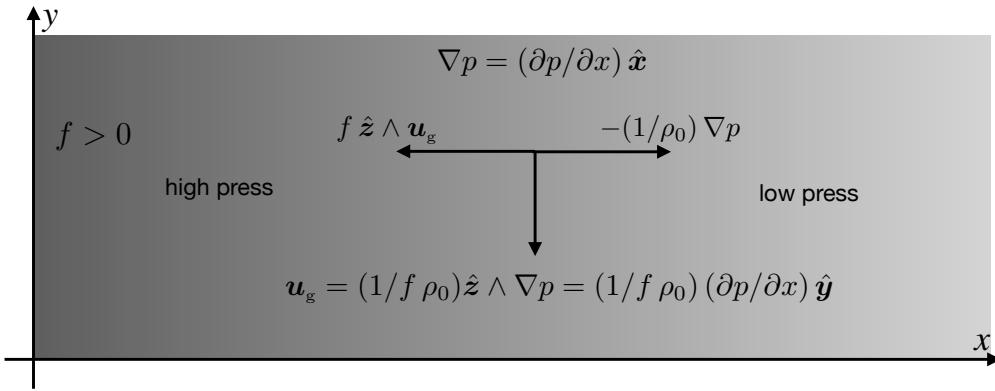


FIGURE 28.11: A southward geostrophic flow in the northern hemisphere ($f > 0$) induced by an eastward pressure gradient acceleration (low pressure to the east) that balances a westward Coriolis acceleration.

We assume flow above the boundary layer is geostrophic and supported by a prescribed depth-independent pressure gradient

$$\mathbf{u}_g = 1/(f \rho_0) \hat{z} \wedge \nabla p. \quad (28.48)$$

A specific example is given by Figure 28.11, where a southward geostrophic flow in the northern hemisphere has an eastward pressure gradient acceleration (low pressure to the east) balanced by a westward Coriolis acceleration.

For flow within the boundary layer, viscous stresses exchange horizontal momentum vertically between the inviscid interior and the bottom no-slip condition. This viscous exchange slows the boundary layer velocity relative to the interior velocity above the boundary layer. Since the pressure gradient is assumed to be depth-independent and prescribed, the slower velocity within the boundary layer means that the Coriolis acceleration is unable to balance the pressure gradient acceleration. This imbalance between pressure acceleration and Coriolis acceleration leads to a down pressure gradient component to the boundary layer velocity. As a result, the velocity spirals downward toward the bottom.

The key assumptions required to produce an analytic Ekman spiral velocity are: (i) the fluid within the Ekman layer has a constant density, ρ_0 ; (ii) the Coriolis parameter is a constant as per the f -plane; (iii) the prescribed pressure gradient is depth independent so that the associated

geostrophic velocity is depth independent; (iv) the eddy viscosity is constant within the boundary layer and zero in the interior region above the boundary layer; (v) the flow is steady. The homogeneous density assumption is motivated by the rather small vertical scale of the bottom Ekman layer (tens of meters) relative to the horizontal scales over which density varies in the ocean bottom (tens to hundreds of kilometers). A constant viscosity is not always realistic since the turbulent viscosity is generally inhomogeneous within boundary layers. Even so, by assuming sufficient time for statistics to become stationary we can expect to make good use of the time averaged viscosity.

Velocity profile within the Ekman layer

Bringing the above assumptions into the frictional geostrophic equations (28.1) leads to

$$f \hat{z} \wedge \mathbf{u} = -(1/\rho_0) \nabla p + \nu^{\text{eddy}} \partial_{zz} \mathbf{u} \quad \text{frictional geostrophic} \quad (28.49\text{a})$$

$$\partial_z (\nabla_z p) = 0 \quad z\text{-independent horizontal pressure gradient} \quad (28.49\text{b})$$

$$\mathbf{u}(\eta_b) = 0 \quad \text{no-slip bottom boundary condition} \quad (28.49\text{c})$$

$$\mathbf{u}(\infty) = \mathbf{u}_g \quad \text{matching to geostrophic interior.} \quad (28.49\text{d})$$

Decomposing the velocity into its geostrophic and ageostrophic components

$$\mathbf{u} = \mathbf{u}_g + \mathbf{u}_e, \quad (28.50)$$

leads to

$$f \hat{z} \wedge \mathbf{u}_e = \nu^{\text{eddy}} \frac{\partial^2 \mathbf{u}_e}{\partial z^2} \quad (28.51\text{a})$$

$$\mathbf{u}_e(\eta_b) = 0 \quad (28.51\text{b})$$

$$\mathbf{u}_e(\infty) = 0. \quad (28.51\text{c})$$

To reach this result we noted that the geostrophic velocity is depth independent within the boundary layer where the density is constant, so that the geostrophic velocity does not contribute to the viscous friction operator. These coupled second order differential equations are equivalent to the two uncoupled fourth order differential equations

$$(f^2 + \nu^2 \partial_{zzzz}) \mathbf{u}_e = 0 \quad (28.52\text{a})$$

$$\mathbf{u}_e(\eta_b) = 0 \quad (28.52\text{b})$$

$$\mathbf{u}_e(\infty) = 0. \quad (28.52\text{c})$$

A solution to these equations renders the Ekman boundary layer profile

$$u = u_g \left[1 - e^{-\Delta z/h_e} \cos(\Delta z/h_e) \right] - v_g e^{-\Delta z/h_e} \sin(\Delta z/h_e) \quad (28.53\text{a})$$

$$v = v_g \left[1 - e^{-\Delta z/h_e} \cos(\Delta z/h_e) \right] + u_g e^{-\Delta z/h_e} \sin(\Delta z/h_e) \quad (28.53\text{b})$$

$$h_e^2 = 2 \nu^{\text{eddy}} / |f| \quad (28.53\text{c})$$

$$\Delta z = z - \eta_b. \quad (28.53\text{d})$$

Properties of the Ekman spiral

The spiral velocity profile (28.53a)-(28.53b) vanishes at the bottom, $\Delta z = 0$ or $z = \eta_b$, reflecting the no-slip bottom boundary condition. It also reduces to the interior geostrophic velocity at $\Delta z = \infty$.

In moving downward through the boundary layer, the boundary layer current deflects to the left of the interior geostrophic current (deflection is to the right in the southern hemisphere). Making use of the integral identities

$$\int_{\eta_b}^{\infty} e^{-\Delta z/h_e} \cos(\Delta z/h_e) d\Delta z = h_e/2 \quad \text{and} \quad \int_{\eta_b}^{\infty} e^{-\Delta z/h_e} \sin(\Delta z/h_e) d\Delta z = -h_e/2 \quad (28.54)$$

leads to the frictionally induced mass transport within the Ekman layer

$$\mathbf{M}_e = \rho_0 \int_{\eta_b}^{\infty} \mathbf{u}_e dz \quad (28.55a)$$

$$= (h_e \rho_0 / 2) [-\hat{\mathbf{x}} (u_g + v_g) + \hat{\mathbf{y}} (u_g - v_g)] \quad (28.55b)$$

$$= f^{-1} \hat{\mathbf{z}} \wedge \boldsymbol{\tau}(\eta_b), \quad (28.55c)$$

where we introduced the bottom stress according to equation (28.44). Furthermore, this stress is given by the Neumann boundary condition placed on the horizontal velocity

$$\boldsymbol{\tau}(\eta_b) = \rho_0 \nu^{\text{eddy}} \left[\frac{\partial \mathbf{u}_e}{\partial z} \right]_{z=\eta_b} = (f h_e / 2) [\hat{\mathbf{x}} (u_g - v_g) + \hat{\mathbf{y}} (u_g + v_g)]. \quad (28.56)$$

Following from equation (28.46) we see that the wind induced mass transport across the top of the boundary layer is

$$\rho_0 w_{\text{Ekman}}^{(j_e)} = -\nabla \cdot \mathbf{M}_e = -\nabla \cdot [f^{-1} \hat{\mathbf{z}} \wedge \boldsymbol{\tau}(\eta_b)] = (h_e / 2) \zeta_g, \quad (28.57)$$

where we made use of the f -plane assumption to set $\nabla \cdot \mathbf{u}_g = 0$ and introduced the relative vorticity of the geostrophic flow

$$\zeta_g = \partial_x v_g - \partial_y u_g. \quad (28.58)$$

Hence, the relative vorticity of the interior geostrophic flow provides a divergence to the horizontal mass transport within the bottom Ekman layer. We illustrate the Ekman mass transport and bottom stress in Figure 28.12.

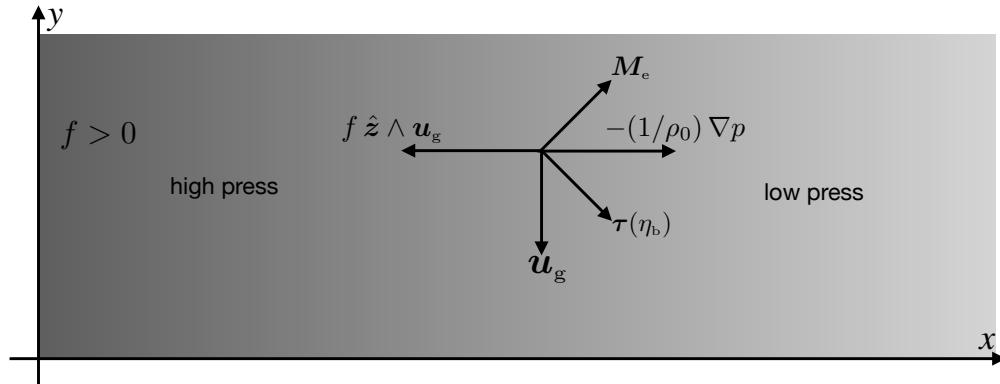


FIGURE 28.12: Mass transport and bottom stress in a homogeneous bottom Ekman layer on an f -plane in the northern hemisphere. Note how the boundary layer stress is directed $\pi/4$ radians to the left of the prescribed interior geostrophic velocity. Furthermore, the horizontal Ekman transport is itself $\pi/2$ radians to the left of the stress, which is then $3\pi/4$ to the left of the geostrophic velocity.

28.5.4 Comments

The Ekman spiral is a striking solution to the linear frictional geostrophic equations in a homogeneous fluid next to a no slip boundary. The spiral profile has been measured in the atmosphere and can be produced in the laboratory. However, it has proven difficult to measure in the ocean (see [Gnanadesikan and Weller \(1995\)](#) for an example). Even so, the effects from Ekman transport are robust features of the theory developed in this chapter.

28.6 Arrested bottom Ekman flows

When studying the vertical velocity profile within the bottom boundary layer in Section 28.5, we assumed the density to be uniform within the boundary layer. A uniform density is commonly assumed for Ekman layers. However, [MacCready and Rhines \(1991, 1993\)](#); [Garrett et al. \(1993\)](#) examined that assumption in their study of bottom boundary layers in the presence of a nonzero vertical density stratification next to sloping topography. They point out the remarkable possibility of a vanishing frictional stress within the sloping bottom boundary layer. That is, rather than nonzero friction leading to a zero flow next to the bottom, flow is arrested by a compensation of pressure gradient accelerations. The vanishing frictional stress motivates the term *slippery Ekman layers*, whereas the equivalent term *arrested Ekman layer* refers to the zero boundary layer flow. We here briefly describe the arrested Ekman layer by studying Figure 28.13.

28.6.1 Description

In both panels of Figure 28.13 we depict a prescribed barotropic pressure acceleration that acts eastward and has a corresponding southward geostrophic flow ($f > 0$) and westward Coriolis acceleration. This configuration is identical to that shown in Figure 28.11 when studying the bottom Ekman spiral over a flat bottom. As the flow enters the bottom boundary layer it slows so that its Coriolis acceleration no longer balances the barotropic pressure (we have seen this imbalance throughout this chapter). The bottom Ekman layer flow sends water cross-slope in the direction down the barotropic pressure gradient. In the configuration shown here, the cross-slope flow advects denser water upslope. In so doing, a baroclinic pressure acceleration develops adjacent to the bottom boundary layer, with the pressure acceleration pointing opposite to the barotropic pressure acceleration. A steady state is realized when the two pressure accelerations balance, in which case flow within the boundary layer halts.

It is remarkable that the dynamical balance for the arrested bottom boundary layer flow does not involve friction. Rather, arrest happens when the two pressure gradients balance and the Coriolis acceleration vanishes. That is, referring to the frictional geostrophic equation (28.1), each term separately vanishes in the arrested state.

28.6.2 Applications

[MacCready and Rhines \(1991, 1993\)](#); [Garrett et al. \(1993\)](#) study the transient adjustment leading to the arrested state. They derive the expression for the time scale for adjustment to the arrest state

$$T_{\text{arrest}} = \frac{|f|}{N^2 s^2}, \quad (28.59)$$

where s is the slope of the bottom and N^2 is the squared buoyancy frequency of the ambient water. The time is less in regions of strong stratification (N^2 relatively large), large topographic slopes

(s^2 large), and low latitudes (f small). Conversely, the infinite time for either $N^2 = 0$ or $s^2 = 0$ indicates the need for both stratification and topographic slopes to render an arrest. Finally, we note the absence of any dissipation parameters (e.g., viscosity) from the time scale. Friction is needed to support the Ekman layer where flow crosses isobars, but the time to reach the arrest state is independent of friction.

Wåhlin et al. (2012) interpreted their observations in the Amundsen Sea according to the arrested Ekman boundary layer and found T_{arrest} to be just a few hours. Additionally, the numerical model study of *Spence et al. (2017)*; *Webb et al. (2019)* point to the ability of barotropic shelf waves around Antarctica to provide an onshore directed barotropic pressure acceleration. Through the arrested Ekman layer mechanism described here, they find that the barotropic pressure is compensated through an upslope transport of relatively warm deep water in regions of the Antarctic Peninsula.



28.7 Exercises

EXERCISE 28.1: EKMAN MASS TRANSPORT IN A SOUTHERN CHANNEL

Consider a southern hemisphere channel (e.g., and idealized Southern Ocean) with a zonal wind stress that has a meridional dependence such as shown in Figure 28.14. Following the discussion in Sections 28.4.1 and 28.4.4, sketch the sense for the horizontal and vertical Ekman mass transport arising from this wind stress. Show the transport for regions to the north and to the south of the wind stress maximum. Ignore the β contribution by assuming the channel is not very wide and by noting that β is smaller in the high latitudes (as relevant to the Southern Ocean).

EXERCISE 28.2: RELATING EKMAN, ROSSBY, AND REYNOLDS NUMBERS

Under certain scalings, we can connect the Ekman number to the Rossby and Reynolds numbers. In detail, we have

$$\text{Ro} = \frac{U}{f_0 L} \quad \text{and} \quad \text{Ek} = \frac{\nu^{\text{eddy}}}{f_0 H^2} \quad \text{and} \quad \text{Re} = \frac{W H}{\nu^{\text{eddy}}}, \quad (28.60)$$

where

$$H = \text{vertical length scale} \quad (28.61a)$$

$$L = \text{horizontal length scale} \quad (28.61b)$$

$$U = \text{horizontal velocity scale} \quad (28.61c)$$

$$W = \text{vertical velocity scale}. \quad (28.61d)$$

and we defined the Reynolds number (Section 19.8) in terms of the vertical viscosity, vertical length scale, and vertical velocity scale. What is the ratio of the Rossby number to Reynolds if the flow is non-divergent?

EXERCISE 28.3: EKMAN BOUNDARY LAYERS AND RAYLEIGH DRAG

Rayleigh drag described in Section 28.2.3 allowed us to study the effects of friction on fluid trajectories. However, Rayleigh drag has no spatial derivatives, which contrasts to the Laplacian friction from Section 28.3.1 and used elsewhere in this chapter. Comment on whether Rayleigh drag can support the no-slip boundary condition and the corresponding development of a boundary layer.



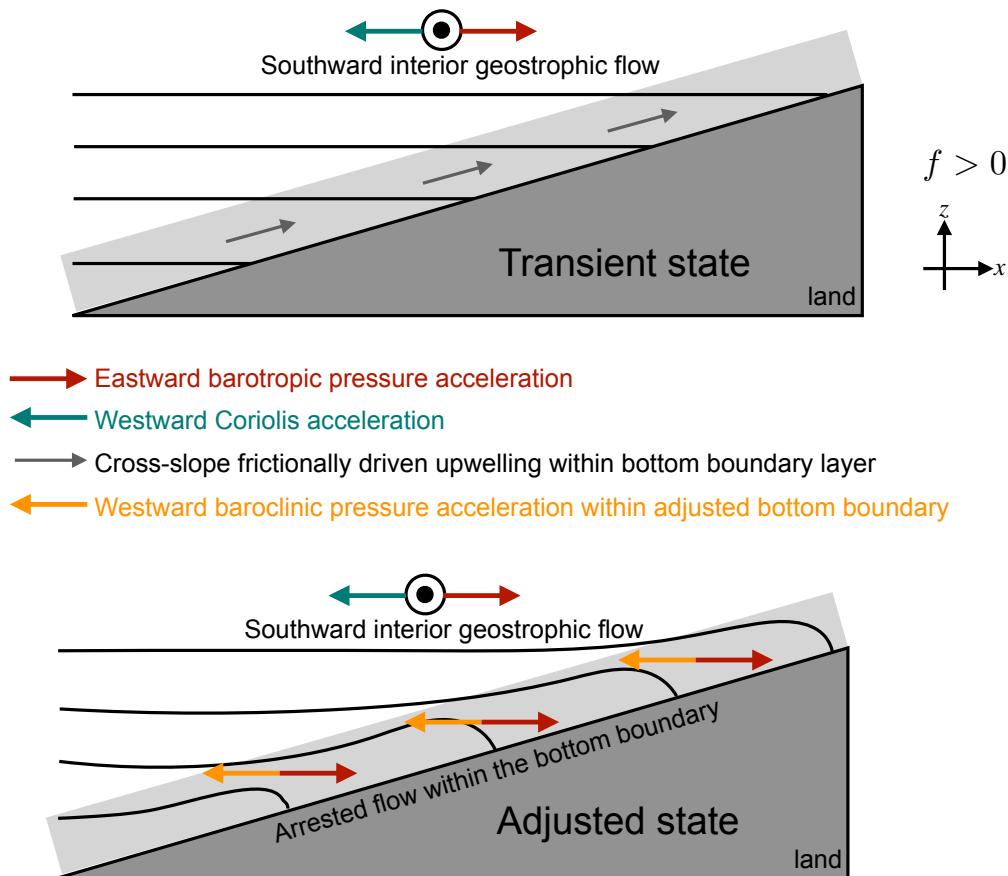


FIGURE 28.13: In both panels we depict a prescribed barotropic pressure acceleration that acts eastward and has a corresponding southward geostrophic flow and westward Coriolis acceleration. The solid lines depict lines of constant buoyancy, also known as isopycnals (see Chapter 25). Top panel: As the flow enters the bottom boundary layer it slows so that its Coriolis acceleration no longer balances the barotropic pressure, thus sending water cross slope in the down pressure gradient direction. Bottom panel: The cross-slope movement of density creates a baroclinic pressure acceleration that counteracts the barotropic pressure acceleration. The steady state is realized when the two pressure accelerations balance, in which case flow within the boundary layer halts. In this halted state there are no frictional stresses since the flow halts due to pressure effects rather than friction. Note that isopycnals are shown intersecting the solid boundary in a perpendicular direction, which is implied by the no-flux bottom boundary condition (in the absence of geothermal heating) as discussed in Section 16.5.2. Geostrophic flow in the opposite direction leads to downwelling along the bottom rather than upwelling. This figure is adapted from Figure 5 of [Wählin et al. \(2012\)](#).

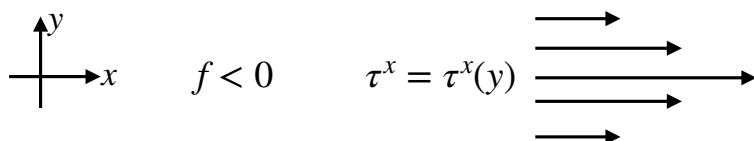


FIGURE 28.14: Zonal wind stress in a southern hemisphere channel with a maximum flanked by lower winds. What is the sense for the associated horizontal and vertical Ekman mass transport following from the discussion in Sections 28.4.1 and 28.4.4?

Space and time dependent gravity

We here formulate the dynamical equations for a geophysical fluid in the presence of a space and time dependent gravitational acceleration. This formulation has application to the study of astronomical tides in the ocean, thus motivating a discussion of the astronomical tidal forcing that follows the treatment given in Chapter 3 of [Pugh \(1987\)](#) and Section 5.15 of [Apel \(1987\)](#), with Chapter 2 of [Brown \(1999\)](#) and Section 17.4 of [Stewart \(2008\)](#) useful pedagogical supplements. Besides tides, a topic of interest to climate science concerns the study of how ocean sea level responds to changes in mass distributions associated with melting land ice. The nontrivial impact that melting land glaciers has on the earth's geoid and earth's rotation ([Farrell and Clark, 1976](#); [Mitrovica et al., 2001](#); [Kopp et al., 2010](#)) further motivates developing the dynamical equations of a liquid ocean in the presence of a space-time dependent gravity.

READER'S GUIDE TO THIS CHAPTER

This chapter assumes an understanding of the equations of motion derived in Chapter 18 as well as the gravitational and planetary centrifugal accelerations from Section 10.12. We dispense with tensor notation in this chapter, with subscripts used here as descriptive labels rather than tensor indices. No other chapter depends on the material in this chapter.

29.1	Gravitational potential	643
29.1.1	Simple geopotential	644
29.1.2	General geopotential	644
29.2	Momentum equation	644
29.3	Primitive equations	645
29.4	Depth independent perturbed geopotential	645
29.5	Forces contributing to ocean tides	646
29.5.1	Tidal acceleration in a spherically symmetric gravity field	646
29.5.2	Heuristics of tidal acceleration on the surface of a sphere	646
29.5.3	Gravitational potential for an idealized earth-moon system	649
29.5.4	Concerning realistic tides	652
29.5.5	Comments	653

29.1 Gravitational potential

In this section we summarize elements of the gravitational force, including the case with a non-constant gravitational acceleration such as occurs from astronomical tidal forcing and changes to the mass distribution of the planet.

29.1.1 Simple geopotential

As detailed in Section 10.12, the effective gravitational field incorporates the effects from the planetary centrifugal acceleration. The effective gravitational field is conservative, so that the gravitational acceleration of a fluid element can be represented as the gradient of a scalar (see Section 10.12.4),

$$\mathbf{g} = -\nabla \Phi, \quad (29.1)$$

with Φ the geopotential. In most applications of this book, the local vertical direction is denoted by

$$z = r - R_e, \quad (29.2)$$

with $z = 0$ the geopotential surface corresponding to a resting ocean and $R_e = 6.367 \times 10^6$ m the average radius of the earth (Section 10.2). The geopotential in this case is given by

$$\Phi = \Phi_0 = g z, \quad (29.3)$$

with $g \approx 9.8 \text{ m s}^{-2}$ the typical value used for the gravitational acceleration at the earth's surface.

29.1.2 General geopotential

Consider a generalized geopotential written in the form

$$\Phi = \Phi_0(r) + \Phi_1(r, \lambda, \phi, t), \quad (29.4)$$

where $\Phi_0(r)$ is the geopotential given by equation (29.3), and Φ_1 incorporates perturbations to the geopotential. For the study of ocean tides, the structure of Φ_1 arises from astronomical perturbations to the earth's gravity field. The calculation of ocean tides arising from astronomical forcing is formulated with a space-time dependent geopotential as in equation (29.4), with the radial dependence of Φ_1 neglected (e.g., Section 9.8 in [Gill, 1982](#)). [Arbic et al. \(2004\)](#) provide a discussion of global tide modelling.

Nontrivial Φ_1 variations also arise from perturbations in terrestrial masses, such as the melting of land ice such as that occurring on Greenland or Antarctica due to climate warming. These mass distribution changes lead to changes in the earth's gravitational field, its rotational moment of inertia, and the deformation of the crust (GRD as in [Gregory et al. \(2019\)](#)). Each of these effects lead to modifications in the *static equilibrium sea level*. In contrast to ocean tides, GRD perturbations associated with melting land ice are not periodic nor readily predictable. Furthermore, as evidenced by Figure 1 in [Mitrovica et al. \(2001\)](#), the amplitude of static equilibrium sea level changes can be far greater than typical open ocean tide fluctuations.

29.2 Momentum equation

As detailed in Section 18.1.3, the inviscid momentum equation for a rotating fluid in a gravitational field is given by

$$\rho \frac{D\mathbf{v}}{Dt} + 2\boldsymbol{\Omega} \wedge \rho \mathbf{v} = -\nabla p - \rho \nabla \Phi. \quad (29.5)$$

In writing the momentum equation in the form (29.5), we have chosen to retain an orientation afforded by the unperturbed geopotential, $\Phi_0(r)$, which are surfaces of constant z . This approach reflects that commonly used to study ocean tides. In the presence of a perturbed geopotential, Φ_1 , the “horizontal” directions defined by surfaces of constant z are no longer parallel to geopotential

surfaces. We thus may interpret the sum $\nabla_z p + \rho \nabla_z \Phi$ as an orientation of the pressure gradient along surfaces of constant geopotential, where the geopotential is determined by $\Phi = \Phi_0 + \Phi_1$, rather than just the unperturbed geopotential Φ_0 .

29.3 Primitive equations

As detailed in Section 23.1, the primitive equations reduce the vertical momentum equation to its static inviscid form, which is the hydrostatic balance

$$\frac{\partial p}{\partial z} = -\rho \frac{\partial \Phi}{\partial z} = -\rho(g + \partial_z \Phi_1). \quad (29.6)$$

The hydrostatic balance is modified from its traditional form for cases where the perturbation geopotential Φ_1 exhibits nontrivial depth dependence. Correspondingly, the horizontal momentum equation (making the Traditional Approximation from Section 23.1) takes the form

$$\rho \frac{D\mathbf{u}}{Dt} + \hat{\mathbf{z}} f \wedge \rho \mathbf{u} = -(\rho \nabla_z \Phi_1 + \nabla_z p) \quad (29.7)$$

where ∇_z is the horizontal gradient taken on surfaces of constant z . In their oceanic Boussinesq form (Chapter 24), the inviscid horizontal momentum equation becomes

$$\frac{D\mathbf{u}}{Dt} + \hat{\mathbf{z}} f \wedge \mathbf{u} = -(1/\rho_0)(\rho_0 \nabla_z \Phi_1 + \nabla_z p) \quad (29.8)$$

where ρ_0 is the constant reference density for a Boussinesq fluid. The Boussinesq form makes the addition of a perturbed geopotential quite straightforward, in which it is gradients in $\rho_0 \Phi_1 + p$ that take the place of gradients in pressure p .

29.4 Depth independent perturbed geopotential

A particularly simple form of Φ_1 occurs when it is depth independent,

$$\Phi_1 = \Phi_1(\lambda, \phi, t), \quad (29.9)$$

in which case the hydrostatic balance (29.6) returns to its traditional form $\partial_z p = -\rho g$. This form is motivated by the scale analysis in Section 29.5.3 where we find that the radial component of the earth's gravitational field greatly exceeds that from the moon or other celestial bodies, so that it is the lateral variation in the gravitational acceleration that drive tidal motions. In this case it is convenient to write the geopotential as

$$\Phi_1 = -g h, \quad (29.10)$$

with $h = h(\lambda, \phi, t)$ the perturbed geopotential height field. The full geopotential is thus written

$$\Phi = g(z - h), \quad (29.11)$$

with this form revealing that the zero of the geopotential is now set by $z = h$ rather than $z = 0$. In the study of ocean tides, h is referred to as the *equilibrium tide*. In geodesy, h is referred to as the *static equilibrium sea level*.

Since the perturbed geopotential is depth independent, it only affects the depth integrated horizontal momentum, and it does so through the term

$$-\int_{\eta_b}^{\eta} \nabla_z \Phi_1 dz = g \int_{\eta_b}^{\eta} \nabla_z h dz = g(-\eta_b + \eta) \nabla_z h. \quad (29.12)$$

Hence, modifications to the geopotential as embodied by the perturbed geopotential height field, $h = h(\lambda, \phi, t)$, are isolated to their impacts on the horizontal pressure gradients acting on the depth integrated horizontal momentum.

29.5 Forces contributing to ocean tides

We here describe the rudiments of forces that contribute to ocean tides as well as solid-earth tides. For simplicity we focus just on the earth-moon system, though note that the sun also plays an analogous role for observed tidal motion.

29.5.1 Tidal acceleration in a spherically symmetric gravity field

Before considering the earth-moon system, we introduce the notion of *tidal acceleration*, which arises on a finite sized body placed within a non-uniform gravitational field. Figure 29.1 depicts this situation where the finite sized body is a narrow rod whose axis points towards the center of a spherically symmetric massive body. One end of the rod experiences a different gravitational acceleration than the other since the gravitational field falls off as the inverse squared distance from the center of the sphere. It is this differential gravitational acceleration that we refer to as the tidal acceleration. As we will see, its key property is that the tidal acceleration falls off as the inverse cube of the distance rather than the more familiar inverse square.

To develop a mathematical expression for the tidal acceleration, focus on the spherically symmetric gravitational field in which the gravitational acceleration at a point is given by (Section 10.12.2)

$$\mathbf{g} = -\frac{GM}{r^2} \hat{\mathbf{r}}, \quad (29.13)$$

where r is the distance from the sphere's center, G is Newton's gravitational constant, M is the mass of the sphere, and $\hat{\mathbf{r}}$ is the radial unit vector. The minus sign indicates that the gravitational acceleration points toward the center of the sphere. For the rod in Figure 29.1, the difference between the gravitational acceleration acting at a point nearest to the sphere (point B) and a point furthest from the sphere (point A) is given by

$$\mathbf{g}(r_B) - \mathbf{g}(r_A) = \mathbf{g}(r_0 - L/2) - \mathbf{g}(r_0 + L/2), \quad (29.14)$$

where r_0 is the distance from the sphere's center to the center of the rod. Assuming the rod is not long, we can expand this difference in a Taylor series about the rod center at r_0 , thus leading to an expression for the tidal acceleration

$$\mathbf{g}(r_B) - \mathbf{g}(r_A) \approx -L \frac{\partial \mathbf{g}}{\partial r} = -2L \frac{GM}{r_0^3} \hat{\mathbf{r}} = (2L/r_0) \mathbf{g}(r_0). \quad (29.15)$$

The key point to conclude from this example is that the tidal acceleration is proportional to the inverse cube of the distance to the center of the sphere. We see this property again when considering in Section 29.5.3 the gravitational acceleration generated from a remote body (e.g., the moon) acting on the surface of a sphere (e.g., the earth).

29.5.2 Heuristics of tidal acceleration on the surface of a sphere

We now consider the tidal acceleration acting on the surface of a smooth massive sphere due to a spherically symmetric gravitational field generated by a neighboring massive body. Figure

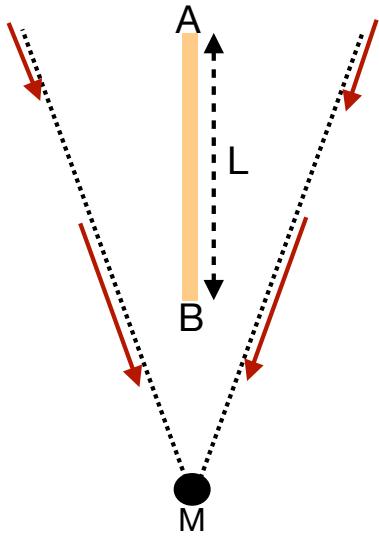


FIGURE 29.1: *Tidal acceleration* is the acceleration that acts on a finite sized object placed in a non-uniform gravitational field. The finite object is here depicted as a narrow rod of length L placed in the gravity field of a spherically symmetric body of mass M . That portion of the rod closer to the gravitating sphere (end B) experiences a stronger gravitational acceleration than the end that is further away (end A). The gradient in the gravitational acceleration constitutes the tidal acceleration acting on the rod.

29.2 depicts this system, which we consider an idealized earth-moon system where each body is assumed homogeneous and spherical. Given that they gravitationally attract one another, it is not astronomically possible for the two bodies to remain spatially fixed. Instead, they orbit around their common center of mass while conserving their angular momentum.

A central question of tidal studies is why there are generally two ocean tides per day (semi-diurnal tides) rather than just one (diurnal tides). We here offer two complementary arguments. The first is based on extending the tidal acceleration discussion of Section 29.5.1, whereas the second follows the more traditional account by considering a balance between gravitational and centrifugal accelerations.

General ideas

Every point on the surface of the earth is attracted to the earth's center by the earth's gravitational field. For a spherical earth, this attractive force is purely radial, so that it cannot lead to lateral motion on the surface of the perfect sphere. We thus conclude that the radial gravitational field is not the cause of tidal motion. Instead, tidal motion arises from a non-radial gravitational field.

The earth-moon gravitational field accelerates the earth and moon toward one another along the axis connecting their centers. Additionally, the spatial dependence of the moon's gravitational field over the earth leads to lateral forces along the earth's surface, thus providing the ingredient for ocean tidal motion. To capture the essence of this force, we examine how the moon's gravitational field acts on a point on the earth relative to its action at the center of the earth.

Sample tidal accelerations on the sphere

Again, we are tasked with computing the tidal acceleration from the moon's gravitational field for selected points on the earth, and we are computing these accelerations relative to the earth center. As for the rod in Figure 29.1, the tidal acceleration at point B relative to the center of the earth

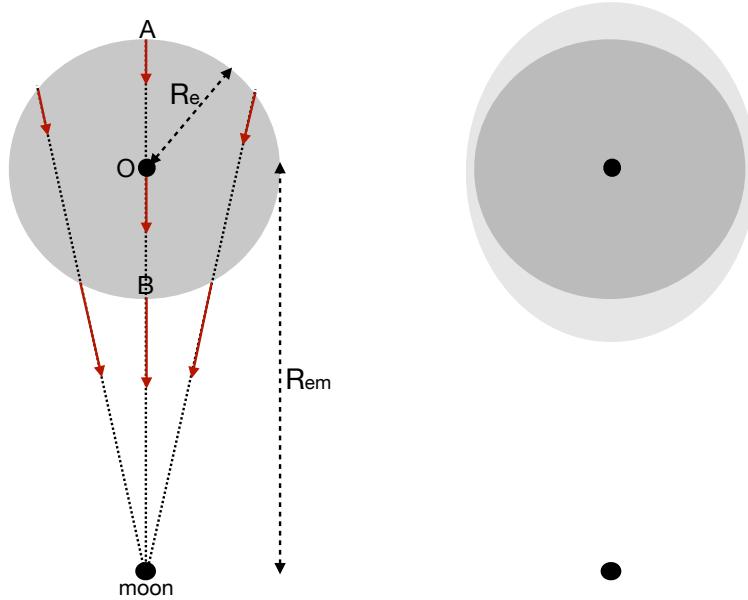


FIGURE 29.2: Illustrating the tidal force on the surface of a sphere. The sphere is an ideal depiction of the earth and the smaller massive object is the moon. The distance between the center of the earth and moon is R_{em} , and the radius of the earth is R_e . The left panel shows representative moon-generated gravitational field lines. Two points along these field lines on the surface of the earth represent the two ends of an imaginary rod as depicted in Figure 29.1. The tidal acceleration acting at point B, relative to the earth's center, points toward the moon (equation (29.16)). In contrast, the tidal acceleration at point A, relative to the earth's center, points in the opposite direction (equation (29.17)). Points on the earth surface between A and B have tidal accelerations with a non-zero component directed along the surface of the earth. Symmetry of the configuration allows us to conclude that a layer of water on the surface of the sphere will accumulate to produce two bulges as shown in the right panel. It is the lateral component of the gravitational acceleration that causes the water to accumulate to produce tidal bulges at points A and B. In contrast, the radial component to the moon's gravitational field has no contribution to the tides. Note that as shown in Section 29.5.3, the bulge shown in the right panel is greatly exaggerated.

is given by

$$\mathbf{g}(r_B) - \mathbf{g}(R_{\text{em}}) = (2R_e/R_{\text{em}}) \mathbf{g}(R_{\text{em}}). \quad (29.16)$$

This acceleration points towards the moon. In contrast, the tidal acceleration at point A relative to the center of the earth is given by

$$\mathbf{g}(r_A) - \mathbf{g}(R_{\text{em}}) = -(2R_e/R_{\text{em}}) \mathbf{g}(R_{\text{em}}), \quad (29.17)$$

which is of equal magnitude but points away from the moon.

The tidal accelerations at points A and B act radially away from the earth's center. Hence, as noted above, these radial forces do not directly lead to tidal motion at those points. However, through symmetry of the configuration, points on the surface of the sphere between A and B have a tidal acceleration from the moon's gravitational field with a nonzero lateral component. These lateral forces lead to the accumulation of water at points A and B. We can compute the gravitational acceleration at intermediate points. However, the trigonometry is somewhat complex and we prefer to compute the forces in Section 29.5.3 through use of the gravitational potential. For the current discussion we appeal to symmetry to conclude that the lateral tidal accelerations act to pile up water at both points A and B as depicted in the second panel of Figure 29.2. This argument, though heuristic, provides the means to understand how a water covered spherical planet has two bulges, rather than one, due to spatial gradients in the moon's gravitational field. We confirm this argument in Section 29.5.3 by explicitly computing the gravitational potential for this idealized

earth-moon system and then taking the gradient to compute the gravitational acceleration (see Figure 29.4).

Including orbital motion

Thus far we have ignored the orbital motion of the earth-moon system around their common center of mass. As we will see, there are no fundamental changes to the above arguments when allowing for orbital motion.

In the absence of dissipation, as assumed here, the earth-moon distance remains constant due to their angular momentum conserving orbital motion. From a force-balance perspective, the two spherical bodies remain in a fixed orbit since the gravitational acceleration acting at their centers is balanced by their respective centrifugal accelerations, where the centrifugal acceleration is computed relative to the center of mass of the two-body system. The gravitational acceleration from the moon, acting at the center of the earth, is given by the *free fall* value $\mathbf{g}(R_{\text{em}})$, which has magnitude $GM_{\text{m}}/R_{\text{em}}^2$ and is directed along the axis connecting the earth and moon centers.

Furthermore, when a body exhibits orbital motion, each point on the body exhibits the same orbital motion and has the same linear velocity. Consequently, each point on the earth possess the same centrifugal acceleration

$$\mathbf{a}_{\text{orbital centrifugal}} = -\mathbf{g}(R_{\text{em}}). \quad (29.18)$$

This property of orbital motion is distinct from the spinning motion of a planet rotating about its axis, whereby points further from the rotational axis have larger centrifugal acceleration (see Section 10.12). To help understand orbital motion, move your hand in a circle while maintaining the arm in a single direction so that the hand exhibits an orbital motion rather than a spinning motion. Notice that all parts of the hand move with the same linear velocity and exhibit the same orbital motion. Hence, each point on the hand has the same centrifugal acceleration.

We can now ask about the acceleration felt by a point on the surface of the earth. The acceleration giving rise to tidal motions is the sum of the gravitational acceleration from the moon plus the centrifugal acceleration due to orbital motion. However, this calculation is identical to that considered previously, which led, for example, to the tidal accelerations for points *B* and *A* as given by equations (29.16) and (29.17). We are thus led to the same result as before.

29.5.3 Gravitational potential for an idealized earth-moon system

We now perform a more thorough calculation of the gravitational acceleration by computing the gradient of the gravitational potential. First recall the discussion of Newton's gravitational law in Section 10.12.2, whereby the gravitational potential for a point at distance r from the center of a spherical earth is given by

$$\Phi_{\text{e}}(r) = -\frac{GM_{\text{e}}}{r}, \quad (29.19)$$

where M_{e} is the mass of the earth. The corresponding radial gravitational acceleration is given by

$$\mathbf{g}_{\text{e}} = -\nabla\Phi_{\text{e}} = -\frac{GM_{\text{e}}\hat{\mathbf{r}}}{r^2}. \quad (29.20)$$

The same considerations hold for the moon's gravitational potential. Hence, referring to Figure 29.3, the moon's gravitational potential evaluated at a distance L from the moon's center is given by

$$\Phi_{\text{m}}(L) = -\frac{GM_{\text{m}}}{L}. \quad (29.21)$$

Trigonometry leads to the law of cosines relation

$$L^2 = (R_{\text{em}} - r \cos \psi)^2 + (r \sin \psi)^2 = R_{\text{em}}^2 + r^2 - 2r R_{\text{em}} \cos \psi, \quad (29.22)$$

where again r is the distance to the earth's center and ψ is the polar angle relative to the \hat{x} axis pointing between the earth and moon centers (see Figure 29.3).

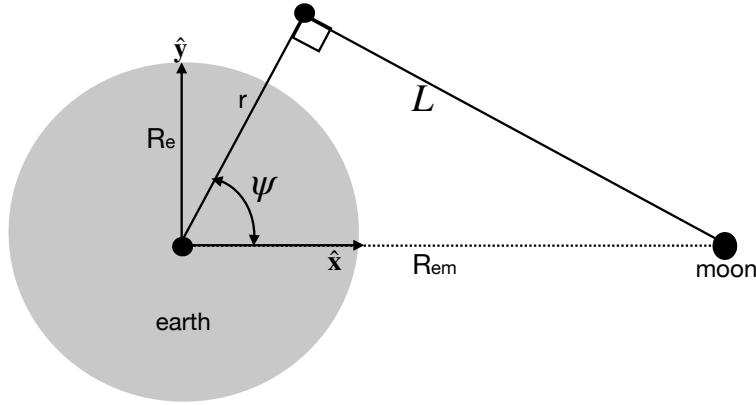


FIGURE 29.3: Geometry of an idealized earth-moon system. The center of the earth is a distance R_{em} from the center of the moon; the moon has a mass M_m ; and the earth has a radius R_e . An arbitrary test point is shown a distance L from the center of the moon, r from the center of the earth, and with a polar angle ψ relative to the \hat{x} axis, where the \hat{x} axis points from the earth center to the moon center. Relative to the earth's center, the test point has Cartesian coordinates $(x, y) = r(\cos \psi, \sin \psi)$. See Section 8.3 for details on relating polar and Cartesian coordinates.

Identifying the leading order contributions

Assuming the test point in Figure 29.3 is closer to the earth than to the moon, we can perform a Taylor series expansion in the small parameter r/R_{em} to render

$$\Phi_m(L) = -\frac{GM_m}{L} = -\frac{GM_m}{R_{\text{em}}} \left[1 + \frac{r \cos \psi}{R_{\text{em}}} + \frac{r^2}{2R_{\text{em}}^2} (3 \cos^2 \psi - 1) + \mathcal{O}(r/R_{\text{em}})^3 \right]. \quad (29.23)$$

We thus identify the leading three terms to the geopotential

$$\Phi_m^{(0)} = -\frac{GM_m}{R_{\text{em}}} \quad (29.24)$$

$$\Phi_m^{(1)} = -\frac{GM_m}{R_{\text{em}}^2} r \cos \psi \quad (29.25)$$

$$\Phi_m^{(2)} = -\frac{GM_m}{2R_{\text{em}}^3} r^2 (3 \cos^2 \psi - 1). \quad (29.26)$$

Assuming the distance between the earth and moon remains fixed, the zeroth order term $\Phi_m^{(0)}$ is a spatial constant and thus leads to no gravitational acceleration. We now examine the gravitational accelerations from the other two terms.

Acceleration maintaining the orbiting earth-moon system

For the first order term, $\Phi_m^{(1)}$, we introduce the Cartesian coordinate as in Figure 29.3 to write

$$\Phi_m^{(1)} = -\frac{GM_m x}{R_{\text{em}}^2}, \quad (29.27)$$

where $x = r \cos \psi$ is the distance along \hat{x} . Hence, the gradient of $\Phi_m^{(1)}$ leads to the gravitational acceleration

$$\mathbf{g}_m^{(1)} = -\nabla \Phi_m^{(1)} = \hat{x} \frac{GM_m}{R_{em}^2}. \quad (29.28)$$

This gravitational acceleration has a constant magnitude at every point in space and it everywhere points in a direction parallel to the earth-moon axis. Furthermore, the magnitude of $\mathbf{g}_m^{(1)}$ equals to that of the moon's gravitational acceleration, \mathbf{g}_m , when evaluated at the earth's center. As seen in Section 29.5.2, the acceleration $\mathbf{g}_m^{(1)}$ maintains the earth in orbit about the center of mass for the earth-moon system; i.e., this is the free fall acceleration towards the moon. Notably, at the earth's surface, the magnitude of $\mathbf{g}_m^{(1)}$ is tiny relative to the gravitational acceleration from the earth itself, with their ratios given by

$$\frac{M_m/R_{em}^2}{M_e/R_e^2} \approx 3.4 \times 10^{-6}, \quad (29.29)$$

where we set

$$M_e = 5.97 \times 10^{24} \text{ kg} \quad M_m = 7.35 \times 10^{22} \text{ kg} = (1/81.2) M_e \quad (29.30a)$$

$$R_e = 6.367 \times 10^6 \text{ m} \quad R_{em} = 3.84 \times 10^8 \text{ m} = 60.3 R_e. \quad (29.30b)$$

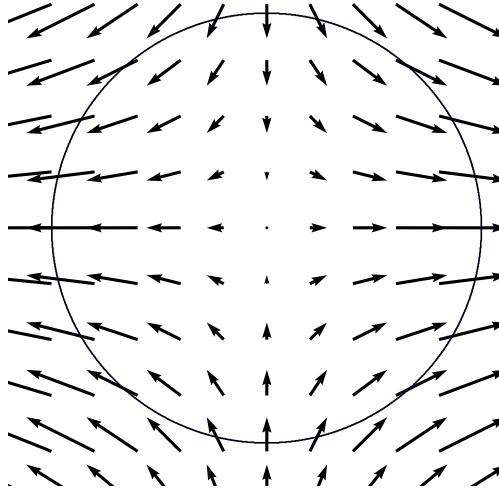


FIGURE 29.4: The tide producing gravitational acceleration $\mathbf{g}_m^{(2)}$ given by equation (29.34). The moon is assumed to be positioned in the equatorial plane of the earth.

Tide producing geopotential

The main tide producing acceleration results from $\Phi_m^{(2)}$. Introducing the second Cartesian coordinate, $y = r \sin \psi$, leads to

$$\Phi_m^{(2)} = -\frac{GM_m}{2R_{em}^3} r^2 (3 \cos^2 \psi - 1) = -\frac{GM_m}{2R_{em}^3} (2x^2 - y^2). \quad (29.31)$$

The corresponding perturbed geopotential height field (see equation (29.11)) is given by

$$h = -\frac{\Phi_m^{(2)}}{g} = \frac{R_e^2}{2R_{em}^3} \frac{M_m}{M_e} r^2 (3 \cos^2 \psi - 1). \quad (29.32)$$

Placing the test point on the earth surface, $r = R_e$, renders

$$h = \frac{R_e^4}{2 R_{\text{em}}^3} \frac{M_m}{M_e} (3 \cos^2 \psi - 1) \approx 2.8 \times 10^{-8} R_e (3 \cos^2 \psi - 1). \quad (29.33)$$

Plugging in numbers for the earth-moon system suggests that the maximum perturbation to the geopotential height arising from the moon's gravity field is roughly 36 cm. Correspondingly, the bulge shown in Figure 29.2 is greatly exaggerated. Note that ocean tidal amplitudes can get much larger (order meters) than this "equilibrium tide" amplitude due to resonances from ocean geometry, with the Bay of Fundy in Nova Scotia a particularly striking example.

Tide producing acceleration

The gravitational acceleration arising from the tidal potential is determined by the gradient of the tidal geopotential

$$\mathbf{g}_m^{(2)} = -\nabla \Phi_m^{(2)} = \frac{GM_m}{R_{\text{em}}^3} (2x \hat{\mathbf{x}} - y \hat{\mathbf{y}}). \quad (29.34)$$

We illustrate the vector field $\mathbf{g}_m^{(2)}$ in Figure 29.4. Note how the accelerations lead to two bulges on opposite sides of the planet. We can write this acceleration using polar coordinates by introducing the polar unit vectors $\hat{\mathbf{r}}$ and $\hat{\psi}$ according to Section 8.3.2

$$\hat{\mathbf{r}} = \hat{\mathbf{x}} \cos \psi + \hat{\mathbf{y}} \sin \psi \quad (29.35a)$$

$$\hat{\psi} = -\hat{\mathbf{x}} \sin \psi + \hat{\mathbf{y}} \cos \psi \quad (29.35b)$$

thus rendering

$$\mathbf{g}_m^{(2)} = \frac{GM_m R_e}{R_{\text{em}}^3} \left[\hat{\mathbf{r}} (3 \cos^2 \psi - 1) - (3/2) \hat{\psi} \sin 2\psi \right], \quad (29.36)$$

where we evaluated the acceleration at the earth surface so that $r = R_e$. Evaluating the acceleration at $\psi = 0, \pi$ verifies the heuristic calculation performed in Section 29.5.2 for points on the earth surface nearest and furthest from the moon. We can further gauge the magnitude of the tidal acceleration by introducing the acceleration due to the earth's gravity field

$$\mathbf{g}_m^{(2)} = g_e \frac{M_m}{M_e} \frac{R_e^3}{R_{\text{em}}^3} \left[\hat{\mathbf{r}} (3 \cos^2 \psi - 1) - (3/2) \hat{\psi} \sin 2\psi \right], \quad (29.37)$$

where $g_e = GM_e/R_e^2$ is the acceleration at the earth's surface from the earth's gravity field. The dimensional prefactor has magnitude $\approx 5.6 \times 10^{-8} g_e$, so that the tidal acceleration is tiny relative to that from the earth's gravity field. It is for this reason that the radial component of the tidal acceleration is largely irrelevant since it is dominated by the far larger radial component of the earth's gravity field. However, the angular component of the tidal acceleration, although small relative to the earth's radial gravitational acceleration, is able to move water along the surface of the planet as indicated by Figure 29.4, thus leading to tidal motion.

29.5.4 Concerning realistic tides

Our discussion of tides has been rather terse, aiming to identify key aspects of the tidal accelerations but giving little attention to details that impact real ocean tides. Here are a few points that must be considered for these purposes.

- As the earth spins under the tidal bulges, there are two high and two low tides per day. Additional orbital motion of the moon adds roughly 50 minutes per day to the diurnal (daily) tide and 25 minutes to the semi-diurnal (twice daily).
- The moon orbits the earth at a latitude of roughly $28.5^\circ N$ rather than within the equatorial plane, so that the tidal bulges are offset from the equator. As the earth spins under the bulges, one of the high tides is generally larger than the other due to the offset. This offset in turn introduces a diurnal component to the tides in addition to the semi-diurnal.
- The sun contributes to tides in a manner similar to the moon. The sun is more massive than the moon, yet it is further away, so that the ratio of the magnitudes for the tidal producing accelerations is given by

$$\frac{\text{moon tidal acceleration}}{\text{sun tidal acceleration}} = \frac{M_m/R_{em}^3}{M_s/R_{es}^3} \approx 2.2 \quad (29.38)$$

where we set

$$M_s = 1.99 \times 10^{30} \text{ kg} \quad R_{es} = 23460 R_e. \quad (29.39)$$

Hence, the moon has an impact on tides that is somewhat more than double that of the sun.

- The gravitational acceleration that leads to the tidal bulge moves around the mid-latitudes at roughly 330 m s^{-1} , which is faster than the $\approx 200 \text{ m s}^{-1}$ wave speed for shallow water gravity waves. Hence, the ocean tidal motion is never equilibrated to the *equilibrium tides* defined by the tidal acceleration. In contrast, solid-earth waves are much faster and so the solid-earth tidal motions are mostly equilibrated with the equilibrium tidal acceleration. Solid-earth tides have an amplitude on the order of 10 cm with wavelengths spanning the planet. Hence, an accurate treatment of ocean tides must take into account the solid-earth tides.
- The movement of ocean mass modifies the earth's gravity field, and this modification is referred to as *self-attraction*. Additionally, movement of the ocean mass alters the *loading* felt by the solid-earth and thus causes the crust to compress and expand. These two terms are referred to as the *self attraction and loading* (SAL) terms.
- Geometry of the ocean plays a leading role in determining tides at a particular location. We have incomplete information about the geometry of ocean basins, such as the presence of deep ridges, troughs, and seamounts. In lieu of such information we can garner useful information based on the analysis of past tides, with that information used to fit sinusoidal waves to the measured time series for use in projecting forward in time.

29.5.5 Comments

A key feature of the tidal producing forces is that it is the lateral (along-earth) component of the moon's tidal gravitational force that produces the earth's tides. These lateral forces cause water to accumulate at the point nearest to and furthest from the moon (points *A* and *B* in Figure 29.2), thus producing the characteristic double-bulge pattern. Notably, many common literature presentations make it appear that it is the radial (i.e., pointing to the earth's center) component of the moon's gravitational force, and its gradient across the earth, that leads to the earth's tidal bulges. But as discussed in Section 29.5.2, radial gravitational forces cannot lead to tidal motions; what is needed is a force that leads to lateral motion. These key notions are nicely emphasized in [this Space Time video](#).



Surface gravity waves and Stokes drift

In this chapter we consider the ocean surface to be a material interface separating the liquid ocean from the atmosphere above, and we examine its small amplitude fluctuations. These fluctuations manifest as linear surface gravity waves whose existence relies on the restoring force from gravity. In formulating the equations for surface gravity waves, we bring together a number of fluid mechanical principles considered earlier in this book so that this analysis supports our understanding of how basic principles are used to describe motion. Notably, we here ignore rotation, instead focusing on ocean surface waves whose lateral extent is too short to be affected much by the Coriolis force.

Our treatment of surface waves is relatively brief, offering only sufficient details to expose their characteristic exponential decay with depth when moving into the ocean interior. This depth decay leads to a general class of fluid particle motion known as *Stokes drift*, where Stokes drift arises for all waves whose amplitude changes in those directions not parallel to the wave phase direction. So although the surface gravity waves are linear, the depth decay in their amplitude leads to a net drift motion of fluid particles and hence to the transport of matter. The Stokes drift exemplified by surface gravity waves provides an example of how averaging at a fixed point in space (Eulerian average) yields distinct behaviors from averaging on a fixed fluid particle (Lagrangian average). The kinematics of Stokes drift has wide application throughout geophysical fluid mechanics.

READER'S GUIDE TO THIS CHAPTER

In deriving the basic equations of surface gravity waves we make use of dynamical ideas from Chapter 18 and elements of the filtered equations from Chapter 23. We also make use of ideas from partial differential equations introduced in Chapter 3. The mathematical description of Stokes drift requires an understanding of Eulerian and Lagrangian kinematic descriptions from Chapter 13. Generalizations of Stokes drift appear in Chapter 49 in our study of wave-mean flow interactions, isopycnal averaging, and the corresponding eddy-induced tracer transport. Finally, we ignore surface tension and their associated capillary waves since our interests concern wave motions with far larger wavelengths than capillary waves (a few centimeters; see Section 19.10.3).

30.1	Surface gravity waves	656
30.1.1	Harmonic scalar potential	656
30.1.2	Equation of motion and Bernoulli's Principle	656
30.1.3	Non-hydrostatic pressure and the shallow water limit	658
30.1.4	Dynamic boundary condition at the free surface	659
30.1.5	Kinematic boundary conditions	659
30.1.6	Summary of the linear equation set	659
30.2	Surface gravity waves in a flat bottom channel	660

30.2.1	Monochromatic wave solution	660
30.2.2	Dispersion relation for surface gravity waves	661
30.2.3	Heuristics of waves in the shortwave/deep water limit	662
30.2.4	Heuristics of waves in the longwave/shallow water limit	662
30.2.5	Further study	662
30.3	Stokes drift	663
30.3.1	General formulation of Stokes drift	663
30.3.2	Stokes drift in surface gravity waves	665
30.3.3	Comments and further study	669
30.4	Exercises	669

30.1 Surface gravity waves

We study linear fluctuations of the ocean free surface on a non-rotating plane, with the free surface assumed to be a material interface separating a homogeneous ocean from a homogeneous atmosphere (Figure 30.1). The mass of the atmosphere is horizontally uniform and static so that surface atmosphere pressure does not contribute to oceanic motion. We develop the boundary value problem describing linear wave motions of the free surface, and characterize physical aspects of the waves.

30.1.1 Harmonic scalar potential

We are interested in fluctuations that have zero vorticity so that the velocity field has zero curl¹

$$\nabla \wedge \mathbf{v} = 0. \quad (30.1)$$

Recall from Section 2.3.2 that vector fields with zero curl can be determined from a scalar potential

$$\mathbf{v} = -\nabla\psi. \quad (30.2)$$

Note that the scalar potential is unspecified up to an arbitrary function of time, since ψ and $\psi + F(t)$ yield the same velocity field. We make use of this *gauge* freedom in Section 30.1.2.

Since the ocean is assumed to have uniform density, the velocity field is non-divergent. Consequently, the scalar potential satisfies Laplace's equation (Section 2.3.3)

$$\nabla \cdot \mathbf{v} = -\nabla \cdot \nabla\psi = \nabla^2\psi = 0. \quad (30.3)$$

ψ is termed a *harmonic function*, with salient properties noted in Section 3.5. To fully specify the scalar potential requires boundary conditions, which enter our development via the equation of motion.

30.1.2 Equation of motion and Bernoulli's Principle

The vector-invariant equation of motion (18.32) for a non-rotating, irrotational, inviscid, uniform density fluid is given by

$$\frac{\partial \mathbf{v}}{\partial t} = -\nabla(\Phi + \mathcal{K} + p/\rho), \quad (30.4)$$

with

$$\mathcal{K} = \frac{\mathbf{v} \cdot \mathbf{v}}{2} \quad (30.5)$$

¹We study fluids with nonzero vorticity in Part VI of this book.

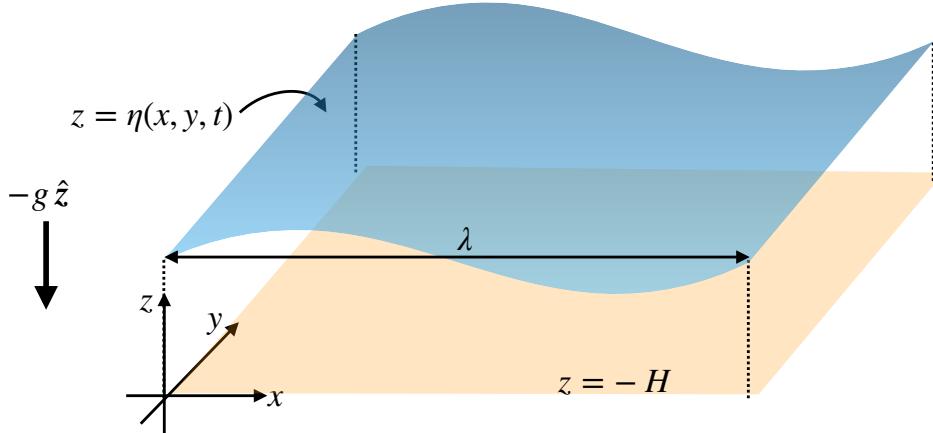


FIGURE 30.1: A depiction of the ocean free surface as a mathematical surface whose linear fluctuations exhibit gravity wave motion due to the restoring effects from a uniform gravitational field, $\mathbf{g} = -g\hat{\mathbf{z}}$. The atmosphere applies a pressure to the ocean due to its mass; however, that mass is assumed to be uniform and static so that it does not affect surface motion. Our formulation of surface waves makes no restrictions on the depth of the fluid relative to the wavelength of the waves, with the resulting waves exhibiting dispersion (i.e., the wave speed is a function of the wavelength). However, assuming the depth to be far smaller than the wavelength, $H \ll \lambda$ recovers the dispersion relation for linear non-dispersive shallow water gravity waves discussed in Section 32.8.3.

the kinetic energy per mass of a fluid element. In the following we assume the simple form of the geopotential (Section 10.12.4)

$$\Phi = g z \quad (30.6)$$

with g the constant gravitational acceleration. Inserting the scalar potential, $\mathbf{v} = -\nabla\psi$, brings the equation of motion (30.4) to the form

$$\nabla(\Phi + \mathcal{K} + p/\rho - \partial_t\psi) = 0. \quad (30.7)$$

This equation means that everywhere in the fluid the dynamical fields satisfy

$$\Phi + \mathcal{K} + p/\rho - \partial_t\psi = C(t), \quad (30.8)$$

for some arbitrary time dependent function $C(t)$. This equation is a particular expression of Bernoulli's theorem encountered in Section 22.6.2.

Removing the time dependent function via a gauge transformation

We ascribe no physical meaning to the arbitrary function $C(t)$ appearing in equation (30.8). In fact, it can be completely removed by exploiting the gauge degree of freedom in the scalar potential as noted following equation (30.2). We do so by introducing a modified scalar potential

$$\Psi(\mathbf{x}, t) = \psi(\mathbf{x}, t) + \int^t C(t') dt'. \quad (30.9)$$

Both ψ and Ψ lead to the same velocity vector

$$\mathbf{v} = -\nabla\psi = -\nabla\Psi \quad (30.10)$$

and as such they are physically indistinguishable. However, Ψ has the mathematical advantage of absorbing the arbitrary time dependent function $C(t)$, thus rendering the simpler expression for the equation of motion

$$\Phi + \mathcal{K} + p/\rho - \partial_t\Psi = 0. \quad (30.11)$$

This equation provides a balance condition that constrains the flow properties at every point within the fluid.

A brief diversion into Bernouilli's Principle for steady flow

The central focus of this analysis is with transient fluctuations of the free surface. Even so, we here comment on the case of steady flow, in which the constraint (30.11) leads to

$$\Phi + \mathcal{K} + p/\rho = 0. \quad (30.12)$$

Furthermore, assuming motion occurs on a constant geopotential then leads to

$$p + \rho \mathcal{K} = \text{constant}. \quad (30.13)$$

Motion on a geopotential means that all fluid particles feel the same gravitational acceleration. Hence, spatial changes in pressure provide the only force that can lead to spatial changes in fluid speed. That is the physical content of this special form of Bernouilli's Principle given by equation (30.13). It says that in regions of relatively high fluid speed the pressure is relatively low, and vice versa. Correspondingly, if the fluid moves into a region of relatively high pressure then it must slow down, whereas the speed increases when moving to a region of low pressure. This principle helps to explain the basic features of steady state flow around obstacles such as airplane wings.

30.1.3 Non-hydrostatic pressure and the shallow water limit

The fluid is non-hydrostatic since

$$\frac{\partial p}{\partial z} = -\rho \frac{\partial \Phi}{\partial z} + \frac{\partial(\partial_t \Psi - \mathcal{K})}{\partial z} = \frac{\partial p_{\text{hydrostatic}}}{\partial z} + \frac{\partial p_{\text{non-hydrostatic}}}{\partial z}. \quad (30.14)$$

Depth variations in the kinetic energy and depth-time variations in the velocity potential lead to deviations from hydrostatic balance. We do not generally expect the flow to be hydrostatically balanced for two reasons: (i) the fluid layer has a uniform density so there is no stratification to suppress vertical accelerations that contribute to non-hydrostatic pressures; (ii) the fluid is nonrotating and so there is no vertical stiffening via the Taylor-Proudman result (Section 26.5.3), with vertical stiffening acting to suppress vertical accelerations that cause deviations from hydrostatic balance.

For the shallow water model we also consider a homogenous density layer. However, as shown in Chapter 31, the hydrostatic balance is fundamental to the dynamics. In particular, in Section 31.2 we see that the hydrostatic balance over a single homogeneous layer leads to horizontal motion that is depth independent throughout the layer. Hence, $\partial_z(\partial_t \Psi) = 0$ and the kinetic energy in the horizontal motions is depth independent, $\partial_z(u^2 + v^2) = 0$. Furthermore, the vertical motion has a linear depth dependence across the shallow water layer (Section 31.2.7) and its magnitude is far smaller than horizontal motions. Therefore, we can drop all contributions to $\partial_z \mathcal{K}$ for the shallow water layer, in which case equation (30.14) reduces to the hydrostatic limit

$$\frac{\partial p}{\partial z} = -\rho \frac{\partial \Phi}{\partial z} = -\rho g \quad \text{hydrostatic limit.} \quad (30.15)$$

So in summary, a homogeneous layer of fluid can have a depth dependence to its horizontal flow. But that depth dependence is driven only through non-hydrostatic pressure forces. The reason is the hydrostatic pressure has a depth-independent horizontal gradient within a homogeneous layer

$$\partial_z (\nabla_z p_{\text{hydrostatic}}) = -g \nabla_z \rho = 0. \quad (30.16)$$

30.1.4 Dynamic boundary condition at the free surface

Return now to the time dependent expression of the equation of motion (30.11). It is a rather remarkable result that applies to any point within the fluid and at any time. In particular, it applies at the free surface, $z = \eta(x, y, t)$ where pressure equals to the atmospheric pressure. As stated earlier, we assume that the atmospheric pressure is constant in space and time so that

$$g\eta + \mathcal{K} - \partial_t \Psi = -p_a/\rho = \text{constant}. \quad (30.17)$$

Without loss of generality we can set this constant to zero.² We are thus left with the boundary condition

$$g\eta + \mathcal{K} - \partial_t \Psi = 0 \quad \text{at } z = \eta. \quad (30.18)$$

At this point we linearize relative to a state of rest with $\eta = 0$, $\mathbf{v} = 0$, and $\partial_t \Psi = 0$. Linear fluctuations about this rest state have small velocities and as such the kinetic energy, which is second order in velocity, can be dropped to leave the linearized boundary condition

$$g\eta = \frac{\partial \Psi}{\partial t} \quad \text{linearized dynamic b.c. at } z = \eta. \quad (30.19)$$

This boundary condition directly connects the free surface to time tendencies of the velocity potential. The free surface rises when the velocity potential has a positive tendency, and vice versa.

30.1.5 Kinematic boundary conditions

The free surface is assumed to be a material interface, meaning that we ignore effects from matter transport across this surface. Consequently, following the discussion of kinematic boundary conditions in Section 15.4.2, we have

$$\frac{\partial \eta}{\partial t} + \mathbf{u} \cdot \nabla \eta = w \quad \text{at } z = \eta. \quad (30.20)$$

Linearizing this boundary condition about the state of rest, and introducing the scalar potential, leads to

$$\frac{\partial \eta}{\partial t} = -\frac{\partial \Psi}{\partial z} \quad \text{linearized kinematic b.c. at } z = \eta. \quad (30.21)$$

This is yet another constraint that links the free surface to the velocity potential.

One final kinematic boundary condition applies just to the velocity potential when it intersects with the rigid solid boundaries. Namely, the no normal flow condition from Section 15.4.1 means that

$$\hat{\mathbf{n}} \cdot \nabla \Psi = 0 \quad \mathbf{x} \text{ on rigid solid boundaries}, \quad (30.22)$$

where $\hat{\mathbf{n}}$ is the outward normal on the solid boundaries.

30.1.6 Summary of the linear equation set

The boundary value problem for the velocity potential and free surface is given by

$$\nabla^2 \Psi = 0 \quad \text{irrotational and non-divergent velocity throughout domain} \quad (30.23a)$$

$$\frac{\partial \Psi}{\partial t} = g\eta \quad \text{linearized dynamic b.c. at } z = \eta \quad (30.23b)$$

$$\frac{\partial \Psi}{\partial z} = -\frac{\partial \eta}{\partial t} \quad \text{linearized kinematic b.c. at } z = \eta \quad (30.23c)$$

$$\hat{\mathbf{n}} \cdot \nabla \Psi = 0 \quad \text{no-normal flow kinematic b.c. on rigid boundaries.} \quad (30.23d)$$

²Alternatively, make use of a further gauge transformation $\Psi' = \Psi - t(p_a/\rho)$ to eliminate the constant.

The first equation holds throughout the fluid whereas the remaining equations hold only at the boundaries.

Although the equations (30.23a)-(30.23d) were derived through linearization, there is an additional nonlinear term to remove in order to allow for analytical treatment of the wave solutions. Namely, when combining the boundary conditions into a single equation we compute the time derivative of equation (30.23b) according to

$$g \frac{\partial \eta}{\partial t} = \left[\frac{\partial}{\partial t} + \frac{\partial \eta}{\partial t} \frac{\partial}{\partial z} \right] \frac{\partial \Psi}{\partial t} \quad (30.24)$$

and then combine with equation (30.23c) to render

$$\left[\frac{\partial}{\partial t} + \frac{\partial \eta}{\partial t} \frac{\partial}{\partial z} \right] \frac{\partial \Psi}{\partial t} = -g \frac{\partial \Psi}{\partial z}. \quad (30.25)$$

With $w \approx \partial \eta / \partial t$ at the free surface, we identify $(\partial \eta / \partial t) \partial_z$ as a vertical advection operator. The corresponding term $(\partial \eta / \partial t) \partial_{zt} \Psi$ is nonlinear and second order in perturbation terms. It is therefore dropped when fully linearizing the system. An equivalent means to realize this linearization is to evaluate the free surface boundary condition at $z = 0$ rather than at $z = \eta(x, y, t)$. For this approximation to be self-consistent requires the amplitude of free surface undulations to be much smaller than the typical wavelengths of the fluctuations

$$|\eta|/\lambda \ll 1. \quad (30.26)$$

In summary, the fully linearized equation set takes the form

$$\nabla^2 \Psi = 0 \quad \text{irrotational and non-divergent velocity throughout domain} \quad (30.27a)$$

$$g \eta = \frac{\partial \Psi}{\partial t} \quad \text{linearized dynamic b.c. at } z = 0 \quad (30.27b)$$

$$\frac{\partial \eta}{\partial t} = -\frac{\partial \Psi}{\partial z} \quad \text{linearized kinematic b.c. at } z = 0 \quad (30.27c)$$

$$\hat{\mathbf{n}} \cdot \nabla \Psi = 0 \quad \text{no-normal flow kinematic b.c. on rigid boundaries.} \quad (30.27d)$$

30.2 Surface gravity waves in a flat bottom channel

We now derive a wave-like solution to the equations (30.27a)-(30.27d) posed in a flat bottom channel domain as illustrated in Figure 30.1. This analysis provides experience with the *separation of variables* method of use for deriving solutions to certain partial differential equations. We are not interested in the most general wave solution. Instead, we aim to determine a particular solution of sufficient generality to expose the underlying physics of the linear wave fluctuations, and in particular to expose the exponential decay of the wave amplitude with depth. Furthermore, given linearity, the superposition principle holds whereby the linear sum of particular solutions is also a solution.

30.2.1 Monochromatic wave solution

We seek a monochromatic wave solution with the wave moving parallel to the channel with radial frequency $\omega = 2\pi/\tau$ and wavenumber $k = 2\pi/\lambda$, with τ the period and λ the wavelength. We

furthermore assume the waves appear in the velocity potential in the shape of a cosine with an undetermined vertical structure

$$\Psi(x, y, z, t) = \Gamma(z) \cos(k x - \omega t). \quad (30.28)$$

Note the assumed absence of y dependence, as motivated by our choice to examine waves moving parallel to the channel and with symmetry across the channel. Furthermore, the cosine and sine functions form a complete set of basis functions for the channel domain, so that they can be superposed to generate any arbitrary wave pattern. This property is the basis for Fourier analysis commonly employed to study waves.

Plugging the ansatz (30.28) into Laplace's equation $\nabla^2 \Psi = \partial_{xx} \Psi + \partial_{zz} \Psi = 0$ leads to the ordinary differential equation satisfied by the vertical structure function

$$\frac{d^2 \Gamma}{dz^2} = k^2 \Gamma \quad -H \leq z \leq 0 \quad (30.29a)$$

$$\frac{d\Gamma}{dz} = 0 \quad \text{at } z = -H, \quad (30.29b)$$

where the bottom boundary condition is required to satisfy the no-normal flow condition (30.27d). We write the solution in the form

$$\Psi = \Psi_o \cosh[k(z + H)] \cos(kx - \omega t), \quad (30.30a)$$

$$\Psi_o = \frac{g \eta_o / \omega}{\cosh(kH)} \quad (30.30b)$$

so that the boundary condition (30.27b) renders the free surface height

$$\eta(x, t) = -\eta_o \sin(kx - \omega t). \quad (30.31)$$

The corresponding velocity field, $\mathbf{v} = -\nabla \Psi$, is given by

$$u = -\Psi_o k \cosh[k(z + H)] \sin(kx - \omega t) \quad (30.32a)$$

$$w = -\Psi_o k \sinh[k(z + H)] \cos(kx - \omega t). \quad (30.32b)$$

30.2.2 Dispersion relation for surface gravity waves

Combining the two $z = 0$ boundary conditions (30.27b) and (30.27c) yields

$$\frac{\partial^2 \Psi}{\partial t^2} + g \frac{\partial \Psi}{\partial z} = 0 \quad z = 0. \quad (30.33)$$

Substituting the wave solution (30.30a) into this relation leads to the constraint

$$\omega^2 = g k \tanh(kH). \quad (30.34)$$

This is the *dispersion relation* that constrains values available for the radial frequency, ω , and wavenumber, k . That is, the surface gravity waves only exist if their frequency and wavenumber are related according to the dispersion relation (30.34). The corresponding phase speed for the wave is given by

$$c = \frac{\omega}{k} = \sqrt{(g/k) \tanh(kH)}. \quad (30.35)$$

We emphasize the two limits: $kH \gg 1$ (shortwaves/deep water) and $kH \ll 1$ (longwave/shallow water), in which the phase speed satisfies

$$\omega \approx \sqrt{gk} \quad c \approx \sqrt{g/k} \quad kH \gg 1 \quad \text{shortwave/deep water limit} \quad (30.36a)$$

$$\omega \approx k \sqrt{gH} \quad c \approx \sqrt{gH} \quad kH \ll 1 \quad \text{longwave/shallow water limit.} \quad (30.36b)$$

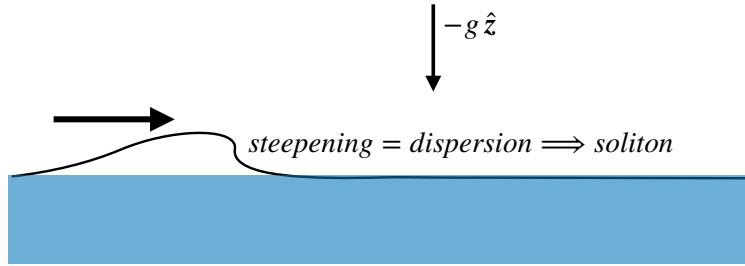


FIGURE 30.2: A soliton in the deep water limit results when the wave dispersion (long waves travel faster than short waves) balances the nonlinear steepening. The result is a soliton, which has an exact analytic expression following from the KdV equation (see [Drazin and Johnson \(1989\)](#)).

30.2.3 Heuristics of waves in the shortwave/deep water limit

The shortwave/deep water waves are notable for having shorter waves travel slower than longer waves. In the event of a perturbation to the fluid, such as from a stone dropped into a pond or a storm on a lake or the ocean, deep water waves are energized. The dispersion relation means that longer waves spread away from the storm center faster than the shorter waves, leading to a self-organization of the wavelengths and corresponding wave packets.

Now imagine a deepwater wave packet (group of waves traveling together) that somehow steepens and takes on a nonlinear form. Fourier decomposing this nonlinear wave into linear deepwater modes requires more shortwave linear modes in the steep region, whereas the less steep portion of the wave requires longer deep water Fourier modes, which travel faster. If the nonlinear steepening on the wave face is exactly balanced by the faster dispersion of the long waves near the wave base and back-side, then the wave pattern remains stable; it does not break. This balance of steepening and dispersion describes the fundamental features of a soliton.

30.2.4 Heuristics of waves in the longwave/shallow water limit

The longwave/shallow water limit is notable for the absence of wave dispersion; i.e., shallow water gravity waves of all wavelengths travel at speed \sqrt{gH} . Tsunamis are the prototypical shallow water waves that travel at speeds well approximated by \sqrt{gH} .

The dispersion relation also means that shallow water gravity waves slow down when the depth shoals, as when approaching a beach. Consequently, as waves reach the shoreline there is a tendency to accumulate wave energy as the deeper waves pile up behind the shallower waves. Furthermore, the steeper part of the wave, being part of a thicker region of the fluid and thus a larger effective H , travels slightly faster than the wave trough. As such, the steeper part of the wave overtakes the trough and, at some point, the assumptions of linearity breakdown and the shallow water waves break on the shore as depicted in Figure 30.3.

30.2.5 Further study

Elements from this section follow Section 54 of the [Fetter and Walecka \(1980\)](#), with their discussion of surface gravity waves also probing topics such as energetics, group velocity, and the inclusion of surface tension as appropriate for deep water waves with wavelengths on the order of centimeters. An accessible development of the theory of solitons can be found in the book by [Drazin and Johnson \(1989\)](#). Ocean surface waves are generally energized by atmospheric winds, in particular large storms. The dynamics of gravity wave breaking in the open ocean and on the beach remains

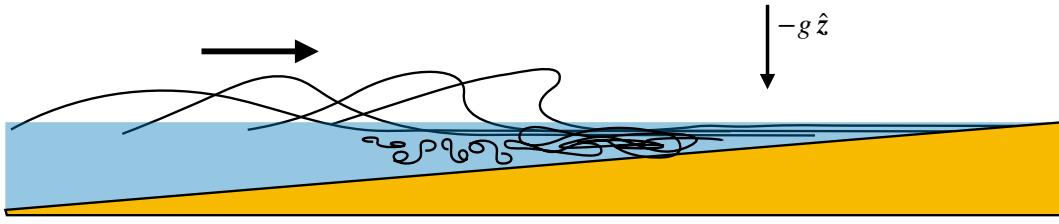


FIGURE 30.3: Shallow water waves approaching a shoreline steepen and eventually break. We can infer this behavior from the dispersion relation, $\sqrt{g H}$, whereby waves in deeper water move slightly faster than those in shallower water, so that the wave energy accumulates near the shore. Furthermore, water on the steeper part of the wave moves slightly faster than water in the trough, due to the difference in thickness of the water. This process causes water on the steeper portion of the wave to travel slightly faster than in the trough, leading to steepening of the waves. Nonlinearities eventually invalidate the assumptions made in deriving the linear wave dispersion relation $c = \sqrt{g H}$. Even so, the qualitative characterization based on the linear analysis allows for a useful heuristic understanding of shallow water wave breaking on the beach.

a decades-long problem in fluid mechanics that stretches the limits of theory, experiment, and simulation. [Bühler \(2014a\)](#) offers a thorough treatment of waves and their effects on mean flows

This video from Prof. A. Hogg provides a pedagogical introduction to shallow water wave breaking as well as deep water solitons realized in a laboratory.

30.3 Stokes drift

We here consider a fluid particle moving as part of a wave field. In homogenous linear wave fields, the particle periodically returns to its original position. However, in the presence of wave inhomogeneities, such as the surface gravity waves considered in this chapter, fluid particles generally oscillate between regions where the undulation in one direction does not match that in the other direction. In this case there is a net particle drift or transport known as *Stokes drift*. This drift occurs even though the phase average of the wave at a fixed spatial point vanishes. Formulating the mathematics of Stokes drift offers a means to explore the differences between averages formed at a fixed space point (Eulerian mean) versus averages following fluid particles (Lagrangian mean). We here introduce these notions, which form part of the rudiments for wave-mean flow interaction theory further pursued in Chapter 49.

30.3.1 General formulation of Stokes drift

Consider a three-dimensional particle trajectory written in Cartesian coordinates,

$$\mathbf{X}(\mathbf{a}, t) = X(\mathbf{a}, t) \hat{\mathbf{x}} + Y(\mathbf{a}, t) \hat{\mathbf{y}} + Z(\mathbf{a}, t) \hat{\mathbf{z}}. \quad (30.37)$$

In the analysis of waves, it is common to assume the material coordinate, \mathbf{a} , is the initial position of a fluid particle so we make that assumption here. As discussed in Section 13.8.1, the particle trajectory is determined by time integrating the particle velocity (also known as the “flow map”)

$$\frac{\partial \mathbf{X}(\mathbf{a}, t)}{\partial t} = \mathbf{v}[\mathbf{X}(\mathbf{a}, t), t] \quad (30.38)$$

so that

$$\mathbf{X}(\mathbf{a}, t) = \mathbf{X}(\mathbf{a}, 0) + \int_0^t \mathbf{v}[\mathbf{X}(\mathbf{a}, t'), t'] dt'. \quad (30.39)$$

This equation is a trivial result of time integrating the flow map. Nonetheless, it is useful to express the content of this equation in words. It says that the position at time t of a fluid particle labelled by the material coordinate \mathbf{a} is given by the initial position of the particle, $\mathbf{X}(\mathbf{a}, 0)$, plus the time integrated movement of the particle following the fluid flow.

We now form a Taylor series computed relative to the initial position of the particle, so that the particle velocity at time t takes on the approximate form

$$v^n[\mathbf{X}(\mathbf{a}, t), t] \approx v^n[\mathbf{X}(\mathbf{a}, 0), t] + \nabla v^n[\mathbf{X}(\mathbf{a}, 0), t] \cdot [\mathbf{X}(\mathbf{a}, t) - \mathbf{X}(\mathbf{a}, 0)] \quad (30.40a)$$

$$= v^n[\mathbf{X}(\mathbf{a}, 0), t] + \nabla v^n[\mathbf{X}(\mathbf{a}, 0), t] \cdot \int_0^t \frac{d\mathbf{X}(\mathbf{a}, t')}{dt'} dt' \quad (30.40b)$$

$$= v^n[\mathbf{X}(\mathbf{a}, 0), t] + \nabla v^n[\mathbf{X}(\mathbf{a}, 0), t] \cdot \int_0^t \mathbf{v}[\mathbf{X}(\mathbf{a}, t'), t'] dt', \quad (30.40c)$$

where the Taylor series was truncated after terms linear in the particle displacement $\mathbf{X}(\mathbf{a}, t) - \mathbf{X}(\mathbf{a}, 0)$. We emphasize two points regarding equation (30.40c).

- How do we interpret $v^n[\mathbf{X}(\mathbf{a}, 0), t]$? This is the n 'th component of the velocity field evaluated at the initial point of the trajectory, $\mathbf{X}(\mathbf{a}, 0)$, at time t . That is, it is the Eulerian velocity evaluated at the fixed Eulerian point $\mathbf{X}(\mathbf{a}, 0)$.
- What determines the accuracy of the Taylor series? A suitable non-dimensional expansion coefficient for the Taylor expansion is the ratio of the particle displacement to the scale, Λ , of inhomogeneities in flow properties

$$\epsilon = \frac{|\mathbf{X}(\mathbf{a}, t) - \mathbf{X}(\mathbf{a}, 0)|}{\Lambda}. \quad (30.41)$$

This ratio is small for the small amplitude waves considered here, whereby the particle displacements are far smaller than inhomogeneities in flow properties.

The integrand on the right hand side of equation (30.40c) is the Lagrangian velocity integrated over the time interval. To within the same order of accuracy as maintained thus far, we can use the Eulerian velocity evaluated at the initial position, thus rendering

$$v^n[\mathbf{X}(\mathbf{a}, t), t] \approx v^n[\mathbf{X}(\mathbf{a}, 0), t] + \nabla v^n[\mathbf{X}(\mathbf{a}, 0), t] \cdot \int_0^t \mathbf{v}[\mathbf{X}(\mathbf{a}, 0), t'] dt', \quad (30.42)$$

with rearrangement leading to

$$v^n[\mathbf{X}(\mathbf{a}, t), t] - v^n[\mathbf{X}(\mathbf{a}, 0), t] \approx \nabla v^n[\mathbf{X}(\mathbf{a}, 0), t] \cdot \int_0^t \mathbf{v}[\mathbf{X}(\mathbf{a}, 0), t'] dt'. \quad (30.43)$$

The left hand side is the difference between the velocity following a fluid particle (the Lagrangian velocity for the moving fluid particle) from the velocity at the initial particle point (the Eulerian velocity at the initial point of the trajectory). The right hand side terms are all evaluated at the initial position, $\mathbf{X}(\mathbf{a}, 0)$. Furthermore, the right hand side is non-zero where the velocity at the initial position has a nonzero gradient (i.e., it is spatially inhomogeneous), with its inhomogeneity projecting onto the time integrated velocity at that point. Equation (30.43) says that the velocity following a fluid particle is modified from the velocity at its initial position if the particle moves through an inhomogeneous velocity field.

The Stokes drift is defined as the difference of the velocities in equation (30.43) when time averaged over a wave period, which we write as

$$v_{(S)}^n[\mathbf{X}(\mathbf{a}, 0), t] = \overline{v^n[\mathbf{X}(\mathbf{a}, t), t] - v^n[\mathbf{X}(\mathbf{a}, 0), t]} \quad (30.44a)$$

$$\approx \nabla v^n[\mathbf{X}(\mathbf{a}, 0), t] \cdot \int_0^t \mathbf{v}[\mathbf{X}(\mathbf{a}, 0), t'] dt'. \quad (30.44b)$$

This expression holds for any arbitrary initial point in the fluid, so that we can write it in a concise Eulerian form that dispenses with trajectories

$$v_{(S)}^n(\mathbf{x}, t) \approx \nabla v^n(\mathbf{x}, t) \cdot \overline{\int_0^t \mathbf{v}(\mathbf{x}, t') dt'}. \quad (30.45)$$

We liken Stokes drift to surfing: the more a particle samples larger amplitude variations in the velocity field (the gradient term), the further it drifts (the integral term). Note that for the case of a transverse wave disturbance, where the particle disturbance is orthogonal to the wavevector, then the Stokes drift vanishes to leading order.

30.3.2 Stokes drift in surface gravity waves

The canonical example of Stokes drift occurs in the near surface ocean, where the surface gravity waves discussed in Section 30.1 create particle motion that is larger near the surface than at depth. For this example, consider a monochromatic wave field in the zonal-vertical directions written in the form

$$\frac{dX}{dt} = U \sin(\omega t) \quad (30.46a)$$

$$\frac{dZ}{dt} = U \cos(\omega t), \quad (30.46b)$$

where $U > 0$ is the speed of the circular particle motion and we use the capital X and Z to denote Cartesian components of the particle trajectory. To simplify the mathematics we perform the analysis in a frame where the waves are stationary, hence the $kx - \omega t$ phase from Section 30.2. Figure 30.4 shows a schematic of the particle trajectories as derived in the following.

Homogeneous flow field

If the background flow is homogeneous, then the speed is a constant, $U = U_o > 0$. Particle trajectories in this case are clockwise in the x-z plane around a circle with radius U_o/ω

$$X(t) - X_o = -\frac{U_o}{\omega} [\cos(\omega t) - 1] \quad (30.47a)$$

$$Z(t) - Z_o = \frac{U_o}{\omega} \sin(\omega t), \quad (30.47b)$$

where the initial position at time $t = 0$ is

$$\mathbf{X}(t = t_o) = \mathbf{X}_o, \quad (30.48)$$

and the center of the circle is

$$\mathbf{X}_{\text{center}} = \left[X_o + \frac{U_o}{\omega} \right] \hat{\mathbf{x}} + Z_o \hat{\mathbf{z}}. \quad (30.49)$$



FIGURE 30.4: Sketch of Stokes drift in the upper ocean with the wave vector in the horizontal direction and clockwise time integrated fluid particle motion induced by the traveling wave. For the case of a wave amplitude that decreases with depth, lateral motion of the particle is larger when the particle is closer to the surface thus leading to a Stokes drift in the direction of the wave. Note that there is zero Stokes drift for the case of a homogeneous wave, in which the wave amplitude is independent of depth. Additionally, and even more trivially, if the particle motion is purely transverse to the wave vector, in this case purely vertical, then the particle merely retraces its motion along a vertical line and does not undergo any lateral Stokes drift.

Inhomogeneous flow field

In the presence of vertical inhomogeneities of the flow field, the wave amplitude becomes a function of depth, $U = U(z)$. The canonical example is where the wave amplitude decreases with depth, as for the surface gravity waves detailed in Section 30.2. In turn, we expect there to be a fluid particle drift in the zonal direction introduced by the vertical wave inhomogeneity. This drift is a particular realization of Stokes drift.

To compute the leading order expression for the Stokes drift, expand U in a Taylor series about the initial position

$$U \approx U_o + R(Z - Z_o) \quad (30.50)$$

where R has units of inverse time and is given by

$$R = \left[\frac{dU}{dZ} \right]_{Z=Z_o}. \quad (30.51)$$

For the surface gravity waves with velocity components (30.32a) and (30.32b), we have

$$U = \Psi_o k \sqrt{\cosh[2k(z + H)]} \quad (30.52)$$

so that with $Z_o = 0$,

$$R = \Psi_o k^2 \frac{\sinh(2kH)}{\sqrt{\cosh(2kH)}}. \quad (30.53)$$

In general, the Taylor series is valid so long as the vertical trajectories maintain the inequality

$$|R| |Z - Z_o| \ll U_o, \quad (30.54)$$

which says that the vertical shear is small

$$|R| \ll \frac{U_o}{|Z - Z_o|}. \quad (30.55)$$

We use the Taylor series expansion (30.50) to solve for the vertical trajectory as determined by

$$\frac{d(Z - Z_o)}{dt} = [U_o + R(Z - Z_o)] \cos(\omega t). \quad (30.56)$$

Rearrangement leads to

$$\int_{Z_o}^Z \frac{d(Z - Z_o)}{U_o + R(Z - Z_o)} = \int_0^t \cos(\omega t) dt. \quad (30.57)$$

The left hand side integral can be computed by changing variables

$$\Sigma = U_o + R(Z - Z_o) \quad (30.58a)$$

$$d\Sigma = R d(Z - Z_o), \quad (30.58b)$$

so that equation (30.57) becomes

$$\int_{U_o}^{\Sigma} \frac{d\Sigma}{\Sigma} = R \int_0^t \cos(\omega t) dt. \quad (30.59)$$

Performing the integrals and evaluating the end points renders

$$\ln \left[1 + \frac{R}{U_o} (Z - Z_o) \right] = \frac{R \sin(\omega t)}{\omega}, \quad (30.60)$$

which yields the exponential solution

$$1 + \frac{R}{U_o} (Z - Z_o) = e^{(R/\omega) \sin(\omega t)} \Rightarrow Z - Z_o = \frac{U_o}{R} \left(-1 + e^{(R/\omega) \sin(\omega t)} \right). \quad (30.61)$$

The vertical particle position is seen to oscillate around its initial position Z_o .

We next consider the zonal particle position, in which case

$$\frac{d(X - X_o)}{dt} = U_o \left[1 + \frac{R}{U_o} (Z - Z_o) \right] \sin(\omega t) \quad (30.62a)$$

$$= U_o e^{(R/\omega) \sin(\omega t)} \sin(\omega t) \quad (30.62b)$$

where we used equation (30.61) for the vertical trajectory. To make progress, we expand the exponential assuming the ratio of inverse time scales, R/ω , is small

$$|R/\omega| \ll 1. \quad (30.63)$$

In this limit, the vertical trajectory retains its unperturbed form (30.47b), and the zonal trajectory satisfies

$$\frac{d(X - X_o)}{dt} \approx U_o \sin(\omega t) \left[1 + \frac{R}{\omega} \sin(\omega t) \right], \quad (30.64)$$

where we dropped terms of order $(R/\omega)^2$. We can understand the scaling in equation (30.63) by noting that the period for the circular motion is given by

$$\tau_{\text{circle}} = \frac{2\pi}{\omega}. \quad (30.65)$$

The inverse time R introduces a time scale for the drift, defined according to

$$\tau_{\text{drift}} = \frac{2\pi}{|R|}. \quad (30.66)$$

A small ratio $|R/\omega|$ thus implies

$$|R/\omega| = \tau_{\text{circle}}/\tau_{\text{drift}} \ll 1. \quad (30.67)$$

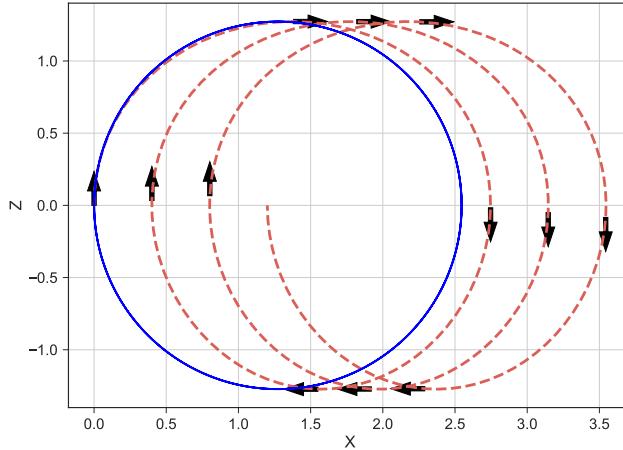


FIGURE 30.5: Example trajectories of fluid particles undergoing Stokes drift for short surface ocean waves. Particle motion is clockwise in the x - z plane. For homogeneous waves, there is zero Stokes drift with circular trajectories given by equations (30.47a) and (30.47b), as depicted here by the blue trajectory. There is a Stokes drift in the presence of vertical shear in the wave amplitude and thus in the particle velocity, with the trajectories for this example given by equation (30.61) for the vertical component and equation (30.69b) for the horizontal component. We set the parameters as follows: $T = 2\pi/\omega = 60$ s, $U_o = 0.1$ m s $^{-1}$, and $R = \omega/10$ and exhibit trajectories over four minutes.

Hence, we are solving for the zonal trajectory in the limit where the time scale for the circular motion is small (i.e., fast oscillations around the circle) relative to the time scale for the drift (i.e., slow drift).

Returning now to the approximate zonal trajectory equation (30.64) yields

$$\frac{d(X - X_o)}{dt} = U_o \sin(\omega t) \left[1 + \frac{R}{\omega} \sin(\omega t) \right] \quad (30.68a)$$

$$= U_o \sin(\omega t) + \frac{U_o R}{2\omega} [1 - \cos(2\omega t)], \quad (30.68b)$$

which integrates to

$$X - X_o = \left(\frac{U_o}{\omega} \right) \left[1 - \cos(\omega t) - \frac{R \sin(2\omega t)}{4\omega} + \frac{R t}{2} \right] \quad (30.69a)$$

$$= \underbrace{\left(\frac{U_o}{\omega} \right) [1 - \cos(\omega t)]}_{\text{homogeneous}} + \underbrace{\frac{U_o R t}{2\omega}}_{\text{Stokes drift}} - \underbrace{\frac{U_o R \sin(2\omega t)}{4\omega^2}}_{\text{higher harmonic}} + \mathcal{O}(R/\omega)^2. \quad (30.69b)$$

The leading order term is the homogeneous motion given by equation (30.47a). The next term is the Stokes drift, followed by a higher order harmonic and then further terms on the order of $(R/\omega)^2$. There is no vertical Stokes drift to this order in (R/ω) , so that the Stokes drift velocity is given by

$$\left[\frac{\mathbf{X} - \mathbf{X}_o}{t} \right]^{\text{drift}} = \frac{R U_o}{2\omega} \hat{\mathbf{x}}. \quad (30.70)$$

The circular motion of the parcels is therefore deformed by the zonal Stokes drift. The drift increases with larger wave amplitude (U_o large); with larger vertical shear (R large); and with longer period waves (ω small). See Figure 30.5 for an illustration based on a particular choice for the dimensional parameters.

30.3.3 Comments and further study

Stokes drift occurs in many guises when studying the motion of fluid particles within wave fields. We revisit elements of Stokes drift in Chapter 49 when studying the rudiments of eddy-induced tracer transport. This video from Prof. Andy Hogg at Australian National University provides an overview of the discussion in this section along with some laboratory experiments to illustrate Stokes drift.

Is Stokes drift a nonlinear wave phenomena? In answering this question we note that Stokes drift occurs with particle motion in linear waves, but the waves must be inhomogeneous such as the surface gravity waves studies in this chapter. However, nonlinearity appears in the form of the particle-following averaging, as can be seen by the expression of Stokes drift given by equation (30.45)

$$\overline{v_{(S)}^n(\mathbf{x}, t)} \approx \nabla v^n(\mathbf{x}, t) \cdot \int_0^t \mathbf{v}(\mathbf{x}, t') dt'. \quad (30.71)$$

The dot product of the velocity gradient with the time integrated velocity (to give the time integrated position) is nonlinear. So although the waves are linear, the Lagrangian kinematics of particle trajectories introduces nonlinearities.



30.4 Exercises

EXERCISE 30.1: STOKES DRIFT FOR ONE-DIMENSIONAL MONOCHROMATIC WAVE

Consider a one-dimensional monochromatic longitudinal wave with velocity

$$u = u_o \sin(k x - \omega t), \quad (30.72)$$

where u_o is the wave amplitude, $k = 2\pi/\lambda$ the wave number, λ the wavelength, $\omega = 2\pi/T$ the radial frequency, T the wave period, and $c = \omega/k = \lambda/T$ the wave speed. A longitudinal wave is one whose particle motions are parallel to the wave vector, which in this exercise are both in the $\hat{\mathbf{x}}$ direction. Determine the wave period averaged Stokes velocity to first order accuracy in the small parameter

$$\epsilon = \frac{u_o}{c} = \frac{u_o k}{\omega} = \frac{u_o T}{\lambda} \quad (30.73)$$

with this parameter the ratio of the wave amplitude to wave speed, or equivalently the ratio of the length scale of particle displacements to the wavelength. Hint: make use of the general result given by equation (30.45).



Part V

Shallow water mechanics

Adiabatic shallow water models consist of constant density homogeneous fluid layers whose interfaces are material (i.e., no matter is transferred between the immiscible layers). Thermodynamic processes are absent from the system, thus allowing us to focus on the dynamics of fluid layers. Pressure is hydrostatic, which means that motion occurs in columns whereby horizontal velocity is independent of the vertical position within a layer, whereas the column can expand and contract so that there is vertical motion within a layer. Horizontal momentum is transferred between shallow water layers via pressure form stresses that act on sloping layer interfaces.

The shallow water system offers a suite of versatile theoretical models of use to deduce impacts on the flow from both rotation and stratification. Consequently, the shallow water model is featured in many areas of geophysical fluid mechanics as well as in ocean and atmosphere applications. In this part of the book, our goal is to develop an understanding of the shallow water model fundamentals and study some basic dynamical features as realized in the model. We have further use of the shallow water model in our study of vorticity in Part VI, balanced flows in Part VII, and generalized vertical coordinates in Part VIII.

We offer an observation concerning the mathematics required for studying the shallow water model. Namely, the shallow water model is a vertically discrete realization of a continuously stratified fluid described by the isopyncal coordinates of Chapter 46. However, the mathematical formalism of generalized vertical coordinates developed in Part VIII is largely unnecessary when working with the shallow water model. The reason for the simplification is that columnar motion within a shallow water layer means that lateral gradients of properties need not be projected along the slope of the layer. In contrast, this projection is needed for a continuously stratified fluid described by generalized vertical coordinates (e.g., see Figure 43.4). So although we appreciate the beauty and power of generalized vertical coordinate mathematics for studying continuously stratified flows, it is liberating to be able to avoid that formalism while still capturing much of the underlying physics associated with stratification (albeit discretely stratified). This feature of the shallow water model greatly adds to its allure and accessibility.

Shallow water models

In this chapter we formulate the kinematic and dynamic equations for a suite of shallow water models. For this purpose we develop the equations describing motion of a single shallow water layer; multiple shallow water layers (stacked shallow water); and reduced gravity models (models with one layer that is dynamically inactive). In formulating these models we also expose their underlying physical basis.

READER'S GUIDE TO THIS CHAPTER

We make use of fluid kinematics and dynamics described in earlier chapters. Our presentation was inspired by Chapter 3 of [Vallis \(2017\)](#) as well as various sections in [Salmon \(1998\)](#).

The horizontal velocity is vertically uniform within a shallow water layer, whereas the vertical velocity and hydrostatic pressure are linear functions of vertical position within the layer. When acting on a property that is vertically uniform within a layer, the gradient operator, ∇ , results in a horizontal vector. To minimize notational clutter, we typically write ∇ for brevity, rather than the more clumsy ∇_k or ∇_z (with k the layer index). The meaning of the resulting vector equations will be clear from the functional dependencies of the fields present in the equations.

31.1	Loose threads	674
31.2	A single shallow water layer	674
31.2.1	Pressure gradient force within the fluid layer	674
31.2.2	Further comments on pressure in a homogeneous layer	676
31.2.3	Momentum equation	677
31.2.4	Thickness equation	677
31.2.5	Bottom kinematic boundary condition	679
31.2.6	Surface kinematic boundary condition	680
31.2.7	Column stretching and vertical velocity	681
31.2.8	Tracer concentration equation	682
31.2.9	Summary comments	683
31.3	Reduced gravity model for the upper ocean	683
31.3.1	Momentum and thickness equations for the active layer	684
31.3.2	Relating undulations of the top and bottom layer interfaces	684
31.3.3	Momentum equation with reduced gravity	685
31.3.4	Some intuition about reduced gravity	685
31.3.5	Further study	686
31.4	Stacked shallow water equations	686
31.4.1	2-layer model formulation	687
31.4.2	N -layer model equations	689

31.5 Non-conservative processes and boundary stresses	690
31.5.1 Dia-surface volume fluxes	690
31.5.2 Frictional and boundary stresses	692
31.6 Exercises	692

31.1 Loose threads

- Internal and external mode as per Raymond's notes.

31.2 A single shallow water layer

Consider a homogeneous layer of fluid in a uniform effective gravitational field (gravity plus planetary centrifugal as in Section 10.12.4), contained on its side by solid boundaries. If there are no lateral force imbalances, then the fluid remains static. Now perturb the fluid so that it has a nonuniform layer thickness, say with a bump in a particular region. Conservation of fluid mass (which translates into volume conservation for a uniform density layer) means that thicker fluid regions must be exactly compensated by thinner fluid regions. Furthermore, layer thickness gradients create pressure differences (thicker water has larger hydrostatic pressure than thinner water), which in turn drives fluid motion. If the fluid has much larger lateral extent than vertical, then the lateral motion occurs as an expanding and contracting column with no depth dependence to the horizontal pressure forces and thus the horizontal motion is depth independent.

The essence of a perfect fluid (i.e., no irreversible processes such as mixing) shallow water flow concerns the motion of fluid columns accelerated by pressure gradients created by layer thickness undulations, and the associated conservation of mass ensuring that the accumulation of fluid in one region is balanced by the depletion of fluid in another. Pressure gradients act to homogenize the layer thickness. However, rotation allows for layer thickness to be non-constant even in a steady state.

31.2.1 Pressure gradient force within the fluid layer

Figure 31.1 shows a single shallow water layer with a generally non-flat bottom and an undulating free surface height. We assume that each column of fluid within the layer is in hydrostatic balance, so that the vertical momentum equation reduces to

$$\frac{\partial p}{\partial z} = -\rho g. \quad (31.1)$$

Recall from Section 23.3 that the hydrostatic balance is consistent with lateral length scales much larger than vertical (small vertical to horizontal aspect ratio), which is satisfied by large-scale geophysical fluid motion. Hence, a shallow water fluid is a relevant idealization if we are considering large horizontal scales relative to the vertical.

Since the fluid density is assumed constant (i.e., the fluid is a homogeneous layer), we can integrate the hydrostatic balance from the surface to an arbitrary vertical position within the layer

$$p(x, y, z, t) = p_a(x, y, t) + g \rho \int_z^\eta dz = p_a(x, y, t) + g \rho [\eta(x, y, t) - z], \quad (31.2)$$

where $p_a(x, y, t)$ is the pressure applied to the layer free surface, say from the overlying atmosphere. Furthermore, the horizontal pressure gradient thus takes the form

$$\nabla_z p = \nabla_z p_a + g \rho \nabla_z \eta. \quad (31.3)$$

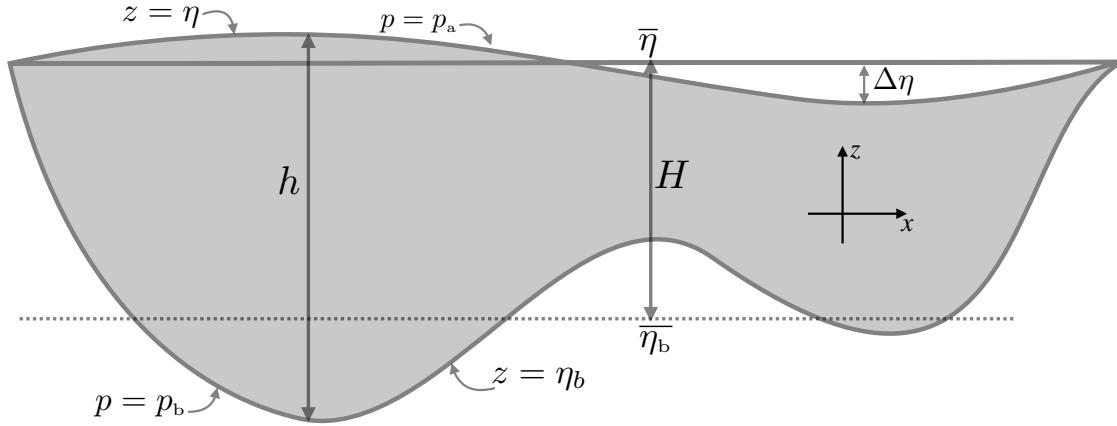


FIGURE 31.1: A single layer of shallow water fluid with thickness $h = \eta - \eta_b$ that extends from the bottom at $z = \eta_b$ to the free surface at $z = \eta$. The area averaged thickness is defined by $H = \bar{y} = A^{-1} \int h \, dx \, dy = A^{-1} \int (\eta - \eta_b) \, dx \, dy = \bar{\eta} - \bar{\eta}_b$, where $A = \int dx \, dy$ is the horizontal area of the layer. The deviation of the free surface from $\bar{\eta}$ is given by $\Delta\eta = \eta - \bar{\eta} = \eta - (\bar{\eta}_b + H)$ so that $\eta(x, y, t) = \eta_b(x, y) + h(x, y, t) = \bar{\eta}_b + H + \Delta\eta(x, y, t)$. Likewise, the deviation of the bottom from $\bar{\eta}_b$ is given by $\Delta\eta_b = \eta_b - \bar{\eta}_b$, so that $h = \eta - \eta_b = H + \Delta\eta - \Delta\eta_b$. Volume conservation for the layer is maintained in the absence of volume boundary fluxes, in which case $\Delta\eta = 0$. Note that the position of the reference height, $z = 0$, is arbitrary. Atmospheric conventions typically choose $z = 0$ so that $\bar{\eta}_b = 0$, $\eta = H + \Delta\eta$, and $\bar{\eta} = H$. Oceanic conventions typically choose $\bar{\eta} = 0$ so that $\eta = \Delta\eta$ and $\bar{\eta}_b = -H$. We are only concerned with fluctuations that leave the free surface monotonic; i.e., we do not consider overturns or breaking waves. This assumption is implied by assuming that each column extending from $\eta_b \leq z \leq \eta$ maintains hydrostatic balance. We refer to this figure countless times in this book when discussing shallow water mechanics.

Since p_a and η are independent of z , there is no need to expose the z subscript on the gradient operator on the right hand side. We thus drop the subscript when no ambiguity results, as per our convention noted at the start of the chapter.

Effective sea level and inverse barometer responses

It is common to ignore the applied surface pressure, p_a , by assuming the fluid above the layer either has zero horizontal variations in pressure, or, more commonly, by assuming the fluid has zero inertia (i.e., zero mass). This assumption must be relaxed if studying the effects of atmospheric pressure variations acting on a single layer ocean. For generality, we sometimes find it useful to introduce an *effective free surface height*

$$\eta^{\text{eff}} = \eta + p_a / (\rho g). \quad (31.4)$$

When the applied pressure is anomalously positive, the free surface is generally depressed under the added weight. A perfect compensation occurs when the free surface exhibits an *inverse barometer response*, whereby the effective sea level remains unchanged. Appendix C to [Griffies and Greatbatch \(2012\)](#) offer a discussion of the inverse barometer response of the ocean to atmospheric loading as well as sea ice loading.

Layer depth independent pressure gradient force

We emphasize that although hydrostatic pressure within a shallow water layer is depth dependent as per equation (31.2), the horizontal pressure gradient, as given by equation (31.3), has no depth dependence within the layer. The acceleration from this hydrostatic pressure force points from highs in the effective sea level to lows in the effective sea level (see Figure 31.2).

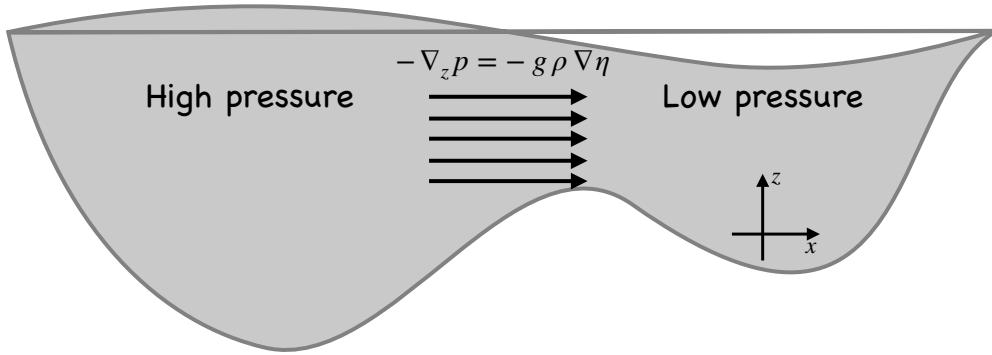


FIGURE 31.2: When there are zero horizontal gradients in the applied surface pressure, $\nabla p_a = 0$, the horizontal acceleration from pressure within a single shallow water fluid layer is determined solely by the surface height, $-\nabla_z p = -g \nabla_z \eta$. The acceleration is uniform throughout the layer and points from sea level highs towards sea level lows. Although bottom topography interacts with the flow and thus affects the shape of the free surface, the topography does not appear explicitly in the horizontal pressure gradient. Instead, we only need to know the shape of the free surface (and the applied pressure p_a) to know the horizontal pressure force throughout the layer.

It is notable that the pressure gradient is solely determined by properties at the upper interface of the layer. That is, we only need to know the shape of the free surface, $z = \eta$, and the applied pressure, p_a , to know the horizontal pressure force acting throughout the layer. But what about the bottom topography? Experience indicates that topography affects the shape of the free surface and thus modifies the pressure field, yet topography does not explicitly appear in the mathematical expression for the horizontal pressure gradient. This characteristic is specific to the use of a pressure gradient body force to describe the role of pressure on the layer momentum. As a complement, we saw in Chapter 20 how to formulate the pressure contact force. It is through the contact force perspective that we see, in Section 32.4, how bottom topography appears in the momentum balance of a shallow water layer. In particular, this perspective exposes the topographic form stresses that mechanically exchange momentum between the layer and the solid earth bottom.

31.2.2 Further comments on pressure in a homogeneous layer

The depth independence of the horizontal pressure gradient within the shallow water layer is a direct result of the assumed hydrostatic nature of the pressure within the layer. To emphasize this point, we certainly can imagine a homogeneous fluid layer in which the horizontal velocity has a vertical shear. For example, in Section 30.1 we study surface gravity waves in a homogeneous fluid layer. Such waves have an amplitude that exponentially decays with depth, and so the horizontal and vertical fluid motion associated with the waves have a non-zero vertical shear. Such motion clearly cannot be caused by a hydrostatic pressure gradient since it has no vertical dependence throughout the layer, in which case

$$\partial_z(\nabla_z p_{\text{hydrostatic}}) = 0. \quad (31.5)$$

Hence, in a homogeneous fluid layer, hydrostatic pressure gradients can only drive a horizontal flow that is depth independent within that layer. So if the vertically sheared horizontal flow is found within a homogeneous fluid layer, and if pressure gradients cause this flow, then it can only be through gradients in the *non-hydrostatic pressure*. As discussed in Section 30.1.3, surface gravity waves indeed involve non-hydrostatic pressure forces that drive the vertical dependence to the wave amplitude.

Moving beyond the homogeneous layer assumption, we saw in Section 23.3.6 that a horizontal gradient in the density leads to a vertically dependent hydrostatic pressure gradient $\partial_z(\nabla_z p_{\text{hydrostatic}}) \neq 0$. This hydrostatic pressure force can impart vertical shears to the horizontal flow. Thermal wind shear is the canonical example whereby vertical shears in the horizontal velocity are present in geostrophically balanced fluids as driven by horizontal density gradients (Section 26.4.4).

31.2.3 Momentum equation

If there is no friction anywhere in the fluid, including at the upper and lower boundaries, then the horizontal momentum is effected only by the Coriolis and pressure forces. Following our discussion of the Traditional Approximation in Section 23.1.3, we retain only the local vertical component to the Coriolis acceleration, which is compatible with the hydrostatic approximation. We are thus led to the horizontal momentum equation

$$\frac{D\mathbf{u}}{Dt} + f \hat{\mathbf{z}} \wedge \mathbf{u} = -\nabla(g\eta + p_a/\rho), \quad (31.6)$$

where

$$\mathbf{v} = (\mathbf{u}, w) \quad (31.7)$$

splits out the horizontal velocity vector, \mathbf{u} , from the vertical velocity component, w .

The Coriolis parameter, $f = f \hat{\mathbf{z}}$, is independent of depth, as is the horizontal pressure force from gradients in the free surface and applied pressure. Consequently, if the horizontal velocity is initially independent of depth, it will remain so for all time. The material time derivative thus only has contributions from the local time derivative and from horizontal advection

$$\frac{D\mathbf{u}}{Dt} = \left[\frac{\partial}{\partial t} + u \frac{\partial}{\partial x} + v \frac{\partial}{\partial y} \right] \mathbf{u} \quad (31.8)$$

so that the shallow water momentum equation (31.6) takes on the form

$$\left[\frac{\partial}{\partial t} + u \frac{\partial}{\partial x} + v \frac{\partial}{\partial y} + f \hat{\mathbf{z}} \wedge \right] \mathbf{u} = -\nabla(g\nabla\eta + p_a/\rho). \quad (31.9)$$

31.2.4 Thickness equation

The mass of a shallow water layer is constant in the absence of sources, sinks, or boundary fluxes. Hence, changes in mass at a particular region in the fluid must arise from mass fluxed across the region boundaries, leaving one region and accumulating in another. For simplicity, we assume that no mass crosses the fluid top (the free surface) or the bottom (the solid earth). We consider the more general case of boundary mass transport in Section 31.5. Note that since the fluid density is constant, mass conservation is the same as volume conservation. Hence, the terms “mass conservation” and “volume conservation” are commonly used interchangeably when working with shallow water models.

Consider an infinitesimally thin (in horizontal cross-section) vertical column of shallow water fluid that is fixed in space and extending from $\eta_b \leq z \leq \eta$. Let the horizontal cross-sectional area be written as dA and the thickness be $h = \eta - \eta_b$ (see Figure 31.3). The total mass of fluid in this column is given by

$$M = \int_{\text{column}} \left[\int_{\eta_b}^{\eta} \rho dz \right] dA = \rho \int_{\text{column}} (\eta - \eta_b) dA = \rho \int_{\text{column}} h dA. \quad (31.10)$$

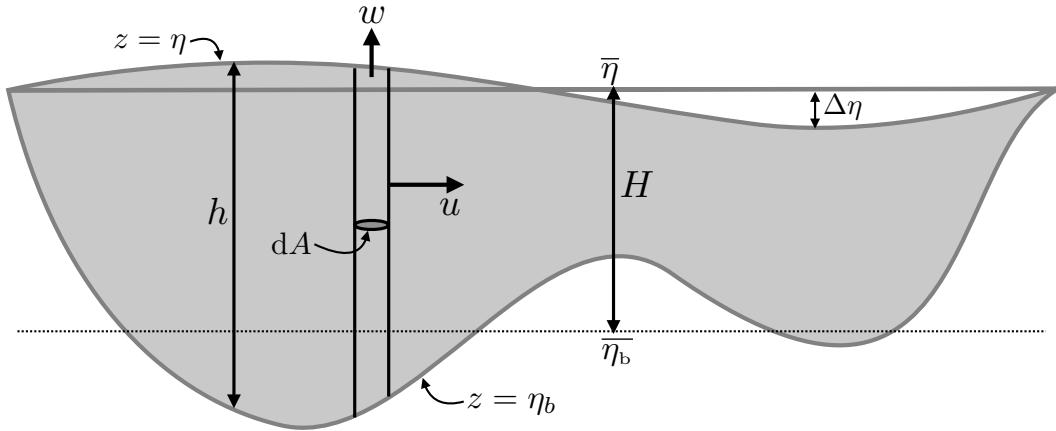


FIGURE 31.3: Mass budget for a column of shallow water fluid with fixed cross-sectional area, dA , constant density, ρ , and thickness, h . The column mass is affected only by horizontal transport (transport within the layer) in the absence of boundary mass fluxes through the top, $z = \eta$, or bottom, $z = \eta_b$. Note that since the density of the layer is constant, then mass equals to the constant density times the volume.

Time changes in the column mass thus arise from time changes in the layer thickness integrated over the horizontal area of the column

$$\frac{dM}{dt} = \rho \int_{\text{column}} \frac{\partial h}{\partial t} dA, \quad (31.11)$$

where

$$\frac{\partial h}{\partial t} = \frac{\partial (\eta - \eta_b)}{\partial t} = \frac{\partial \eta}{\partial t}, \quad (31.12)$$

since the bottom topography at $z = \eta_b(x, y)$ is static.

General derivation

The mass within a fluid column changes due to mass crossing the vertical column boundaries (again, we assume here that no mass crosses the top or bottom interfaces). The mass flux penetrating the vertical boundary is given by

$$\text{mass per time entering column} = -\rho \oint_{\text{column}} \mathbf{u} \cdot \hat{\mathbf{n}} dS, \quad (31.13)$$

where $\hat{\mathbf{n}}$ is the outward normal at the column boundary, and dS is the area element along the column boundary. The area integral is computed over the boundary of the column, which involves a vertical integral and a circumferential line integral

$$\text{mass per time entering column} = -\rho \oint_{\text{column}} \left[\int \mathbf{u} \cdot \hat{\mathbf{n}} dz \right] dl, \quad (31.14)$$

where dl is the infinitesimal line element around the column circumference. Since $\hat{\mathbf{n}} \cdot \mathbf{u}$ is depth independent, we can perform the vertical integral to render

$$-\rho \oint_{\text{column}} \left[\int \mathbf{u} \cdot \hat{\mathbf{n}} dz \right] dl = -\rho \oint_{\text{column}} h \mathbf{u} \cdot \hat{\mathbf{n}} dl = -\rho \int_{\text{column}} \nabla \cdot (h \mathbf{u}) dA, \quad (31.15)$$

where the second equality follows from the divergence theorem applied to the horizontal cross-sectional area of the column. Equating this result to the mass time tendency (31.11), and noting that the horizontal cross-sectional area is arbitrary, yields an equation for the layer thickness

$$\frac{\partial h}{\partial t} + \nabla \cdot (h \mathbf{u}) = 0. \quad (31.16)$$

This result means that the thickness of fluid at a fixed location increases if there is a convergence of thickness onto that location, and decreases if thickness diverges from the location. We may also write the thickness equation (31.16) in the material form

$$\frac{Dh}{Dt} = -h \nabla \cdot \mathbf{u}. \quad (31.17)$$

Hence, thickness of a material fluid column increases in regions where the horizontal flow converges.

Special case with a rectangular column

To further our understanding of the second step in equation (31.15), consider the special case of a rectangular column, for which the mass per time of fluid entering the column is given by

mass per time entering column

$$= -\rho \int_{\text{column}} [(u h)_{\text{east}} - (u h)_{\text{west}}] dy - \rho \int_{\text{column}} [(v h)_{\text{north}} - (v h)_{\text{south}}] dx. \quad (31.18)$$

Taking the limit as the horizontal cross-section of the column becomes infinitesimal leads to

mass per time entering column

$$= -\rho \int_{\text{column}} \left[\frac{\partial(u h)}{\partial x} + \frac{\partial(v h)}{\partial y} \right] dx dy = -\rho \int_{\text{column}} \nabla \cdot (h \mathbf{u}) dA, \quad (31.19)$$

thus recovering the result (31.15).

31.2.5 Bottom kinematic boundary condition

Kinematic boundary conditions arise from geometric constraints placed on the fluid system. We consider here the kinematic boundary condition at the bottom interface in the case where there is no flow through this interface, and follow up in Section 31.2.6 with the surface kinematic boundary condition.¹

The ocean bottom is located at a vertical position, $z = \eta_b(x, y)$. This location can equivalently be specified mathematically by the surface

$$s(x, y, z) = z - \eta_b(x, y) = 0. \quad (31.20)$$

The outward normal (pointing from the fluid into the rock) at this surface is given by

$$\hat{\mathbf{n}} = -\frac{\nabla s}{|\nabla s|} = \frac{\nabla \eta_b - \hat{\mathbf{z}}}{\sqrt{1 + \nabla \eta_b \cdot \nabla \eta_b}}. \quad (31.21)$$

¹From the discussion of fluid kinematics in Part III, we use the term *material surface* for any continuous surface or interface that is impenetrable to the flow of matter or thermal energy (mechanical energy can be transferred via pressure forces). In Section 15.4 we derived the kinematic boundary conditions for a fluid at interfaces. We here apply those ideas to the shallow water system, so that the presentation in Sections 31.2.5 and 31.2.6 offer a review of Section 15.4 as applied to a shallow water layer.

If the bottom is impenetrable to fluid then the velocity field is constrained to satisfy the no-normal flow boundary condition

$$\mathbf{v} \cdot \hat{\mathbf{n}} = 0 \quad \text{at } z = \eta_b. \quad (31.22)$$

Making use of the bottom outward normal (31.21) leads to

$$w = \mathbf{u} \cdot \nabla \eta_b \quad \text{at } z = \eta_b. \quad (31.23)$$

For a flat bottom, with $\nabla \eta_b = 0$, the no-normal flow condition means that $w(\eta_b) = 0$. For the case of nontrivial bottom topography, $w(\eta_b) = 0$ remains if flow occurs along lines of constant topography; i.e., along isobaths, in which case $\mathbf{u} \cdot \nabla \eta_b = 0$. But more generally, sloping bottoms lead to a nonzero vertical velocity component. Dynamically, a nonzero bottom vertical velocity arises from forces at the bottom that cause the horizontal velocity to cross isobaths, $\mathbf{u} \cdot \nabla \eta_b \neq 0$.

The kinematic result (31.23) is written in an Eulerian sense, with the velocity constrained to satisfy this relation at each point along the bottom interface. It has a dual material interpretation based on acknowledging that the bottom interface is a material surface. A fluid element on the bottom at $s = z - \eta_b = 0$ will thus remain there; it does not cross the bottom interface. Rather, it can at most move tangentially to the bottom.² We can ensure the no-normal flow constraint by setting

$$\frac{Ds}{Dt} = \frac{D(z - \eta_b)}{Dt} = 0 \quad \text{at } z = \eta_b. \quad (31.24)$$

Rearrangement of this result leads to the Eulerian constraint (31.23). Equivalently, we can write this boundary condition in the form

$$w = \frac{D\eta_b}{Dt} \quad \text{at } z = \eta_b. \quad (31.25)$$

Since $\eta_b = \eta_b(x, y)$, this expression of the kinematic boundary condition is identical to equation (31.23).

31.2.6 Surface kinematic boundary condition

We here assume the surface boundary is a material surface to derive the surface kinematic boundary condition, with Section 31.5 considering the slightly more general case of volume crossing this surface. As a material surface, the surface kinematic boundary condition follows analogously to the bottom. However, there is a fundamentally new feature in that the layer's upper free surface is a time dependent moving boundary. We studied such boundaries in Section 15.4.2 when detailing the kinematic boundary conditions for a material surface. We here review some of that discussion.

The free surface is located at a vertical position $z = \eta(x, y, t)$. Equivalently, the free surface can be specified by a surface of constant s , where

$$s(x, y, z, t) = z - \eta(x, y, t) = 0. \quad (31.26)$$

The outward normal to the free surface is thus given by

$$\hat{\mathbf{n}} = \frac{\nabla s}{|\nabla s|} = \frac{\hat{\mathbf{z}} - \nabla \eta}{\sqrt{1 + \nabla \eta \cdot \nabla \eta}}. \quad (31.27)$$

²Details of the tangential motion along a material boundary require dynamical information such as boundary stresses (see Chapter 19). We are not concerned with such dynamical information here, rather our concern is solely with kinematics.

We must account for motion of the surface when formulating the no-normal flow condition. To do so, write this no-normal flow condition as

$$(\mathbf{v} - \mathbf{v}^{(s)}) \cdot \hat{\mathbf{n}} = 0 \quad \text{at } z = \eta(x, y, t), \quad (31.28)$$

where $\mathbf{v}^{(s)}$ is the velocity of a point on the ocean surface. The velocity of a point fixed on an arbitrary surface with specified s satisfies

$$\frac{\partial s}{\partial t} + \mathbf{v}^{(s)} \cdot \nabla s = 0. \quad (31.29)$$

As defined, $\mathbf{v}^{(s)}$ advects a fluid element in a manner to always keep the element fixed on the constant s surface. With $\hat{\mathbf{n}} = \nabla s / |\nabla s|$, we have

$$\mathbf{v}^{(s)} \cdot \hat{\mathbf{n}} = -\frac{\partial_t s}{|\nabla s|} = \frac{\partial_t \eta}{\sqrt{1 + \nabla \eta \cdot \nabla \eta}}. \quad (31.30)$$

Hence, if the surface remains static, then $\mathbf{v}^{(s)} \cdot \hat{\mathbf{n}} = 0$. But more generally, the surface is moving, and that movement is fundamental to the surface kinematic boundary condition.

Making use of the result (31.30) in the no-normal flow constraint (31.28) then leads to the surface kinematic boundary condition

$$w - \mathbf{u} \cdot \nabla \eta = \frac{\partial \eta}{\partial t} \quad \text{at } z = \eta. \quad (31.31)$$

As for the bottom kinematic boundary condition written as (31.24), we can interpret the surface kinematic condition (31.31) materially, in which case

$$\frac{Ds}{Dt} = \frac{D(z - \eta)}{Dt} = 0 \quad \text{at } z = \eta. \quad (31.32)$$

That is, in the absence of flow across the surface boundary, that surface remains material. We can write this boundary condition in the equivalent form

$$w = \frac{D\eta}{Dt} = \frac{D\Delta\eta}{Dt} \quad \text{at } z = \eta. \quad (31.33)$$

31.2.7 Column stretching and vertical velocity

Since the fluid has constant density, we know that the velocity has zero divergence

$$\nabla \cdot \mathbf{u} + \partial_z w = 0 \implies \partial_z w = -\nabla \cdot \mathbf{u}. \quad (31.34)$$

This result also follows since material fluid elements in the constant density shallow water layer maintain a constant volume (see Section 17.1). Furthermore, since the horizontal velocity has no depth dependence, we can vertically integrate the continuity equation from the bottom to an arbitrary depth within the layer to render

$$w(z) = w(\eta_b) - (z - \eta_b) \nabla \cdot \mathbf{u}, \quad (31.35)$$

so that the vertical velocity is a linear function of depth. Applying this equation at the ocean surface yields

$$w(\eta) = w(\eta_b) - (\eta - \eta_b) \nabla \cdot \mathbf{u}. \quad (31.36)$$

Eliminating the horizontal convergence between equations (31.35) and (31.36) leads to

$$w(z) - w(\eta_b) = \left[\frac{z - \eta_b}{\eta - \eta_b} \right] [w(\eta) - w(\eta_b)]. \quad (31.37)$$

Making use of the surface kinematic boundary condition (31.33) and bottom kinematic boundary condition (31.25) renders the material form

$$\frac{1}{z - \eta_b} \left[\frac{D(z - \eta_b)}{Dt} \right] = \frac{1}{\eta - \eta_b} \left[\frac{D(\eta - \eta_b)}{Dt} \right]. \quad (31.38)$$

Finally, introducing the layer thickness $h = \eta - \eta_b$ yields the material conservation law

$$\frac{D}{Dt} \left[\frac{z - \eta_b}{h} \right] = 0. \quad (31.39)$$

Again, $h = \eta - \eta_b$ is the layer thickness and $z - \eta_b$ is the height of a fluid element from the bottom interface (see Figure 31.1). Consequently, equation (31.39) means that the ratio of the fluid element height above the bottom to the layer thickness remains constant as the fluid element moves through the shallow water fluid. That is, a column of shallow water fluid stretches or squeezes uniformly within a shallow water layer. Shallow water mechanics thus comprises the mechanics of coherent vertical fluid columns moving within a fluid layer. This constrained behaviour results from the linear depth dependence of the vertical velocity, which itself is a result of the depth independence of the horizontal velocity, and which follows from the depth independence of the horizontal gradient of hydrostatic pressure.

31.2.8 Tracer concentration equation

Suppose there is a material substance, a tracer, contained within a shallow water layer.³ We expect the tracer concentration, ψ , to have a non-uniform vertical structure within the layer, in addition to having horizontal structure: $\psi = \psi(x, y, z, t)$. If the tracer is advected through the layer without any diffusion, then the concentration satisfies the perfect fluid tracer equation

$$\frac{\partial \psi}{\partial t} + \mathbf{u} \cdot \nabla_z \psi + w \partial_z \psi = 0. \quad (31.40)$$

For a shallow water layer, where the horizontal velocity has no depth dependence within a layer, we find it sufficient to study the depth averaged tracer concentration within the layer,

$$C(x, y, t) \equiv \frac{1}{h} \int_{\eta_b}^{\eta} \psi(x, y, z, t) dz. \quad (31.41)$$

To develop the evolution equation for C , we vertically integrate the tracer equation (31.40) over the layer and make use of Leibniz's rule (Section 16.3.4)

$$\int_{\eta_b}^{\eta} \frac{\partial \psi}{\partial t} dz = \partial_t(h C) - \psi(\eta) \partial_t \eta \quad (31.42a)$$

$$\int_{\eta_b}^{\eta} \mathbf{u} \cdot \nabla \psi dz = \nabla \cdot (h C \mathbf{u}) - \psi(\eta) \mathbf{u}(\eta) \cdot \nabla \eta + \psi(\eta_b) \mathbf{u}(\eta_b) \cdot \nabla \eta_b - h C \nabla \cdot \mathbf{u} \quad (31.42b)$$

$$\int_{\eta_b}^{\eta} \psi \partial_z w dz = w(\eta) \psi(\eta) - w(\eta_b) \psi(\eta_b) - h C \partial_z w. \quad (31.42c)$$

³We developed the theory of material tracers in Chapter 16.

Use of the kinematic boundary conditions from Sections 31.2.5 and 31.2.6, and the three dimensional continuity equation, $\nabla \cdot \mathbf{u} + \partial_z w = 0$, renders the equation for the layer depth averaged tracer concentration

$$\frac{\partial(hC)}{\partial t} + \nabla \cdot (hC\mathbf{u}) = 0 \iff \frac{\partial C}{\partial t} + \mathbf{u} \cdot \nabla C = \frac{DC}{Dt} = 0. \quad (31.43)$$

As a self-consistency check, note that if the tracer concentration has a horizontally uniform value, then the Eulerian flux-form of the tracer equation (31.43) reduces to the thickness equation (31.16).

31.2.9 Summary comments

Key physical assumptions for the shallow water fluid

The shallow water fluid model is based on the following two assumptions.

- The fluid layer has a uniform density, which then means the fluid is incompressible.
- The pressure is hydrostatic.

Since the pressure gradient is depth-independent within the layer, an initial horizontal velocity that is depth independent will remain depth independent. That is, the fluid moves as columns within the layer, with the columns stretching and squashing depending on the divergence or convergence of column towards the column.

The term “shallow” refers to the small vertical to horizontal aspect ratio, $H/L \ll 1$, which in turn is consistent with the hydrostatic approximation (Section 23.3). The term “water” refers to the incompressible nature of the fluid, which is a more relevant approximation for water than for the atmosphere (see Section 24.1). Nonetheless, the shallow water model has direct applications to many features of both the atmosphere and ocean circulation, and as such it is widely used across the atmosphere and ocean sciences.

Shallow water fluid columns are not Taylor columns

The columnar motion of fluid within a shallow water layer is reminiscent of the Taylor columns discussed in Section 26.5.3. However, the columnar motion of fluid within a Taylor column is fundamentally distinct from that of a column of shallow water fluid. Motion in a Taylor column holds for homogeneous fluids undergoing rapid rotation and moving over a flat surface. The horizontal fluid velocity within a Taylor column is non-divergent so that there is no vertical motion of the fluid. These properties allow one to interpret a Taylor column as a fluid mechanical realization of a column of rigid matter much like a solid body.

In contrast, shallow water fluid columns do not rely on rotation, but instead arise from the hydrostatic balance maintained within each homogeneous layer. Additionally, shallow water columns are not rigid. Rather, they stretch and squash in the presence of nonzero divergence in the horizontal flow, thus leading to vertical motion of fluid within the column. Finally, shallow water columns remain coherent even as they move over topography, and yet, again, they can stretch and squash.

31.3 Reduced gravity model for the upper ocean

The *reduced gravity model* describes an active layer of uniform density, ρ_1 , above a stagnant layer of density, ρ_2 , and below a fluid of zero density, $\rho_0 = 0$. It is often referred to as the 1.5 layer model. This theoretical model has been used, to some success, as an idealization of the upper ocean circulation whereby an active layer (e.g., the region above the ocean pycnocline), sits above

an inactive layer (the abyss) of zero motion. In this way, we introduce the *level of no motion*, below which (baroclinic) currents vanish.

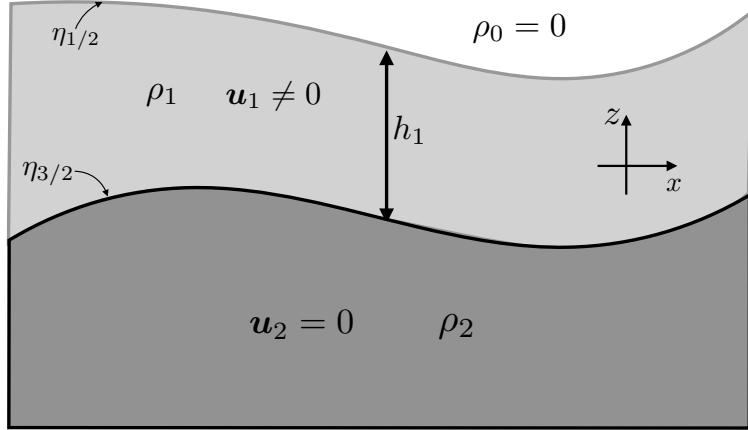


FIGURE 31.4: Reduced gravity model of shallow water fluid. The lower layer with density ρ_2 is dynamically inactive and thus has a zero velocity. The upper layer is dynamically active with thickness $h = \eta_{1/2} - \eta_{3/2}$ and density ρ_1 . The dynamically active layer is bounded above by a zero density layer, $\rho_0 = 0$. The lower inactive layer is assumed to be infinitely deep so that its continuity equation can be ignored; i.e., even though there are zero currents within this layer it can still undulate.

31.3.1 Momentum and thickness equations for the active layer

We develop the momentum equations for the reduced gravity model by making use of the hydrostatic balance, in which pressure at a depth, z , in the upper layer is computed as (see Figure 31.4)

$$p_1(x, y, z, t) = p_{1/2}(x, y, t) + g \rho_1 [\eta_{1/2}(x, y, t) - z], \quad (31.44)$$

where we denote an interface value by a half-index, so that $\eta_{1/2}$ and $p_{1/2}$ are the interface height and pressure at the upper layer interface.⁴ Since the fluid above the upper layer is assumed to have zero density, we set

$$p_{1/2} = 0. \quad (31.45)$$

We immediately see that the horizontal momentum equation for the upper (active) layer is written

$$\frac{D\mathbf{u}_1}{Dt} + f \hat{\mathbf{z}} \wedge \mathbf{u}_1 = -g \nabla \eta_{1/2}. \quad (31.46)$$

The equations for the upper layer are completed by volume conservation in the form of the thickness equation

$$\frac{Dh_1}{Dt} = -h_1 \nabla \cdot \mathbf{u}_1. \quad (31.47)$$

31.3.2 Relating undulations of the top and bottom layer interfaces

The pressure in the lower stagnant layer is given by the weight per horizontal area of fluid above it

$$p_2(x, y, z, t) = g \rho_1 (\eta_{1/2} - \eta_{3/2}) + g \rho_2 (\eta_{1/2} - z). \quad (31.48)$$

⁴There is no general consensus on this notation, with some treatments, such as [Vallis \(2017\)](#) and [Cushman-Roisin and Beckers \(2011\)](#) using an integer to label an interface quantity, whereas some numerical model papers (e.g., [Griffies et al. \(2020\)](#)) use the half-index. We prefer the half-index since it removes any ambiguity concerning the ordering of the interfaces relative to the layer.

However, for the reduced gravity model we assume the lower layer is motionless. To maintain zero motion in the lower layer requires the horizontal pressure gradient in this layer to vanish

$$\nabla_z p_2 = g \rho_1 \nabla(\eta_{1/2} - \eta_{3/2}) + g \rho_2 \nabla \eta_{1/2} = 0. \quad (31.49)$$

This relation provides a constraint that links undulations of the top and bottom interfaces of the dynamically active layer

$$\eta_{1/2} = -\eta_{3/2} \left[\frac{\rho_2 - \rho_1}{\rho_1} \right] + \text{constant}. \quad (31.50)$$

The density ratio on the right hand side is positive but typically much smaller than unity. Hence, the relation (31.50) means that undulations of the free surface, $\eta_{1/2}$, are of opposite sign and of much smaller amplitude than undulations in the lower interface, $\eta_{3/2}$. This behavior is typical for undulations of the pycnocline region of the ocean and the free surface as depicted in Figure 31.5.

31.3.3 Momentum equation with reduced gravity

Relation (31.50) can be used to write the momentum equation (31.46) in the form

$$\frac{D\mathbf{u}_1}{Dt} + f \hat{\mathbf{z}} \wedge \mathbf{u}_1 = +g_{3/2}^r \nabla \eta_{3/2}, \quad (31.51)$$

where

$$g_{3/2}^r = g \left[\frac{\rho_2 - \rho_1}{\rho_1} \right] \ll g \quad (31.52)$$

defines the *reduced gravity* between layer-1 and layer-2. We use the half-integer notation for the reduced gravity since it multiplies the gradient of the interface height, $\eta_{3/2}$. It is common to make use of the momentum equation in the form (31.51), rather than the original form (31.46). The reason is that historically, ocean hydrography measurements⁵ have allowed for an estimate of the pycnocline slope, $\nabla \eta_{3/2}$, whereas it was not until precise satellite altimetry measurements starting in the 1990s that we could estimate the far smaller sea level slope, $\nabla \eta_{1/2}$.

31.3.4 Some intuition about reduced gravity

We here offer some words about the reduced gravity. For that purpose, return to relation (31.50), take the spatial gradient and multiply by the gravitational acceleration to yield

$$g \nabla \eta_{1/2} = -g^r \nabla \eta_{3/2}. \quad (31.53)$$

As noted above, since $g^r \ll g$, the free surface slopes are much smaller than interior slopes: $|\nabla \eta_{1/2}| \ll |\nabla \eta_{3/2}|$. We can understand the large slope differences by recognizing that the gravitational acceleration in the interior is partially compensated by buoyancy forces (Chapter 25). Hence, the interior interface is afforded less resistance to undulations. In contrast, the upper interface, $\eta_{1/2}$, directly contacts the atmosphere so that the corresponding reduced gravity felt by the upper interface is much larger. In fact, for the present discussion the upper interface's reduced gravity equals to g since we assume the atmosphere has zero mass. Even with a massive atmosphere, the upper interface's reduced gravity is close to g since the atmosphere is roughly 1000 times less dense than seawater. With a larger reduced gravity, the upper interface experiences more resistance to motion than the interior interface so that its slope has a much smaller magnitude.

⁵In oceanography, hydrography refers to measurements of temperature, salinity, and pressure; see [Talley et al. \(2011\)](#).

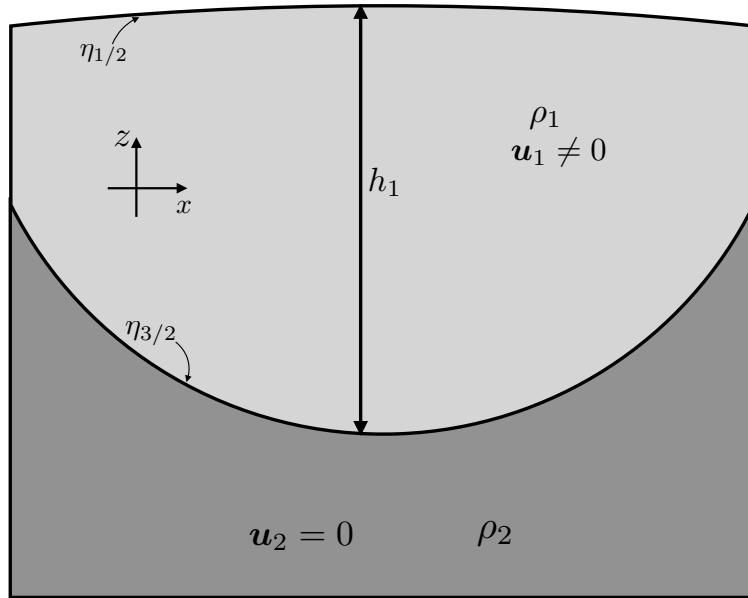


FIGURE 31.5: A vertical slice through a reduced gravity, or 1.5 layer, ocean in hydrostatic balance. Shown here is a plug of dynamically active light water, as may occur in a warm core eddy to the subtropical gyres, sitting on top of heavy water of zero motion. The free surface corresponds to $\eta_{1/2}$ in Figure 31.4, whereas the pycnocline (heavy black line) corresponds to the lower interface $\eta_{3/2}$ of Figure 31.4. The sea surface experiences an applied pressure $p = p_a$, assumed to be uniform for this idealized situation. Note how sea level is a maximum above the pycnocline minimum, with this geometry reflected in equation (31.50). In the ocean, the slope of the pycnocline is about 100-300 times larger than the slope of the sea level. That is, sea level may show undulations on the order of a metre, whereas the pycnocline undulations are on the order of 100-300 m.

31.3.5 Further study

The material in this section is a summary of that in Section 3.2 of [Vallis \(2017\)](#). [Tomczak and Godfrey \(1994\)](#) make use of the reduced gravity model for interpreting aspects of the observed ocean. Additional use is made by [Griffies et al. \(2014\)](#) for interpreting patterns of sea level in the ocean.

31.4 Stacked shallow water equations

In studies of shallow water fluids, much of the formalism developed for a single layer can be readily extended to an arbitrary number of layers. We here pursue this extension and to thereby expose the underlying kinematics and dynamics of stacked shallow water models. We assume the layers are immiscible so that matter and thermal properties are not exchanged between the layers. Instead, the layers couple only through mechanical forces arising from pressure. The notation for our derivations is depicted in Figure 31.6 in the case of three active layers.

In Section 46.1 we develop the equations for a continuously stratified Boussinesq fluid making use of isopycnal vertical coordinates. Although the vertical stratification is continuous in that case, we see in that discussion that the isopycnal equations are isomorphic to the stacked shallow water equations. Hence, besides being of intrinsic interest as a versatile theoretical model, the stacked shallow water model offers a suitable step towards studies of a continuously stratified fluid using isopycnal coordinates.

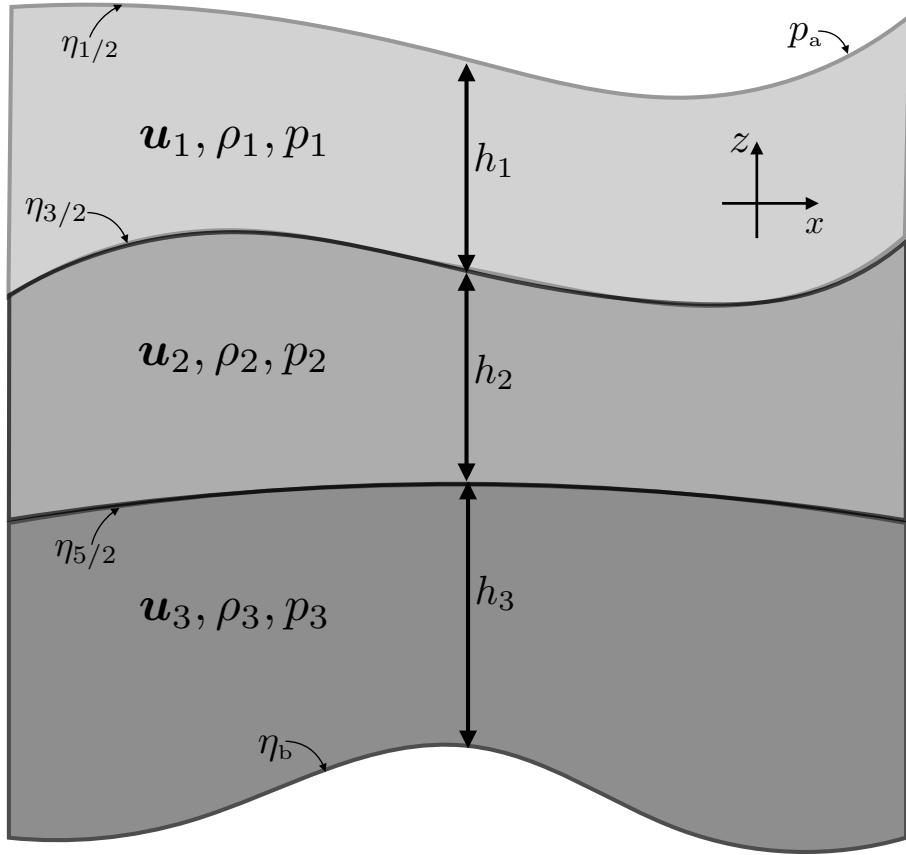


FIGURE 31.6: Three dynamically active layers of stacked shallow water fluid. The notation corresponds to that for the reduced gravity model of Figure 31.4, yet here with three dynamically active layers. The “atmosphere” above the layers is assumed to apply a pressure, p_a , to the upper surface, and the total thickness of a column is $h_1 + h_2 + h_3 = \eta_{1/2} - \eta_b$.

31.4.1 2-layer model formulation

We here display the equations for two layers, thus offering the seeds for an extension to N layers in Section 31.4.2.

Thickness and tracer equations

Each shallow water layer satisfies its own independent thickness equation and tracer equation, representing the conservation of volume and tracer content for each layer

$$\frac{\partial h_1}{\partial t} + \nabla \cdot (h_1 \mathbf{u}_1) = 0 \quad (31.54)$$

$$\frac{\partial h_2}{\partial t} + \nabla \cdot (h_2 \mathbf{u}_2) = 0 \quad (31.55)$$

$$\frac{\partial(h_1 C_1)}{\partial t} + \nabla \cdot (h_1 C_1 \mathbf{u}_1) = 0 \quad (31.56)$$

$$\frac{\partial(h_2 C_2)}{\partial t} + \nabla \cdot (h_2 C_2 \mathbf{u}_2) = 0. \quad (31.57)$$

We emphasize that there is no explicit coupling between these equations, as each layer separately must satisfy volume conservation and tracer conservation. However, the velocities are coupled through the pressure force, as we now discuss.

Pressure within a layer

To compute the pressure within a layer, we make use of the hydrostatic balance and integrate down from the surface, which results in the pressure fields

$$p_1 = \rho_1 g (\eta_{1/2} - z) + p_a \quad (31.58)$$

$$p_2 = \rho_1 g (\eta_{1/2} - \eta_{3/2}) + \rho_2 g (\eta_{3/2} - z) + p_a. \quad (31.59)$$

It is convenient to write pressure in layer-two using the reduced gravity, which leads to

$$p_2 - p_a = \rho_1 g (\eta_{1/2} - \eta_{3/2}) + \rho_2 g (\eta_{3/2} - z) \quad (31.60a)$$

$$= g \eta_{3/2} (\rho_2 - \rho_1) + g \rho_1 \eta_{1/2} - g \rho_2 z \quad (31.60b)$$

$$= \rho_1 \left[g \eta_{1/2} + g \eta_{3/2} \frac{\rho_2 - \rho_1}{\rho_1} \right] - g \rho_2 z \quad (31.60c)$$

$$= \rho_1 (g_{1/2}^r \eta_{1/2} + g_{3/2}^r \eta_{3/2}) - g \rho_2 z, \quad (31.60d)$$

with the reduced gravities given by

$$g_{1/2}^r = g \left[\frac{\rho_1 - \rho_0}{\rho_1} \right] \approx g \quad \text{and} \quad g_{3/2}^r = g \left[\frac{\rho_2 - \rho_1}{\rho_1} \right], \quad (31.61)$$

where we assumed that $\rho_0 \ll \rho_1$ so that the reduced gravity $g_{1/2}^r \approx g$.

Horizontal velocity equations

The horizontal velocity equations for the two layers take the form

$$\rho_1 \left[\frac{D_1 \mathbf{u}_1}{Dt} + f \hat{\mathbf{z}} \wedge \mathbf{u}_1 \right] = -\nabla p_1 \quad (31.62a)$$

$$\rho_2 \left[\frac{D_2 \mathbf{u}_2}{Dt} + f \hat{\mathbf{z}} \wedge \mathbf{u}_2 \right] = -\nabla p_2, \quad (31.62b)$$

where we introduced the layer material time derivatives

$$\frac{D_k}{Dt} = \frac{\partial}{\partial t} + \mathbf{u}_k \cdot \nabla \quad \text{for } k = 1, 2. \quad (31.63)$$

Making use of expressions (31.58) and (31.60d) for layer pressures leads to the horizontal momentum equations

$$\frac{D_1 \mathbf{u}_1}{Dt} + f \hat{\mathbf{z}} \wedge \mathbf{u}_1 = -g \nabla [\eta_{1/2} + p_a / (g \rho_1)] \quad (31.64a)$$

$$\frac{D_2 \mathbf{u}_2}{Dt} + f \hat{\mathbf{z}} \wedge \mathbf{u}_2 = -\frac{\rho_1}{\rho_2} \nabla (g_{1/2}^r \eta_{1/2} + g_{3/2}^r \eta_{3/2} + p_a / \rho_1). \quad (31.64b)$$

Finally, it is convenient to express pressure in terms of layer thicknesses, h_1 and h_2 , since the layer thicknesses are the prognostic fields determined by time stepping the thickness equations (31.54) and (31.55). We thus write

$$\eta_{1/2} = \eta_b + h_1 + h_2 \quad \text{and} \quad \eta_{3/2} = \eta_b + h_2, \quad (31.65)$$

so that

$$p_1 = \rho_1 g (\eta_b + h_1 + h_2) + p_a - g \rho_1 z \quad (31.66a)$$

$$p_2 = \rho_1 \left[g_{1/2}^r (\eta_b + h_1 + h_2) + g_{3/2}^r (\eta_b + h_2) \right] + p_a - g \rho_2 z, \quad (31.66b)$$

thus resulting in the horizontal momentum equations

$$\frac{D_1 \mathbf{u}_1}{Dt} + f \hat{\mathbf{z}} \wedge \mathbf{u}_1 = -g \nabla [\eta_b + h_1 + h_2 + p_a / (g \rho_1)] \quad (31.67a)$$

$$\frac{D_2 \mathbf{u}_2}{Dt} + f \hat{\mathbf{z}} \wedge \mathbf{u}_2 = -\frac{\rho_1}{\rho_2} \nabla \left[g_{1/2}^r (\eta_b + h_1 + h_2) + g_{3/2}^r (\eta_b + h_2) + p_a / \rho_1 \right]. \quad (31.67b)$$

Notice how layer thickness from one layer is coupled to the other layer through the pressure gradient. In this way, changes in the thickness of one layer have a direct impact on pressure forces and flow in the adjacent layer. Finally, the Boussinesq approximation sets the density ratio ρ_1/ρ_2 to unity so that

$$\frac{D_1 \mathbf{u}_1}{Dt} + f \hat{\mathbf{z}} \wedge \mathbf{u}_1 = -g \nabla [\eta_b + h_1 + h_2 + p_a / (g \rho_1)] \quad (31.68a)$$

$$\frac{D_2 \mathbf{u}_2}{Dt} + f \hat{\mathbf{z}} \wedge \mathbf{u}_2 = -\nabla \left[g_{1/2}^r (\eta_b + h_1 + h_2) + g_{3/2}^r (\eta_b + h_2) + p_a / \rho_1 \right]. \quad (31.68b)$$

Vertical shear in horizontal velocities

The difference in layer velocities, $\mathbf{u}_1 - \mathbf{u}_2$, represents the vertical shear in the layers. Assuming $g_{1/2}^r = g$ reveals that this difference is affected by a pressure gradient arising just from bottom topography and the interior layer thickness, h_2

$$\frac{D_1 \mathbf{u}_1}{Dt} - \frac{D_2 \mathbf{u}_2}{Dt} + f \hat{\mathbf{z}} \wedge (\mathbf{u}_1 - \mathbf{u}_2) = g_{3/2}^r \nabla (\eta_b + h_2). \quad (31.69)$$

That is, the vertical shear does not directly feel undulations of the free surface, $\eta_{1/2}$, or the applied pressure, p_a . Rather, it feels these surface undulations only indirectly via nonlinear terms appearing in the advection on the left hand side. We further discuss this result in Section 32.2.2 by introducing thermal wind and the Margules' Relation.

31.4.2 *N*-layer model equations

The 2-layer equations (31.68a) and (31.68b) can be readily generalized to N -layers. The thickness equation and tracer equation represent volume and tracer conservation for each layer

$$\frac{\partial h_k}{\partial t} + \nabla \cdot (\mathbf{u}_k h_k) = 0 \quad (31.70a)$$

$$\frac{\partial (h_k C_k)}{\partial t} + \nabla \cdot (\mathbf{u}_k h_k C_k) = 0, \quad (31.70b)$$

where $k = 1, N$ is the discrete layer index and there is no implied summation on this label.⁶

⁶To help distinguish the layer index, k , from a tensor index, k , we write the layer index using an upright Roman font whereas a tensor index is slanted.

Some work is needed to generalize the pressure gradient appearing in the velocity equation. For that purpose, write the interface height, $\eta_{k+1/2}$, as

$$\eta_{k+1/2} = \eta_b + \sum_{j=k+1}^N h_j \quad \text{with } \eta_{N+1/2} = \eta_b \text{ and } \eta_{1/2} = \eta. \quad (31.71)$$

For example, the layer interfaces with $N = 3$ layers are given by

$$\eta_{1/2} = \eta_b + h_1 + h_2 + h_3 \quad \eta_{3/2} = \eta_b + h_2 + h_3 \quad \eta_{5/2} = \eta_b + h_3 \quad \eta_{7/2} = \eta_b. \quad (31.72)$$

In turn, the hydrostatic pressure within layer- k is given by

$$p_k = -g \rho_k z + p_a + \rho_1 \sum_{j=0}^{k-1} g_{j+1/2}^r \eta_{j+1/2} = p_{k-1/2} + g \rho_k (\eta_{k-1/2} - z), \quad (31.73)$$

where the reduced gravities are

$$g_{j+1/2}^r = g \frac{\rho_{j+1} - \rho_j}{\rho_1} > 0 \quad \text{with} \quad g_{1/2}^r = g \quad \text{if} \quad \rho_0 = \rho_a = 0, \quad (31.74)$$

with ρ_a the density of the overlying atmosphere. The half-integer pressures are evaluated on the corresponding interface with the hydrostatic balance yielding the pressure difference

$$p_{k+1/2} - p_{k-1/2} = g \rho_k h_k = g \rho_k (\eta_{k-1/2} - \eta_{k+1/2}). \quad (31.75)$$

Since the terms $g \rho_k z$ appearing in the layer pressures have zero horizontal gradient, they do not contribute to horizontal accelerations. Hence, in some treatments it can be useful to make use of the *dynamic pressures*

$$p_k^{\text{dyn}} \equiv \rho_1 \sum_{j=0}^{k-1} g_{j+1/2}^r \eta_{j+1/2} \implies \nabla p_k = \nabla p_k^{\text{dyn}}. \quad (31.76)$$

The corresponding velocity equation for an arbitrary layer- k is given by

$$\frac{D_k \mathbf{u}_k}{Dt} + f \hat{z} \wedge \mathbf{u}_k = -\nabla p_k / \rho_1 \iff [\partial_t + (\mathbf{u}_k \cdot \nabla)] \mathbf{u}_k + f \hat{z} \wedge \mathbf{u}_k = -\nabla p_k / \rho_1. \quad (31.77)$$

31.5 Non-conservative processes and boundary stresses

Much of our focus on shallow water mechanics concerns the adiabatic case in which the layers are immiscible and there are no frictional processes. Even so, we will find occasion to consider non-conservative processes, such as the transfer of volume between layers via dia-surface fluxes, as well as irreversible momentum transfer associated with friction and boundary stresses. In this section we consider these processes for the single shallow water layer, with an extension to stacked shallow water straightforward.

31.5.1 Dia-surface volume fluxes

Consider the case of dia-surface transfer as occurs across the ocean surface through evaporation, precipitation, and river runoff, or as occurs for interior layers in the presence of irreversible mixing processes. Let $w^{(\eta)}$ measure the volume per time per horizontal area of fluid entering across the

surface interface of the shallow water layer, with $w^{(\dot{\eta})}$ having dimensions of length per time and a sign convention so that $w^{(\dot{\eta})} > 0$ means volume enters the shallow water layer. For example, $w^{(\dot{\eta})} > 0$ could represent the addition of rain to the surface layer.⁷ To keep the analysis simple, assume the mass contained in $w^{(\dot{\eta})} \neq 0$ has the same density and same velocity as the shallow water layer. Hence, there is no modification to the layer density. This assumption must be relaxed for realistic layered models.

Surface kinematic boundary condition

The kinematic boundary condition (31.32) expresses the material nature of the free surface in the absence of boundary volume flux. In the presence of a non-zero boundary flux, $w^{(\dot{\eta})} \neq 0$, we follow the formulation of the kinematic boundary condition for the ocean free surface with a mass flux (equation (15.79)). This result renders the boundary condition

$$\frac{D(z - \eta)}{Dt} = -w^{(\dot{\eta})} \quad \text{at } z = \eta. \quad (31.78)$$

In effect, this relation defines the surface boundary flux. This boundary condition can be written in the equivalent form

$$\frac{D\eta}{Dt} = w(\eta) + w^{(\dot{\eta})}. \quad (31.79)$$

Stretching of a vertical column

Equation (31.39) shows that in the absence of volume sources, a column of shallow water fluid stretches or squeezes uniformly. What happens when $w^{(\dot{\eta})} \neq 0$? Following through the derivation of the thickness equation in Section 31.2.4, we are trivially led to the following generalization

$$\frac{\partial h}{\partial t} + \nabla \cdot (h \mathbf{u}) = w^{(\dot{\eta})}, \quad (31.80)$$

or equivalently

$$\frac{Dh}{Dt} = -h \nabla \cdot \mathbf{u} + w^{(\dot{\eta})}. \quad (31.81)$$

Equation (31.37) is derived through assuming the horizontal velocity has no depth dependence within a shallow water layer, and the fluid has a three-dimensional non-divergent velocity. It holds whether the ocean surface is material or non-material, and thus forms a suitable starting point for generalization to the case of $w^{(\dot{\eta})} \neq 0$

$$w(z) - w(\eta_b) = \left[\frac{z - \eta_b}{\eta - \eta_b} \right] [w(\eta) - w(\eta_b)]. \quad (31.82)$$

We now make use of the new surface kinematic boundary condition (31.79) and the unmodified bottom kinematic boundary condition (31.25) to write the material form

$$\frac{1}{z - \eta_b} \left[\frac{D(z - \eta_b)}{Dt} \right] = \frac{1}{\eta - \eta_b} \left[\frac{D(\eta - \eta_b)}{Dt} - w^{(\dot{\eta})} \right]. \quad (31.83)$$

Rearrangement thus leads to the material conservation statement

$$\left[\frac{z - \eta_b}{h} \right]^{-1} \frac{D}{Dt} \left[\frac{z - \eta_b}{h} \right] = -\frac{w^{(\dot{\eta})}}{h}, \quad (31.84)$$

⁷This sign convention follows that used for water mass transformation as developed in Chapter 52 and defined by equation (52.38).

where $h = \eta - \eta_b$ is the layer thickness. This result can be written in the more concise form

$$\frac{D}{Dt} \left[\ln \left(\frac{z - \eta_b}{h} \right) \right] = -\frac{w^{(\dot{\eta})}}{h}. \quad (31.85)$$

Either way, we see that in the presence of a surface boundary volume flux, a column of shallow water fluid no longer stretches or squeezes uniformly. Instead, for $w^{(\dot{\eta})} > 0$, a fluid parcel moves down within the column as more water is added to the top of the layer.

31.5.2 Frictional and boundary stresses

As discussed in Chapter 19, accelerations from friction appear in the momentum equation via the divergence of the frictional stress tensor. For a shallow water fluid we may choose to include a frictional stress proportional to lateral shears within each layer much like the friction operators discussed in Section 19.7. Additionally, we may need to include boundary stresses such as those from winds and bottom friction. To incorporate boundary stresses, consider the Boussinesq form (Section 24.5.2) of the vertical stress divergence appearing in the momentum equation as an acceleration

$$\mathbf{F} = \frac{1}{\rho_0} \frac{\partial \boldsymbol{\tau}}{\partial z}, \quad (31.86)$$

where $\boldsymbol{\tau}$ is the horizontal stress imparted to the layer interface due to turbulent boundary layer processes. Since a shallow water layer is homogeneous, we can only apply this boundary stress uniformly throughout the layer (as a body stress) and so it appears as

$$\mathbf{F} = \frac{1}{\rho_0} \frac{\boldsymbol{\tau}_{k-1/2} - \boldsymbol{\tau}_{k+1/2}}{h_k}, \quad (31.87)$$

where h_k is the layer thickness, $\boldsymbol{\tau}_{k-1/2}$ is the stress at the upper interface (e.g., wind stress) and $\boldsymbol{\tau}_{k+1/2}$ is the stress at the lower interface (e.g., bottom drag).



31.6 Exercises

EXERCISE 31.1: RELATIONS FOR VERTICAL VELOCITY (EXERCISE (3.2) OF *Vallis (2006)*)

Show that the vertical velocity within a shallow water system is given by

$$w = \left[\frac{z - \eta_b}{h} \right] \frac{Dh}{Dt} + \frac{D\eta_b}{Dt}. \quad (31.88)$$

Interpret the result, showing that it gives sensible answers at the top and bottom of the fluid layer.

EXERCISE 31.2: NON-ROTATING HYDRAULIC CONTROL

Consider the steady flow in a non-rotating shallow water layer where the flow is purely one-dimensional in the zonal direction.

- (a) Show that the steady flow satisfies the balance

$$\partial_x h [1 - Fr^2] = \partial_x \eta_b \quad (31.89)$$

where the Froude number is given by

$$\text{Fr} = \frac{u}{\sqrt{gh}}. \quad (31.90)$$

The Froude number is the ratio of the speed for a fluid particle to the speed of a shallow water gravity wave.

- (b) Discuss the case in which $\partial_x \eta_b = 0$ yet $\partial_x h \neq 0$. This case is known as *hydraulic control*.

EXERCISE 31.3: SHALLOW WATER EQUATIONS WITH TIDES

In Chapter 29 we derive the equations for a primitive equation ocean in the presence of astronomical forcing that leads to tides. Specialize the general results from that chapter to derive the thickness and momentum equations for a single layer of shallow water fluid in the presence of astronomical tidal forcing. As in Section 29.4, assume the perturbation geopotential is depth independent.

EXERCISE 31.4: INVERTED REDUCED GRAVITY MODEL

Derive the shallow water equations for a single moving layer of fluid of density ρ_2 above a rigid floor, with this moving layer below a stagnant fluid of density ρ_1 , with $\rho_1 < \rho_2$. Show that as $\rho_1/\rho_2 \rightarrow 0$ the single layer shallow water equations emerge. Make use of notation from the three-layer system shown in Figure 31.6. This model might be used to study flow in the atmosphere well above the boundary layer, or the abyssal ocean well below the pycnocline.



Shallow water dynamics

In this chapter we further our dynamical understanding of the shallow water fluid model. Our study includes geostrophy, thermal wind (as expressed by Margules' relation), momentum budgets, form stress, mechanical energy, and available potential energy. We also develop three case studies: the steady force balance in a zonally reentrant channel, the angular momentum dynamics of a rotating tank of shallow water fluid, and linear gravity waves in a single shallow water layer.

READER'S GUIDE TO THIS CHAPTER

This chapter builds from the formulations in Chapter 31 as well as the geostrophic mechanics from Chapter 26 and the pressure form stress from Chapter 20. We make use of the dynamical results in this chapter for many of the subsequent chapters. Notationally, we follow the same convention for the gradient operator noted at the start of Chapter 31.

32.1	Loose threads	696
32.2	Geostrophic balance and thermal wind	696
32.2.1	Geostrophy for a single layer	697
32.2.2	Geostrophy and thermal wind for two layers	697
32.2.3	Geostrophic transport within layers	698
32.2.4	Flow within a geostrophic eddy	699
32.2.5	Comments	700
32.3	Thickness weighted momentum equation	701
32.3.1	Single layer equations	701
32.3.2	Geostrophic and ageostrophic contributions	702
32.3.3	Form stresses acting on a shallow water column	703
32.3.4	Comments on the two pressure force formulations	705
32.4	Contact pressure forces in shallow water layers	705
32.4.1	Pressure contact force and pressure body force	706
32.4.2	N -layer equations	706
32.4.3	Contact pressure force along vertical sides	706
32.4.4	Contact pressure force along the top and bottom interfaces	708
32.4.5	Form stress	709
32.4.6	Net contact pressure force on a shallow water column	709
32.4.7	Contact pressure force summed over all layers	710
32.4.8	Horizontal pressure force with potential energy gradients	710
32.4.9	Momentum equation with contact pressure forces	711
32.4.10	Further reading	714
32.5	Mechanical energy	714
32.5.1	Gravitational potential energy	714

32.5.2	Kinetic energy	715
32.5.3	Mechanical energy	715
32.5.4	Available potential energy	716
32.6	Steady momentum balance in a zonal channel	717
32.6.1	Volume transport for steady flow	717
32.6.2	Steady meridional balances	718
32.6.3	Steady zonal balance	719
32.6.4	The role of frictional bottom drag	719
32.6.5	Correlation between surface height and topographic slope	720
32.6.6	Connection to meridional geostrophic transport	721
32.6.7	Sinusoidal example	721
32.6.8	Comments	723
32.7	Angular momentum in a rotating tank	724
32.7.1	Angular momentum for a column of shallow water fluid	724
32.7.2	Material time evolution of the angular momentum	725
32.7.3	Materially invariant angular momentum	726
32.7.4	Comments	726
32.8	Inertia-gravity waves in a single shallow water layer	726
32.8.1	Linearized shallow water equations	726
32.8.2	Relative vorticity of linear fluctuations	727
32.8.3	Non-rotating gravity waves	728
32.8.4	Potential vorticity of linear fluctuations	730
32.8.5	Inertia-gravity waves	730
32.8.6	Inertia-gravity wave dispersion relation	731
32.8.7	Horizontally longitudinal and transverse components	733
32.8.8	Comments	733
32.9	Inertia-gravity waves generated by topography	734
32.9.1	Linearized equations	734
32.9.2	Galilean transformation	734
32.9.3	Wave equation and relation between amplitudes	735
32.10	Exercises	736

32.1 Loose threads

- More schematics
- Work through a two-layer example for the zonal channel in Section 32.6 to describe the features exhibited by Olbers and Hughes in their schematic of the Southern Ocean circulation. Discuss topographic form stress.
- Two-layer gravity waves and topographic forcing to show that with two layers one gets a phase shift and thus a nonzero topographic form stress.

32.2 Geostrophic balance and thermal wind

As described in Chapter 26, geostrophic balance arises from neglecting the material time derivative in the inviscid horizontal momentum equation, which is a sensible assumption when the Rossby number is small. The resulting balance between Coriolis and pressure accelerations constitutes the geostrophic balance. We consider here the implications of geostrophy for one and two-layer shallow water systems.

32.2.1 Geostrophy for a single layer

Ignoring atmospheric pressure ($p_a = 0$) leads to the geostrophic balance for a single shallow water layer

$$f \hat{z} \wedge \mathbf{u}_g = -g \nabla \eta \implies f \mathbf{u}_g = g \hat{z} \wedge \nabla \eta, \quad (32.1)$$

or in component form

$$u_g = -\frac{g}{f} \frac{\partial \eta}{\partial y} \quad \text{and} \quad v_g = \frac{g}{f} \frac{\partial \eta}{\partial x}. \quad (32.2)$$

Consequently, the shallow water layer geostrophic current is balanced by the gradient of the free surface (sea level). In the northern hemisphere, where $f > 0$, geostrophic shallow water currents flow counter-clockwise around negative sea level anomalies (low pressure) and clockwise around positive sea level anomalies (high pressure). The opposite orientation holds in the southern hemisphere, where $f < 0$. Figure 32.1 shows a schematic of geostrophic balance for a single shallow water layer.

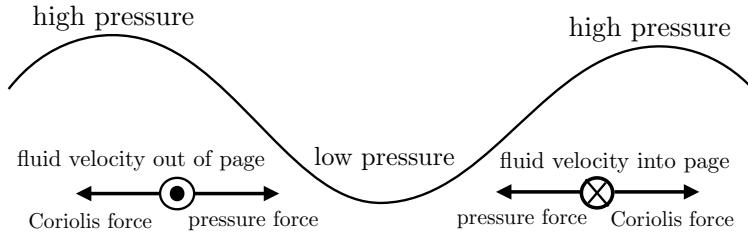


FIGURE 32.1: Side view of geostrophic balance for a single shallow water layer, here shown with two high pressure centers surrounding a low pressure center. The Coriolis force balances the pressure gradient force. In the northern hemisphere, where $f > 0$, geostrophic flow is counter-clockwise around a low pressure center and clockwise around a high pressure center.

32.2.2 Geostrophy and thermal wind for two layers

Now consider two shallow water layers as in Figure 31.6. Recall the layer pressure equations (31.58) and (31.59), which leads to the pressure difference

$$p_1 - p_2 = g \eta_{3/2} (\rho_1 - \rho_2) + g z (\rho_2 - \rho_1) = g_{3/2}^r \rho_1 (z - \eta_{3/2}), \quad (32.3)$$

where the reduced gravity is given by

$$g_{3/2}^r = g \left[\frac{\rho_2 - \rho_1}{\rho_1} \right] > 0. \quad (32.4)$$

The density difference $\rho_2 - \rho_1$ is generally much smaller than either density, so that $g_{3/2}^r \ll g$. For a Boussinesq shallow water system, the momentum equations are given by

$$\frac{D_1 \mathbf{u}_1}{Dt} + f \hat{z} \wedge \mathbf{u}_1 = -\rho_1^{-1} \nabla p_1 \quad (32.5a)$$

$$\frac{D_2 \mathbf{u}_2}{Dt} + f \hat{z} \wedge \mathbf{u}_2 = -\rho_1^{-1} \nabla p_2, \quad (32.5b)$$

where we used the top layer density as the Boussinesq reference density, and where we introduced the material time derivatives for each layer

$$\frac{D_n}{Dt} = \frac{\partial}{\partial t} + \mathbf{u}_n \cdot \nabla. \quad (32.6)$$

Making use of the pressure difference (32.3) renders

$$\frac{D_1 \mathbf{u}_1}{Dt} - \frac{D_2 \mathbf{u}_2}{Dt} + f \hat{\mathbf{z}} \wedge \Delta \mathbf{u} = -\rho_1^{-1} \nabla(p_1 - p_2) = g_{3/2}^r \nabla \eta_{3/2}, \quad (32.7)$$

where

$$\Delta \mathbf{u} = \mathbf{u}_1 - \mathbf{u}_2 \quad (32.8)$$

is the vertical difference of the layer horizontal velocities. We see that the difference in the geostrophic velocities for the two layers is proportional to the slope of the interface between the two layers

$$f \hat{\mathbf{z}} \wedge \Delta \mathbf{u}_g = g_{3/2}^r \nabla \eta_{3/2}, \quad (32.9)$$

which is equivalent to

$$f \Delta \mathbf{u}_g = -\hat{\mathbf{z}} \wedge g_{3/2}^r \nabla \eta_{3/2} \implies \Delta u_g = +\frac{g_{3/2}^r}{f} \frac{\partial \eta_{3/2}}{\partial y} \quad \text{and} \quad \Delta v_g = -\frac{g_{3/2}^r}{f} \frac{\partial \eta_{3/2}}{\partial x}. \quad (32.10)$$

These equations represent the *Margules' relation*. It applies at any interface between two shallow water fluid layers. It says that the vertical difference between the layer geostrophic velocities is proportional to the interface slope. When the slope is large, the vertical difference in the geostrophic velocity is large. Also, the velocity difference is large when the reduced gravity is large; i.e., when the density difference is large. We illustrate this relation in Figure 32.2. The Margules relation is a discrete (two-layer) version of the thermal wind relation discussed in Section 26.5.5.

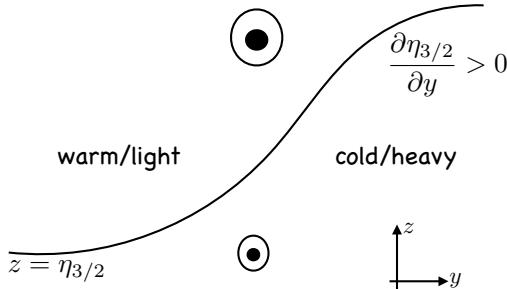


FIGURE 32.2: Illustrating Margules' relation for the northern hemisphere ($f > 0$). Here we show the interface between a two-layer shallow water model with a cold/heavy layer to the right and a warm/light layer to the left. The slope of the interface is positive, $\partial \eta_{3/2} / \partial y > 0$, thus leading to an increase in the eastward zonal geostrophic velocity moving upward, as depicted by the circles with a dot. This orientation corresponds to the northern hemisphere atmospheric jet stream, whereby the interface between cold/heavy air to the north and warm/light air to the south leads to a zonal thermal wind jet.

32.2.3 Geostrophic transport within layers

We are often interested in computing the net volume transport within a layer of fluid in order to measure how much the fluid is moving across a particular region. For an N -layer shallow water fluid this transport is written

$$\mathbf{U} = \int \mathbf{u} dz = \sum_{k=1}^N \mathbf{u}_k h_k, \quad (32.11)$$

where \mathbf{u}_k is the layer horizontal velocity and h_k the layer thickness. For many purposes it is sufficient to compute the transport due to the geostrophic motion, in which case

$$\mathbf{u}_k = (\rho_1 f)^{-1} \hat{\mathbf{z}} \wedge \nabla p_k, \quad (32.12)$$

where ρ_1 is the layer-1 density, which is also used as the reference density for the Boussinesq approximation.

Inserting the geostrophic velocity (32.12) into the depth integrated transport equation (32.11) yields

$$\mathbf{U}_g = \int_{\eta_b}^{\eta} \mathbf{u}_g dz = (\rho_1 f)^{-1} \hat{\mathbf{z}} \wedge \sum_{k=1}^N h_k \nabla p_k \quad (32.13)$$

As an exercise, let us work through the geostrophic transport for three layers (assuming $p_a = 0$), in which the layer pressures are written

$$p_1 = \rho_1 g (\eta_{1/2} - z) \quad (32.14a)$$

$$p_2 = \rho_1 g (\eta_{1/2} - \eta_{3/2}) + \rho_2 g (\eta_{1/2} - z) \quad (32.14b)$$

$$p_3 = \rho_1 g (\eta_{1/2} - \eta_{3/2}) + \rho_2 g (\eta_{3/2} - \eta_{5/2}) + \rho_3 g (\eta_{5/2} - z), \quad (32.14c)$$

which can be written in terms of the reduced gravities according to

$$p_1 = \rho_1 g (\eta_{1/2} - z) \quad (32.15a)$$

$$p_2 = \rho_1 (g_{1/2}^r \eta_{1/2} + g_{3/2}^r \eta_{3/2}) - \rho_2 g z \quad (32.15b)$$

$$p_3 = \rho_1 (g \eta_{1/2} + g_{3/2}^r \eta_{3/2} + g_{5/2}^r \eta_{5/2}) - \rho_3 g z. \quad (32.15c)$$

With $g_{1/2}^r = g$ (consistent with $p_a = 0$) we can express the geostrophic transport within the three layers as

$$h_1 \mathbf{u}_1 = \frac{h_1}{f} \hat{\mathbf{z}} \wedge \nabla (g_{1/2}^r \eta_{1/2}) \quad (32.16a)$$

$$h_2 \mathbf{u}_2 = \frac{h_2}{f} \hat{\mathbf{z}} \wedge (g_{1/2}^r \nabla \eta_{1/2} + g_{3/2}^r \nabla \eta_{3/2}) \quad (32.16b)$$

$$h_3 \mathbf{u}_3 = \frac{h_3}{f} \hat{\mathbf{z}} \wedge (g_{1/2}^r \nabla \eta_{1/2} + g_{3/2}^r \nabla \eta_{3/2} + g_{5/2}^r \nabla \eta_{5/2}), \quad (32.16c)$$

in which case the depth integrated geostrophic transport is

$$\mathbf{U}_g = \frac{1}{f} \left[g_{1/2}^r (h_1 + h_2 + h_3) \nabla \eta_{1/2} + g_{3/2}^r (h_2 + h_3) \nabla \eta_{3/2} + g_{5/2}^r h_3 \nabla \eta_{5/2} \right], \quad (32.17)$$

where $h_1 + h_2 + h_3 = \eta_{1/2} - \eta_b$ is the total depth of a fluid column spanning the three layers. This expression for \mathbf{U}_g displays the cascade of contributions from each of the layer interfaces. The top interface contributes as a depth-integral along with the full contribution from gravity, whereas the interior layers have successively less thickness weighting along with a weighting by their respective reduced gravities.

32.2.4 Flow within a geostrophic eddy

The ocean and atmosphere are highly turbulent fluids, with turbulent features extending from the small scales (millimeters) to large scales (hundreds to thousands of kilometers). The larger scale turbulent features feel the earth's rotation and thus maintain a force balance close to geostrophic.¹

¹As noted in Section 27.6, the gradient wind balance provides a more accurate approximation to flows in ocean and atmospheric eddies by also including the centrifugal acceleration associated with the curved motion. Even so, the geostrophic balance provides a sufficient approximation for many purposes and it will be used here, along with thermal wind.

We here outline some features of an ocean *geostrophic eddy* as idealized using the reduced gravity model of Section 31.3. Figure 32.3 shows a vertical-zonal slice through the upper portion of an ocean eddy in the middle latitude northern hemisphere ($f > 0$) ocean. The central region consists of a geostrophic eddy, sometimes also referred to as an *ocean mesoscale eddy*. The signature of the eddy is a depression in the free surface height and upward deformation of the pycnocline. The lateral scale of the eddy is on the order of the internal deformation scale (see Exercise 32.11).

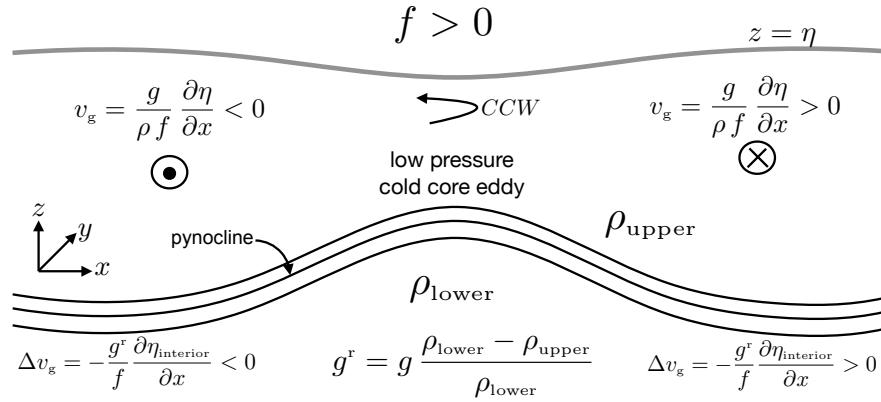


FIGURE 32.3: Vertical-zonal slice through a northern hemisphere mid-latitude cold-core ocean eddy looking from the south to the north (north is into the page). The ocean free surface is depressed down in the middle so that flow is cyclonic (counter-clockwise in north) within the upper portion of the eddy where $\rho = \rho_{\text{upper}}$. The pycnocline (region of enhanced vertical density gradient) is deformed upward, and the baroclinic flow implied by Margules' relation (32.10) is indicated, making use of the reduced gravity $g^r = g(\rho_{\text{lower}} - \rho_{\text{upper}})/\rho_{\text{upper}}$.

The ocean eddy in Figure 32.3 is an anomalously dense cyclonic mesoscale eddy with the dense water causing the pycnocline to deviate upward. If density is dominated by temperature, then the eddy is a *cold core eddy*. Under geostrophic balance, water circulates counter-clockwise in the northern hemisphere within the upper portion of the cold core eddy, in the region where the pressure gradient force is dominated by the free surface undulation. In this case we see say the eddy is cyclonic. According to the reduced gravity model from Section 31.3 (see in particular Figure 31.5), the ratio of the free surface undulation to the pycnocline undulation scales like the reduced gravity, so that a meter undulation of the free surface corresponds to roughly 100 m undulation of the pycnocline. The same ideas hold for a *warm core eddy* (Figure 32.4), with undulations complementing those in the cold core and thus supporting anti-cyclonic thermal wind flow.

In presenting the idealized rendition of an ocean eddy in Figure 32.3, we are assuming a reduced gravity model is sufficient and that the atmosphere has no significant horizontal pressure gradients over the scale of the eddy. Under these assumptions, we make use of the Margules' relation (32.10) to deduce the thermal wind flow in the upper layer relative to the layer below. For the left side of the eddy, where $\partial\eta_{\text{interior}}/\partial x > 0$, the implied meridional geostrophic velocity is southward, consistent with the orientation of the flow implied by the sea surface gradient. Conversely, on the right side of the eddy, where $\partial\eta_{\text{interior}}/\partial x < 0$, the flow is northward.

32.2.5 Comments

An alternative definition of reduced gravity for the 2-layer system uses the average density,

$$\bar{\rho}^{12} = (\rho_1 + \rho_2)/2, \quad (32.18)$$

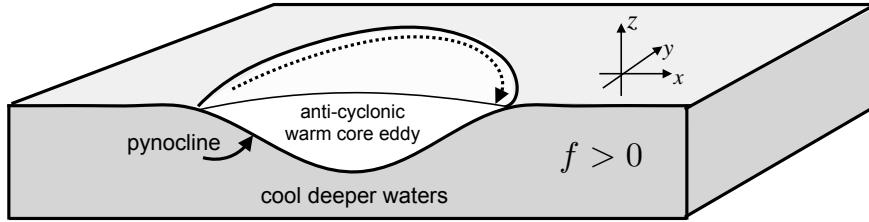


FIGURE 32.4: Schematic of a warm core (light water) geostrophic ocean eddy as idealized by a reduced gravity model. The geostrophic/thermal wind flow is anti-cyclonic within the eddy, which contrasts to the cyclonic flow for a cold core eddy as depicted in Figure 32.3. The eddy is characterized by a slight expansion of the free surface (high pressure) and a relatively larger depression of the pycnocline.

in which case the modified reduced gravity is

$$\bar{g}_{3/2} = g \left[\frac{\rho_2 - \rho_1}{\bar{\rho}^{12}} \right]. \quad (32.19)$$

Correspondingly, the pressure gradient on the right hand side of the momentum equations (32.5a) and (32.5b) have $1/\rho_1$ replaced by $1/\bar{\rho}^{12}$. Even so, the equation for the vertical shear evolution, (32.7), remains unchanged. For more than two layers, it is typical to take the reference density equal to that of the top layer, as done in our three layer example in Section 32.2.3.

32.3 Thickness weighted momentum equation

In this section we derive the momentum equation for the depth integrated velocity in a shallow water layer, thus allowing for a measure of the momentum budget in a column of shallow water fluid within a layer. This formulation proves particularly useful when studying forces acting on the layer, such as those from pressure form stresses, and has proven of use for studies of rotating hydraulics such as in the book by [Pratt and Whitehead \(2008\)](#).

For a single layer, the shallow water column extends from the surface to the bottom of the layer (see Figure 32.5)

$$\int_{\eta_b}^{\eta} \mathbf{u} dz = \mathbf{u} h, \quad (32.20)$$

so that the column momentum equals to $\mathbf{u} h \rho dx dy$. The resulting momentum equation is written in an Eulerian flux form. Doing so facilitates analysis of forces acting on the layer and hence of the layer momentum budget. In Section 32.6 we illustrate the momentum budget for a zonal channel. We also show in Section 32.4.2 that the N -layer equations are isomorphic to the single layer, thus allowing for concepts developed for a single layer to be readily extended to multiple layers.

32.3.1 Single layer equations

Recall the velocity and thickness equations written using the material time operator

$$\frac{D\mathbf{u}}{Dt} + f \hat{z} \wedge \mathbf{u} = -g \nabla \eta \quad \text{and} \quad \frac{Dh}{Dt} = -h \nabla \cdot \mathbf{u}. \quad (32.21)$$

Combining these two equations allows us to write the thickness weighted material acceleration as

$$h \frac{D\mathbf{u}}{Dt} = h \frac{D\mathbf{u}}{Dt} + \mathbf{u} \left[\frac{Dh}{Dt} + h \nabla \cdot \mathbf{u} \right] = \frac{\partial(h \mathbf{u})}{\partial t} + \nabla \cdot [h \mathbf{u} \otimes \mathbf{u}], \quad (32.22)$$

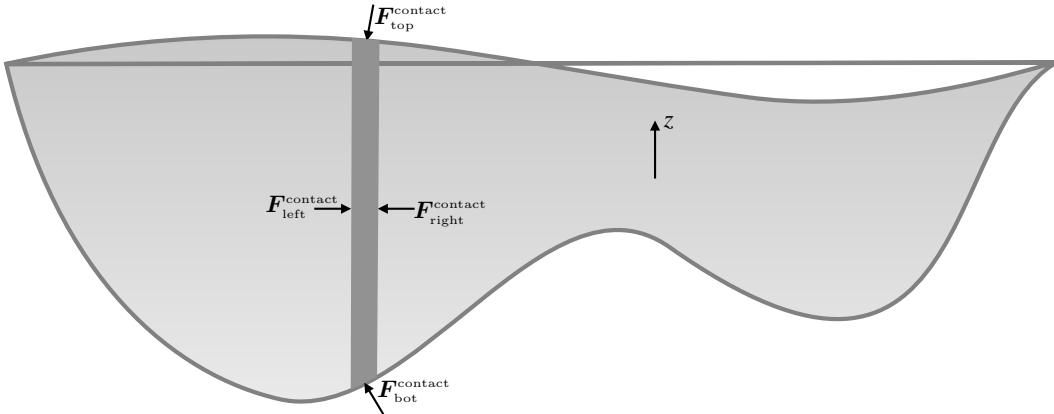


FIGURE 32.5: Momentum of a column of a single layer shallow water fluid is affected by contact forces at the column boundaries (pressure and friction), as well as body forces acting throughout the column (Coriolis and gravity).

so that the thickness weighted equation takes the vector form

$$\frac{\partial(h \mathbf{u})}{\partial t} + \nabla \cdot [h \mathbf{u} \otimes \mathbf{u}] + f \hat{\mathbf{z}} \wedge (h \mathbf{u}) = -g h \nabla \eta. \quad (32.23)$$

The Cartesian tensor form of the outer product is²

$$[\mathbf{u} \otimes \mathbf{u}]_{mn} = u_m u_n \quad \text{for } m, n = 1, 2, \quad (32.24)$$

with $h \mathbf{u} \otimes \mathbf{u}$ the specialization to the shallow water system of the *kinetic stress* appearing in the continuously stratified momentum equation (19.56). The component form of the thickness weighted momentum equation (32.23) is

$$\frac{\partial(h u)}{\partial t} + \partial_x(h u^2) + \partial_y(h u v) - v h f = -g h \partial_x \eta \quad (32.25a)$$

$$\frac{\partial(h v)}{\partial t} + \partial_x(h u v) + \partial_y(h v^2) + u h f = -g h \partial_y \eta. \quad (32.25b)$$

32.3.2 Geostrophic and ageostrophic contributions

Bringing the Coriolis terms to the right hand side of equations (32.25a) and (32.25b) renders

$$\frac{\partial(h u)}{\partial t} + \partial_x(h u^2) + \partial_y(h u v) = h (-g \partial_x \eta + v f) \quad (32.26a)$$

$$\frac{\partial(h v)}{\partial t} + \partial_x(h u v) + \partial_y(h v^2) + u h f = h (-g \partial_y \eta - u f). \quad (32.26b)$$

In the absence of rotation, the right hand side has contributions only from the thickness weighted pressure gradient. For the case of rotation it sometimes proves useful to decompose velocity into its geostrophic and ageostrophic components

$$f u = f (u_a + u_g) = f u_a - g \partial \eta / \partial y \quad (32.27a)$$

$$f v = f (v_a + v_g) = f v_a + g \partial \eta / \partial x, \quad (32.27b)$$

²Equation (32.24) is the only place in this chapter where a subscript refers to a tensor label. Otherwise, subscripts refer to a shallow water layer index as in Section 32.4.2.

in which case equations (32.26a) and (32.26b) become

$$\frac{\partial(h u)}{\partial t} + \partial_x(h u^2) + \partial_y(h u v) = h f v_a \quad (32.28a)$$

$$\frac{\partial(h v)}{\partial t} + \partial_x(h u v) + \partial_y(h v^2) = -h f u_a. \quad (32.28b)$$

One should be careful *not* to take the $f = 0$ limit of these equations since one might spuriously conclude there is no free surface contribution. Instead, equations (32.26a) and (32.26b) should be the basis for the $f = 0$ limit.

32.3.3 Form stresses acting on a shallow water column

The kinetic stress contributes to momentum evolution in equations (32.25a) and (32.25b) via its divergence. In contrast, the pressure stress contributes as a thickness weighted pressure gradient body stress. In this subsection, and in all of Section 32.4, we formulate pressure as a contact stress, in which case it also contributes to momentum evolution as a divergence. In so doing, we provide an Eulerian flux-form conservation law for momentum that supports analysis and interpretation.

Reintroducing atmospheric pressure to symmetrize the forces acting on the layer

To expose both the surface and bottom form stresses, we reintroduce the atmospheric pressure, p_a , and thus make use of the effective sea level (31.4)

$$\eta^{\text{eff}} = \eta + p_a/(\rho g) = \eta_b + h + p_a/(\rho g), \quad (32.29)$$

with the corresponding thickness weighted horizontal momentum equation

$$\frac{\partial(h \mathbf{u})}{\partial t} + \nabla \cdot [h \mathbf{u} \otimes \mathbf{u}] + f \hat{\mathbf{z}} \wedge (h \mathbf{u}) = -g h \nabla \eta^{\text{eff}}. \quad (32.30)$$

Exposing the contact pressure stresses

The free surface height equals $\eta = \eta_b + h$, in which case the momentum equation (32.30) is

$$\frac{\partial(h \mathbf{u})}{\partial t} + \nabla \cdot [h \mathbf{u} \otimes \mathbf{u}] + f \hat{\mathbf{z}} \wedge (h \mathbf{u}) = -(g/2) \nabla h^2 - g h \nabla [\eta_b + p_a/(\rho g)]. \quad (32.31)$$

To help interpret this equation it is convenient to write the boundary terms on the right hand side as

$$-g h \nabla [\eta_b + p_a/(\rho g)] = -\nabla(h p_a/\rho) + (p_a/\rho) \nabla(\eta - \eta_b) - g h \nabla \eta_b \quad (32.32a)$$

$$= -\nabla(h p_a/\rho) + (p_a/\rho) \nabla \eta - (g h + p_a/\rho) \nabla \eta_b \quad (32.32b)$$

$$= -\nabla(h p_a/\rho) + \rho^{-1} (p_a \nabla \eta - p_b \nabla \eta_b), \quad (32.32c)$$

so that

$$-(g/2) \nabla h^2 - g h \nabla [\eta_b + p_a/(\rho g)] = -\nabla[(g/2) h^2 + h p_a/\rho] + \rho^{-1} (p_a \nabla \eta - p_b \nabla \eta_b). \quad (32.33)$$

The first term on the right hand side is the gradient of the layer integrated hydrostatic pressure

$$P \equiv \int_{\eta_b}^{\eta} [p_a + \rho g (\eta - z)] dz = h (\rho g h/2 + p_a), \quad (32.34)$$

and the second term exposes the form stresses acting at the surface and bottom of the layer. With these expressions, the horizontal thickness weighted momentum equation (32.31) becomes

$$\frac{\partial(h \mathbf{u})}{\partial t} + \nabla \cdot [h \mathbf{u} \otimes \mathbf{u} + \mathbb{I} P/\rho] + f \hat{\mathbf{z}} \wedge (h \mathbf{u}) = (p_a \nabla \eta - p_b \nabla \eta_b)/\rho, \quad (32.35)$$

where \mathbb{I} is the unit tensor. Exposing the zonal and meridional components renders

$$\frac{\partial(h u)}{\partial t} + \partial_x(h u^2 + P/\rho) + \partial_y(h u v) - v h f = (p_a \partial_x \eta - p_b \partial_x \eta_b)/\rho \quad (32.36a)$$

$$\frac{\partial(h v)}{\partial t} + \partial_x(h u v) + \partial_y(h v^2 + P/\rho) + u h f = (p_a \partial_y \eta - p_b \partial_y \eta_b)/\rho. \quad (32.36b)$$

The horizontal pressure gradient appears as a continuous operator since we assumed an infinitesimal horizontal cross-sectional area for the fluid column. In contrast, the form stresses appear as a vertical finite difference across the layer interfaces since we are integrating over the thickness of a finite layer. Furthermore, note how the pressure contributions appear in a flux-form, which contrasts to the body force version that appears as thickness weighted pressure gradient.

Kinetic stresses and contact pressure stresses combined into a momentum flux

To anticipate the thickness weighted momentum equation for the stacked shallow water model in Section 32.4, write the finite difference of the form stresses as

$$p_a \nabla \eta - p_b \nabla \eta_b = p_{1/2} \nabla \eta_{1/2} - p_{3/2} \nabla \eta_{3/2} \equiv \delta_k(p_{k-1/2} \nabla \eta_{k-1/2}). \quad (32.37)$$

We here introduced the layer interface difference operator

$$\delta_k(\Phi_{k-1/2}) = \Phi_{k-1/2} - \Phi_{k+1/2} = -(\Phi_{k+1/2} - \Phi_{k-1/2}), \quad (32.38)$$

with the backward difference motivated since k increases downward whereas $\hat{\mathbf{z}}$ points upward. In the following, we choose to define the difference operator to only act on interface fields. Hence, any layer quantity, such as the layer thickness, commutes with the interface operator

$$\delta_k(h \Phi_{k-1/2}) = h (\Phi_{k-1/2} - \Phi_{k+1/2}). \quad (32.39)$$

With the above notation, the component momentum equations (32.36a) and (32.36b) take on the matrix-vector form

$$\begin{bmatrix} \partial_t(h u) - h f v \\ \partial_t(h v) + h f u \end{bmatrix} = - \begin{bmatrix} \partial_x & \partial_y & h^{-1} \delta_k \end{bmatrix} \begin{bmatrix} D_1^{(u)} & D_1^{(v)} & 0 \\ D_2^{(u)} & D_2^{(v)} & 0 \\ D_3^{(u)} & D_3^{(v)} & 0 \end{bmatrix}. \quad (32.40)$$

The 3×3 matrix is a second order tensor with the first and second columns consisting of the momentum fluxes

$$\mathbf{D}^{(u)} = (h u^2 + P/\rho) \hat{\mathbf{x}} + h u v \hat{\mathbf{y}} - (p_{k-1/2} \partial_x \eta_{k-1/2}/\rho) h \hat{\mathbf{z}} \quad (32.41a)$$

$$\mathbf{D}^{(v)} = h u v \hat{\mathbf{x}} + (h v^2 + P/\rho) \hat{\mathbf{y}} - (p_{k-1/2} \partial_y \eta_{k-1/2}/\rho) h \hat{\mathbf{z}}, \quad (32.41b)$$

where we suppressed unnecessary layer indices. The horizontal flux components are given by minus the thickness weighted kinetic stress, $h \mathbf{u} \otimes \mathbf{u}$, plus the depth integrated contact pressure acting on the vertical sides of the shallow water column. The vertical flux component contains the pressure

form stresses acting on the top and bottom interfaces, with these interfacial form stresses leading to the vertical transfer of horizontal form stresses across the layer boundaries. These fluxes allow us to write the thickness weighted zonal and meridional momentum equations as

$$\frac{\partial(h u)}{\partial t} - v h f = -(\hat{x} \partial_x + \hat{y} \partial_y + \hat{z} h^{-1} \delta_k) \cdot \mathbf{D}^{(u)} \quad (32.42a)$$

$$\frac{\partial(h v)}{\partial t} + u h f = -(\hat{x} \partial_x + \hat{y} \partial_y + \hat{z} h^{-1} \delta_k) \cdot \mathbf{D}^{(v)}. \quad (32.42b)$$

In this form of the momentum equation, contributions from contact stresses (kinetic stresses and form stresses) appear as the convergence of these stresses with the divergence operator built as combination of the continuous horizontal gradient operator along with a finite difference vertical operator. The third column of the tensor (32.40) is identically zero and so it can be readily dropped. However, we include it to connect with the *Eliassen-Palm flux tensor* as detailed by [Maddison and Marshall \(2013\)](#). We return to equations (32.42a) and (32.42b) in Section 32.4.9 for the stacked shallow water model.

32.3.4 Comments on the two pressure force formulations

The momentum equations (32.42a) and (32.42b) are written as a flux-form Eulerian conservation law, with only the Coriolis force appearing as a body force. This formulation follows that for Cauchy's equation of motion as discussed in Section 18.1.3. We make use of these flux-form momentum equations in Section 32.6 when discussing force balances in a zonally periodic channel, as well as in Chapter 47 when developing the thickness weighted averaged shallow water equations. Before doing so, we focus in Section 32.4 by further unpacking the contact force version of pressure as it appears in the shallow water model.

What has been gained by writing the momentum equation as the thickness weighted forms (32.42a) and (32.42b) versus the non-flux form velocity equation (32.21)? Indeed, the thickness weighted formulation is arguably less elegant and takes more effort to derive. A key reason we consider the thickness weighted equations is that they provide a venue to study how pressure contact forces alter momentum of a shallow water fluid column through interactions with the bottom, the surface, and adjacent vertical columns. We put this perspective to use in Section 32.6 when studying the force balances on a zonally reentrant channel. Additionally, in Section 32.4 we further pursue the contact force perspective by studying how pressure form stresses appear within a stacked shallow water model. In general, we find it is useful to avail oneself of a variety of perspectives when studying mechanical systems, with each perspective often revealing unique and novel insights.

32.4 Contact pressure forces in shallow water layers

For a finite region of fluid, if the boundary area integrated contact pressure stress is nonzero, then pressure accelerates the region. In this section we study the physics and maths of contact pressure forces as they appear in the stacked shallow water model. As revealed by this study, the columnar motion of the shallow water fluid is fundamental to the analysis. Namely, the contact pressure approach is realized by studying the thickness weighted velocity equations of motion, which determine evolution of the momentum per horizontal area of a shallow water fluid column. We introduced the thickness weighted approach in Section 32.3.3 for a single shallow water layer, and it led to the flux-form Eulerian momentum equations (32.42a) and (32.42b). The single layer

results are reproduced here for the stacked shallow water model, yet only after furthering our understanding of how pressure forces act to move momentum through shallow water layers.

32.4.1 Pressure contact force and pressure body force

As discussed in Section 19.1.3, the connection between the body force and contact force expressions of the pressure force arise through an application of Gauss's divergence theorem to scalar fields (see Section 2.7.2)

$$\mathbf{F}_{\mathcal{R}}^{\text{press}} = - \int_{\mathcal{R}} \nabla p \, dV = - \oint_{\partial\mathcal{R}} p \hat{\mathbf{n}} \, d\mathcal{S}. \quad (32.43)$$

The first expression on the right hand side is a volume integral of the pressure gradient over the fluid region, \mathcal{R} . This expression provides the body force version of the pressure force. The second expression is a surface area integral over the region boundary, $\partial\mathcal{R}$, whose outward normal is $\hat{\mathbf{n}}$. This second expression provides the contact force version of the pressure force. Neither expression is more or less fundamental. Instead, they offer complementary insights into how pressure acts to modify the momentum of a fluid, with general notions of this complementarity the topic of Chapter 20. We here pursue the contact force perspective as a means to understand the *pressure form stress* or *interfacial form stress* acting between layers of a shallow water fluid. There is also a pressure form stress acting between a fluid layer and the solid earth (topographic form stress), as well as between a fluid layer and the overlying atmosphere when the atmosphere has a non-zero mass (atmosphere form stress).

32.4.2 N -layer equations

Our starting point for the development in this section is the N -layer shallow water thickness and velocity equations derived in Section 31.4.2

$$\frac{\partial h_k}{\partial t} + \nabla \cdot (h_k \mathbf{u}_k) = 0 \quad \text{and} \quad [\partial_t + (\mathbf{u}_k \cdot \nabla)] \mathbf{u}_k + f \hat{\mathbf{z}} \wedge \mathbf{u}_k = -(1/\rho_1) \nabla p_k, \quad (32.44)$$

where $k = 1, N$ is the layer index with no implied summation over this index, ρ_1 is the top layer density that also serves as the Boussinesq reference density, and equation (31.73) gives the horizontal pressure gradient acceleration. Equations (32.44) are isomorphic to the single layer equations considered in Section 32.3.1. Hence, the thickness weighted velocity equation is a simple generalization of the single layer equation (32.23)

$$\frac{\partial(h_k \mathbf{u}_k)}{\partial t} + \nabla \cdot [h_k \mathbf{u}_k \otimes \mathbf{u}_k] + f \hat{\mathbf{z}} \wedge (h_k \mathbf{u}_k) = -(h_k/\rho_1) \nabla p_k, \quad (32.45)$$

where, again, there is no implied summation over the layer index, k . We commonly refer to the thickness weighted equation (32.45) as the momentum equation since $\rho dx dy h_k \mathbf{u}_k$ is the horizontal momentum of a shallow water fluid column,

32.4.3 Contact pressure force along vertical sides

We now build up our understanding of pressure form stresses acting in a stacked shallow water fluid, with the essence of this discussion following that encountered for the single layer in Section 32.3.3. Our interest concerns the pressure acting on the boundaries of a fluid column within a shallow water layer, such as shown in Figure 32.6.

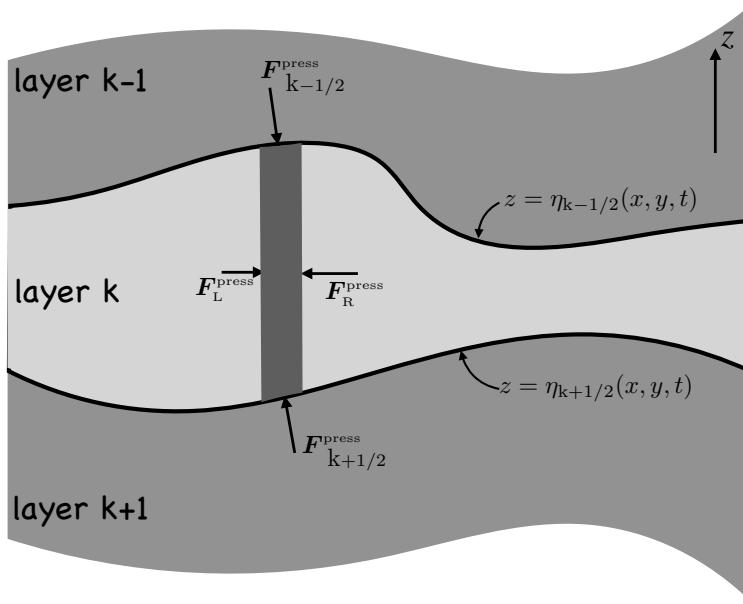


FIGURE 32.6: A schematic of the contact pressure force per area acting on the boundaries of a vertical column region within a shallow water layer of density ρ_k . Since fluid moves as vertical columns in a shallow water layer, we focus on the pressure forces acting on this column. The horizontal cross-sectional area of the column is depth independent. The interface at the lower boundary is at the vertical position $z = \eta_{k+1/2}$, and the upper interface is at $z = \eta_{k-1/2}$. In accordance with Newton's third law, pressures are continuous across each of the $\eta_{k\pm 1/2}$ layer interfaces so that the pressure forces are equal in magnitude yet oppositely directed on the opposite sides to the interfaces. The layer thickness is the difference between the interface positions, $h_k = \eta_{k-1/2} - \eta_{k+1/2}$. The boundaries of the dark gray columnar region feel a contact pressure force acting inward. The left side of the column experiences a pressure p_L ; the right side experiences p_R ; the upper interface has a pressure $p_{k-1/2}$ acting between the layer $k-1$ and layer k , and the lower interface has a pressure $p_{k+1/2}$ acting between the layer $k+1$ and layer k . The net pressure acting on the column is computed as the area integral of the pressure acting around the full extent of the column boundaries. The horizontal components of the stress are known as *interfacial form stresses*. This figure is identical to Figure 20.5 used to discuss the general notions of pressure form stress.

The pressure at a vertical position within the shallow water layer- k is given by

$$p_k(z) = \rho_k g (\eta_{k-1/2} - z) + p_{k-1/2}. \quad (32.46)$$

Integrating this pressure over the layer thickness yields

$$P_k \equiv \int_{\eta_{k+1/2}}^{\eta_{k-1/2}} p_k(z) dz \quad (32.47a)$$

$$= g \rho_k \left[\eta_{k-1/2} h_k - (1/2) (\eta_{k-1/2}^2 - \eta_{k+1/2}^2) \right] + p_{k-1/2} h_k \quad (32.47b)$$

$$= h_k (g \rho_k h_k / 2 + p_{k-1/2}). \quad (32.47c)$$

Since pressure is a linear function of z within a layer, the vertically averaged hydrostatic pressure within a layer, P_k/h_k , equals to the pressure at the upper interface, $p_{k-1/2}$, plus one-half the weight per area of the layer, $g \rho_k h_k / 2$.

The zonal pressure force acting on the column sides is the difference between the pressure integrated across the left and right zonal faces of the column. Assuming the fluid column to have

an infinitesimal horizontal cross-sectional area $dx dy$, we find the zonal pressure force is given by

$$dy \int_{\eta_{k+1/2}}^{\eta_{k-1/2}} (p_L - p_R) dz = -dx dy \left[(g/2) \rho_k \frac{\partial h_k^2}{\partial x} + \frac{\partial(h_k p_{k-1/2})}{\partial x} \right] \quad (32.48a)$$

$$= -dx dy \partial_x \left[(g/2) \rho_k h_k^2 + h_k p_{k-1/2} \right] \quad (32.48b)$$

$$= -dx dy \partial_x P_k. \quad (32.48c)$$

The analogous result holds for the meridional direction, thus rendering the net contact pressure force acting on the vertical sides of the column

$$\mathbf{F}_{\text{sides}}^{\text{press}} = -dx dy \nabla P_k. \quad (32.49)$$

Hence, the contact force on the vertical sides of the column is given by the gradient of the layer vertically integrated pressure, with the vertical integral given by equation (32.47c). It is notable that this semi-discrete exercise reveals no more information than already contained within the integral theorem (32.43). Nonetheless, it is useful to see how the integral theorem manifests within discrete shallow water layers.

32.4.4 Contact pressure force along the top and bottom interfaces

Now consider the contact pressure force acting on the top interface. This interface is generally sloped, so that the contact force has a component in both the vertical and horizontal directions. The vertical component to the pressure force maintains hydrostatic balance with the contact pressure at the lower boundary interface. The horizontal component provides a horizontal acceleration, with this acceleration (sign and magnitude) determined by the slope of the interface. Following our discussion in Section 20.3, we term the horizontal pressure acting on the sloped interface the *interfacial form stress*. In addition to acting between two fluid layers with sloped interfaces, form stress also acts between a fluid and the solid-earth bottom, as well as the fluid and the atmosphere (see Section 20.2).

To mathematically characterize the pressure force on the top interface, $z = \eta_{k-1/2}$, requires the outward normal

$$\hat{\mathbf{n}}_{k-1/2} = \frac{\nabla(z - \eta_{k-1/2})}{|\nabla(z - \eta_{k-1/2})|} = \frac{\hat{\mathbf{z}} - \nabla\eta_{k-1/2}}{\sqrt{1 + (\nabla\eta_{k-1/2})^2}}. \quad (32.50)$$

Temporarily assume the interface slope to have a zero projection in the $\hat{\mathbf{y}}$ direction. In this case, the outward normal is

$$\hat{\mathbf{n}}_{k-1/2} = \frac{\hat{\mathbf{z}} - \hat{\mathbf{x}} \partial_x \eta_{k-1/2}}{\sqrt{1 + (\partial_x \eta_{k-1/2})^2}} \quad (32.51a)$$

$$= \frac{\hat{\mathbf{z}} - \hat{\mathbf{x}} \tan \varphi_{k-1/2}}{\sqrt{1 + \tan^2 \varphi_{k-1/2}}} \quad (32.51b)$$

$$= (\hat{\mathbf{z}} - \hat{\mathbf{x}} \tan \varphi_{k-1/2}) \cos \varphi_{k-1/2}, \quad (32.51c)$$

where we defined the interface slope as

$$\frac{\partial \eta_{k-1/2}}{\partial x} = \tan \varphi_{k-1/2}, \quad (32.52)$$

with $\varphi_{k-1/2}$ the angle between the horizontal plane and the interface. Trigonometry leads to an expression for the area of the top of the column³

$$d\mathcal{S}_{k-1/2} = \frac{dx dy}{\cos \varphi_{k-1/2}}, \quad (32.53)$$

so that the product of the area and the outward normal is given by

$$\hat{\mathbf{n}}_{k-1/2} d\mathcal{S}_{k-1/2} = dx dy (\hat{\mathbf{z}} - \hat{\mathbf{x}} \partial_x \eta_{k-1/2}). \quad (32.54)$$

This result generalizes to an interface slope that projects into both horizontal directions

$$\hat{\mathbf{n}}_{k-1/2} d\mathcal{S}_{k-1/2} = dx dy (\hat{\mathbf{z}} - \nabla \eta_{k-1/2}), \quad (32.55)$$

so that the contact pressure force acting on layer-k at its top interface is given by

$$\mathbf{F}_{\text{top}}^{\text{press}} = -dx dy (\hat{\mathbf{z}} - \nabla \eta_{k-1/2}) p_{k-1/2}. \quad (32.56)$$

Analogous considerations lead to the contact pressure force acting on layer-k at the bottom of the column

$$\mathbf{F}_{\text{bot}}^{\text{press}} = dx dy (\hat{\mathbf{z}} - \nabla \eta_{k+1/2}) p_{k+1/2}. \quad (32.57)$$

32.4.5 Form stress

As noted earlier, form stress is the horizontal projection of the contact pressure acting on the sloped top or bottom interface of the fluid column (Chapter 20). The corresponding forces acting on layer-k is the horizontal area element multiplied by the form stress

$$\mathbf{F}_{\text{top face}}^{\text{form stress}} = dx dy (p_{k-1/2} \nabla \eta_{k-1/2}) \quad (32.58)$$

$$\mathbf{F}_{\text{bot face}}^{\text{form stress}} = -dx dy (p_{k+1/2} \nabla \eta_{k+1/2}). \quad (32.59)$$

These forces render a mechanically reversible vertical exchange of horizontal momentum. This momentum exchange occurs without any exchange of matter. Rather, it an inviscid exchange that occurs according to Newton's third law principle of action/reaction.

For a specific case, consider a $k - 1/2$ interface that slopes upward in the $\hat{\mathbf{x}}$ direction (e.g., see Figure 32.6). Form stress acting at the interface provides a $+ \hat{\mathbf{x}}$ directed acceleration on the column. For the $k + 1/2$ interface, a negatively sloped interface also experiences a $+ \hat{\mathbf{x}}$ directed acceleration.

32.4.6 Net contact pressure force on a shallow water column

Summing the contact pressure forces (32.49), (32.56), and (32.57), and dividing by the horizontal area, leads to the net pressure force per horizontal area acting on a column within layer-k

$$\frac{\mathbf{F}_{\text{net},k}^{\text{press}}}{dx dy} = -\nabla P_k - (\hat{\mathbf{z}} - \nabla \eta_{k-1/2}) p_{k-1/2} + (\hat{\mathbf{z}} - \nabla \eta_{k+1/2}) p_{k+1/2} \quad (32.60a)$$

$$= \rho_k g \hat{\mathbf{z}} - \nabla P_k + p_{k-1/2} \nabla \eta_{k-1/2} - p_{k+1/2} \nabla \eta_{k+1/2}. \quad (32.60b)$$

To reach this result we made use of the hydrostatic relation for the vertical pressure difference across a layer

$$p_{k+1/2} - p_{k-1/2} = \rho_k g h_k, \quad (32.61)$$

³Equation (32.53) was also found in Section 15.4.3 when developing the kinematic boundary condition for a material interface.

and introduced the mass of a column within layer k

$$M_k = g \rho_k h_k dx dy. \quad (32.62)$$

The vertical component of the net contact pressure force balances the weight of the column within the layer, which is expected since the shallow water fluid is in hydrostatic balance. The horizontal contact pressure force arises from a horizontal gradient plus the form stress at the surface and bottom interfaces. The gradient term is removed when integrating horizontally over the full domain given that the thickness of the layer vanishes upon reaching the coastlines (e.g., see Figure 20.7). The resulting net force on the full domain arises just from the weight of the fluid acting in the vertical, plus form stress at the surface and bottom. We discuss this point more in Section 32.4.7.

32.4.7 Contact pressure force summed over all layers

Summing the contact pressure force (32.60b) over all layers reveals the contact forces on the interior layer interfaces vanish, as per Newton's third law (see Section 20.1), thus leaving just the form stress at the surface and bottom and the contact pressure force acting on the vertical sides. Dividing by the horizontal area of the column leads to the net pressure force per area

$$\frac{1}{dx dy} \sum_{k=1}^N \mathbf{F}_{\text{net},k}^{\text{press}} = g \hat{\mathbf{z}} \sum_{k=1}^N \rho_k h_k + p_a \nabla \eta_{1/2} - p_b \nabla \eta_b - \sum_{k=1}^N \nabla P_k \quad (32.63a)$$

$$= (p_b - p_a) \hat{\mathbf{z}} + p_a \nabla \eta_{1/2} - p_b \nabla \eta_b - \sum_{k=1}^N \nabla P_k, \quad (32.63b)$$

where we wrote the total mass per area within the column as the difference between the bottom pressure and applied surface pressure

$$\sum_{k=1}^N \rho_k h_k = (p_b - p_a)/g. \quad (32.64)$$

The horizontal components to the applied and bottom pressure terms in equation (32.63b) arise from pressure form stresses applied to the interfaces at the top and bottom of the column. The vertical component arises from the net weight per area of the fluid. The summation term is the horizontal gradient of the depth integrated pressure form stress applied along the vertical sides of the column.

32.4.8 Horizontal pressure force with potential energy gradients

There is another means to express the horizontal pressure force. Here, we expose the gravitational potential energy per horizontal area for a column of fluid within a shallow water layer

$$\mathcal{P}_k = g \rho_k \int_{\eta_{k+1/2}}^{\eta_{k-1/2}} z dz = (g \rho_k/2) (\eta_{k-1/2}^2 - \eta_{k+1/2}^2). \quad (32.65)$$

Use of the layer gravitational potential energy brings the layer vertical integral of the hydrostatic pressure from Section 32.4.3 into

$$\int_{\eta_{k+1/2}}^{\eta_{k-1/2}} p_k(z) dz = g \rho_k h_k^2/2 + h_k p_{k-1/2} = \mathcal{P}_k - g \rho_k h_k \eta_{k+1/2} + h_k p_{k-1/2}. \quad (32.66a)$$

Making use of this result in equation (32.60b), along with a few lines of algebra, yields the net horizontal contact pressure force acting on a shallow water column

$$-\nabla P_k + \delta_k(p_{k-1/2}\nabla\eta_{k-1/2}) = -\nabla\mathcal{P}_k - \delta_k(\eta_{k-1/2}\nabla p_{k-1/2}). \quad (32.67)$$

To reach the identity (32.67) requires the hydrostatic relation, $p_{k+1/2} - p_{k-1/2} = g\rho_k h_k$, and the layer thickness, $h_k = \eta_{k-1/2} - \eta_{k+1/2}$. A consistency check notes that the curl of both sides to equation (32.67) are the same. The identity (32.67) suggests we define the form stress and its dual

$$\mathbf{F}^{\text{form}} = p\nabla\eta \quad \text{and} \quad \mathbf{F}^{\text{dual form}} = -\eta\nabla p. \quad (32.68)$$

These two stresses have the same curl, and thus impart the same pressure torque on a column of fluid (Chapter 35)

$$\nabla \wedge (p\nabla\eta) = \nabla \wedge (-\eta\nabla p). \quad (32.69)$$

However, these stresses are distinct and as such cannot be arbitrarily interchanged.⁴

32.4.9 Momentum equation with contact pressure forces

Comparing the body force version and the contact force version

Recall that the thickness weighted velocity equation (32.45), as written in terms of the pressure gradient body force, is given by⁵

$$\frac{\partial(h_k \mathbf{u}_k)}{\partial t} + \nabla \cdot [h_k \mathbf{u}_k \otimes \mathbf{u}_k] + f \hat{\mathbf{z}} \wedge (h_k \mathbf{u}_k) = -(h_k/\rho_1) \nabla_z p_k, \quad (32.70)$$

again with no implied summation over the layer label, k. Alternatively, we can make use of the net contact pressure force (32.60b) so that

$$\frac{\partial(h_k \mathbf{u}_k)}{\partial t} + \nabla \cdot (h_k \mathbf{u}_k \otimes \mathbf{u}_k + \mathbb{I} P_k/\rho_1) + f \hat{\mathbf{z}} \wedge (h_k \mathbf{u}_k) = \delta_k(p_{k-1/2}\nabla\eta_{k-1/2})/\rho_1, \quad (32.71)$$

where \mathbb{I} is the identity tensor, P_k is the layer integrated pressure given by equation (32.47c), and δ_k is the difference operator defined by equation (32.38). Choosing the contact pressure force as in equation (32.67) to expose the potential energy brings the momentum equation (32.71) into the alternative form

$$\frac{\partial(h_k \mathbf{u}_k)}{\partial t} + \nabla \cdot [h_k \mathbf{u}_k \otimes \mathbf{u}_k + \mathbb{I} \mathcal{P}_k/\rho_1] + f \hat{\mathbf{z}} \wedge (h_k \mathbf{u}_k) = -\delta_k(\eta_{k-1/2}\nabla p_{k-1/2})/\rho_1, \quad (32.72)$$

Equations (32.70), (32.71), and (32.72) allows us to identify the body force and contact force versions of the thickness weighted horizontal pressure acceleration

$$-h_k \nabla_z p_k = -\nabla P_k + \delta_k(p_{k-1/2}\nabla\eta_{k-1/2}) = -\nabla\mathcal{P}_k - \delta_k(\eta_{k-1/2}\nabla p_{k-1/2}). \quad (32.73)$$

⁴It is notable that much of the literature refers to $-\eta\nabla p$ as the form stress rather than the dual form stress. This usage presumably originates from the common application of zonal averages for studying atmospheric motions, whereby $\eta \partial_x p^x = -p \partial_x \eta^x$. But this identity does not hold for arbitrary averaging operators so it is generally necessary to distinguish the form stress from the dual form stress.

⁵In equation (32.70) we wrote the gradient on the pressure as ∇_z since we are only interested in the horizontal gradient acting on $p_k(x, y, z)$. All other objects in equation (32.70) are just a function of horizontal position within a layer, so that ∇ acting on them reduces to ∇_z . Hence, the subscript on the gradient operator, ∇_z , is exposed only when it acts on a function of z , such as for $p_k(x, y, z)$. Since $p_k(x, y, z)$ is a linear function of z , its horizontal gradient is depth independent within the layer, as illustrated in Figure 31.1.

The balance of pressure torques acting on a shallow water column

A necessary (but not sufficient) check of the identity (32.73) can be found by verifying that the curl agrees for each expression

$$-\nabla \wedge \delta_k(\eta_{k-1/2} \nabla p_{k-1/2}) = \nabla \wedge \delta_k(p_{k-1/2} \nabla \eta_{k-1/2}) \quad (32.74a)$$

$$= \delta_k[\nabla \wedge (p_{k-1/2} \nabla \eta_{k-1/2})] \quad (32.74b)$$

$$= \delta_k[\nabla p_{k-1/2} \wedge \nabla \eta_{k-1/2}] \quad (32.74c)$$

$$= \nabla p_{k-1/2} \wedge \nabla \eta_{k-1/2} - \nabla(p_{k-1/2} + g \rho_k h_k) \wedge \nabla \eta_{k+1/2} \quad (32.74d)$$

$$= \nabla p_{k-1/2} \wedge \nabla h_k - g \rho_k \nabla h_k \wedge \nabla \eta_{k+1/2} \quad (32.74e)$$

$$= \nabla(p_{k-1/2} + g \rho_k \eta_{k+1/2}) \wedge \nabla h_k \quad (32.74f)$$

$$= \nabla(p_{k-1/2} - g \rho_k h_k + g \rho_k \eta_{k-1/2}) \wedge \nabla h_k \quad (32.74g)$$

$$= \nabla_z[p_{k-1/2} + g \rho_k (\eta_{k-1/2} - z)] \wedge \nabla h_k \quad (32.74h)$$

$$= \nabla_z p_k \wedge \nabla h_k \quad (32.74i)$$

$$= -\nabla \wedge (h_k \nabla_z p_k), \quad (32.74j)$$

which concurs with the curl of the left hand side of equation (32.73). To reach this result we set $h_k = \eta_{k-1/2} - \eta_{k+1/2}$ and used equation (32.46) for the pressure within a shallow water layer: $p_k(z) = \rho_k g (\eta_{k-1/2} - z) + p_{k-1/2}$.

Anticipating our discussion of vorticity for the shallow water fluid in Section 35.2, we observe that the identity derived above

$$\nabla \wedge \delta_k(p_{k-1/2} \nabla \eta_{k-1/2}) = -\nabla \wedge (h_k \nabla_z p_k) \quad (32.75)$$

says that the difference in interfacial pressure torques acting on the top and bottom of a shallow water layer precisely balances minus the torque arising from the thickness weighted horizontal pressure gradient acting within the layer. This rather remarkable fine tuning of the interfacial and interior pressure torques is a direct consequence of assuming that the fluid motion is restricted to extensible vertical columns within each shallow water layer. This balance is not maintained within a three dimensional fluid, where fluid columns can generally tilt and bend (Chapter 36).

Layer summed momentum equation

Taking the vertical sum of the layer-k momentum equation (32.71) leads to the column integrated horizontal momentum equation

$$\frac{\partial \mathbf{U}}{\partial t} + f \hat{\mathbf{z}} \wedge \mathbf{U} + \nabla \cdot \left[\sum_{k=1}^N (h_k \mathbf{u}_k \otimes \mathbf{u}_k + \mathbb{I} P_k / \rho_1) \right] = [p_a \nabla \eta_{1/2} - p_b \nabla \eta_b] / \rho_1, \quad (32.76)$$

where \mathbf{U} is the depth integrated horizontal velocity given by equation (32.11). The same vertical sum for equation (32.72) leads to

$$\frac{\partial \mathbf{U}}{\partial t} + f \hat{\mathbf{z}} \wedge \mathbf{U} + \nabla \cdot \left[\sum_{k=1}^N (h_k \mathbf{u}_k \otimes \mathbf{u}_k + \mathbb{I} \mathcal{P}_k / \rho_1) \right] = [-\eta_{1/2} \nabla p_a + \eta_b \nabla p_b] / \rho_1. \quad (32.77)$$

The right hand side of equations (32.76) and (32.77) exposes the pressure form stresses (and dual form stresses) acting on the ocean surface and bottom.

Decomposing into depth averaged and depth deviation velocities

For detailed analyses of the depth integrated (layer summed) momentum and vorticity budgets, it is of interest to introduce the depth averaged operator and its deviation

$$\bar{\Phi}^z = \frac{\sum_{k=1}^N h_k \Phi_k}{\sum_{k=1}^N h_k} \quad \text{and} \quad \Phi'_k = \Phi_k - \bar{\Phi}^z \quad (32.78)$$

so that the depth integrated kinetic stress in equation (32.76) is

$$\sum_{k=1}^N h_k \mathbf{u}_k \otimes \mathbf{u}_k = D [\bar{\mathbf{u}}^z \otimes \bar{\mathbf{u}}^z + \bar{\mathbf{u}}' \otimes \bar{\mathbf{u}}'] \quad \text{where } D = \sum_{k=1}^N h_k. \quad (32.79)$$

The velocity \mathbf{u}'_k is the deviation from the layer- k velocity from the depth average velocity, and we refer to it as the *internal velocity*, whereas the depth averaged velocity, $\bar{\mathbf{u}}^z$, is the *external velocity*.⁶ The identity (32.79) reveals that the depth integrated kinetic stress can be decomposed into a stress arising from internal-internal velocity interactions plus external-external velocity interactions. By construction, there are no cross-terms (i.e., no internal-external terms) appearing in this depth integrated stress.

Momentum fluxes

Following the single layer discussion in Section 32.3.3, we write the momentum equation (32.71) in the form

$$\frac{\partial(h u)}{\partial t} - v h f = -(\hat{\mathbf{x}} \partial_x + \hat{\mathbf{y}} \partial_y + \hat{\mathbf{z}} h^{-1} \delta_k) \cdot \mathbf{D}^{(u)} \quad (32.80a)$$

$$\frac{\partial(h v)}{\partial t} + u h f = -(\hat{\mathbf{x}} \partial_x + \hat{\mathbf{y}} \partial_y + \hat{\mathbf{z}} h^{-1} \delta_k) \cdot \mathbf{D}^{(v)}, \quad (32.80b)$$

where we dropped the k label for brevity and introduced the layer momentum fluxes

$$\mathbf{D}^{(u)} = (h u^2 + P/\rho_1) \hat{\mathbf{x}} + h u v \hat{\mathbf{y}} - (p_{k-1/2} \partial_x \eta_{k-1/2}/\rho_1) h \hat{\mathbf{z}} \quad (32.81a)$$

$$\mathbf{D}^{(v)} = h u v \hat{\mathbf{x}} + (h v^2 + P/\rho_1) \hat{\mathbf{y}} - (p_{k-1/2} \partial_y \eta_{k-1/2}/\rho_1) h \hat{\mathbf{z}}. \quad (32.81b)$$

Likewise, we can write the momentum equation (32.72) in the component form

$$\frac{\partial(h u)}{\partial t} - v h f = -(\hat{\mathbf{x}} \partial_x + \hat{\mathbf{y}} \partial_y + \hat{\mathbf{z}} h^{-1} \delta_k) \cdot \mathbf{E}^{(u)} \quad (32.82a)$$

$$\frac{\partial(h v)}{\partial t} + u h f = -(\hat{\mathbf{x}} \partial_x + \hat{\mathbf{y}} \partial_y + \hat{\mathbf{z}} h^{-1} \delta_k) \cdot \mathbf{E}^{(v)}, \quad (32.82b)$$

where the dual layer momentum fluxes are given by

$$\mathbf{E}^{(u)} = (h u^2 + \mathcal{P}/\rho_1) \hat{\mathbf{x}} + h u v \hat{\mathbf{y}} + (\eta_{k-1/2} \partial_x p_{k-1/2}/\rho_1) h \hat{\mathbf{z}} \quad (32.83a)$$

$$\mathbf{E}^{(v)} = h u v \hat{\mathbf{x}} + (h v^2 + \mathcal{P}/\rho_1) \hat{\mathbf{y}} + (\eta_{k-1/2} \partial_y p_{k-1/2}/\rho_1) h \hat{\mathbf{z}}. \quad (32.83b)$$

Besides swapping the depth integrated pressure, P , for the potential energy, \mathcal{P} , the dual momentum fluxes, $\mathbf{E}^{(u)}$ and $\mathbf{E}^{(v)}$, make use of the dual form stress, $\eta_{k-1/2} \nabla p_{k-1/2}$, rather than the form stress, $-p_{k-1/2} \nabla \eta_{k-1/2}$. Upon performing an eddy-mean flow decomposition as in Section 47.6, the eddy correlation portion of the fluxes (32.83a) and (32.83b) lead to the shallow water *Eliassen-Palm* fluxes, which are rows in the Eliassen-Palm flux tensor.

⁶It is also common in the oceanography literature to refer to \mathbf{u}'_k as the *baroclinic velocity* and $\bar{\mathbf{u}}^z$ as the *barotropic velocity*.

32.4.10 Further reading

Ward and Hogg (2011) and *Barthel et al. (2017)* offer pedagogical treatments of the stacked shallow water equations in the context of idealized simulations that lend insight into the dynamical balances. *Maddison and Marshall (2013)* study the Eliassen-Palm flux tensor for continuously stratified quasi-geostrophy as well as the Boussinesq hydrostatic equations.

32.5 Mechanical energy

We here study the mechanical energy for a single shallow water layer sitting on top of a generally non-flat bottom. As part of this discussion we consider the available potential energy (APE).

32.5.1 Gravitational potential energy

The gravitational potential energy per horizontal area of a shallow water fluid of constant density is given by

$$\mathcal{P} = g \rho \int_{\eta_b}^{\eta} z \, dz \quad (32.84a)$$

$$= \frac{g \rho}{2} (\eta^2 - \eta_b^2) \quad (32.84b)$$

$$= g \rho h (\eta - h/2). \quad (32.84c)$$

Note how the gravitational potential energy vanishes when $\eta^2 = \eta_b^2$. For the case $\eta = \eta_b$, there is no fluid so we expect the potential energy to vanish. For the case $\eta = -\eta_b > 0$, there is the same amount of fluid above $z = 0$ as below, in which case the potential energy vanishes since we are computing it relative to the $z = 0$ reference state. Furthermore, in the flat bottom case, $\eta_b = 0$ so that $h = \eta - \eta_b = \eta$, in which case the potential energy (32.84c) reduces to

$$\mathcal{P}_{\text{flat}} = g \rho h^2 / 2. \quad (32.85)$$

Taking the material time tendency of the potential energy in equation (32.84b) yields

$$\frac{D\mathcal{P}}{Dt} = g \rho \left[\eta \frac{D\eta}{Dt} - \eta_b \frac{D\eta_b}{Dt} \right] = g \rho (\eta w_\eta - \eta_b w_b), \quad (32.86)$$

where we used equations for the vertical velocity component from Section 31.2.7. This equation states that the potential energy increases if the thickness of the layer increases through vertical motion along the top and bottom interfaces. Making further use of these equations renders

$$\frac{D\mathcal{P}}{Dt} = g \rho (\eta w_\eta - \eta_b w_b) \quad (32.87a)$$

$$= g \rho [\eta (w_b - h \nabla \cdot \mathbf{u}) - \eta_b w_b] \quad (32.87b)$$

$$= g \rho [w_b (\eta - \eta_b) - \eta h \nabla \cdot \mathbf{u}] \quad (32.87c)$$

$$= g \rho h (w_b - \eta \nabla \cdot \mathbf{u}), \quad (32.87d)$$

where we used the definition $h = \eta - \eta_b$. We make use of this form of the potential energy change for developing the mechanical energy budget in Section 32.86.

32.5.2 Kinetic energy

As for the flat bottom case, the kinetic energy per horizontal area is

$$\mathcal{K} = \frac{1}{2} \int_{\eta_b}^{\eta} \rho \mathbf{u} \cdot \mathbf{u} dz = \frac{1}{2} \rho h \mathbf{u} \cdot \mathbf{u}, \quad (32.88)$$

which has a material time derivative given by

$$\frac{D\mathcal{K}}{Dt} = -g \rho h \mathbf{u} \cdot \nabla \eta - \frac{\rho h \mathbf{u} \cdot \mathbf{u}}{2} \nabla \cdot \mathbf{u} \quad (32.89a)$$

$$= -g \rho h \mathbf{u} \cdot \nabla \eta - \mathcal{K} \nabla \cdot \mathbf{u}. \quad (32.89b)$$

$$= -g \rho h \mathbf{u} \cdot \nabla \eta + \frac{\mathcal{K}}{h} \frac{Dh}{Dt}. \quad (32.89c)$$

The first term on the right hand side is the projection of the horizontal velocity onto the horizontal pressure gradient acceleration, thus indicating that kinetic energy increases if the flow has a component that is directed down the pressure gradient. The second term increases kinetic energy for cases where the thickness increases via the convergence of horizontal velocity.

32.5.3 Mechanical energy

The material time derivative of the mechanical energy per horizontal area is given by

$$\frac{D(\mathcal{K} + \mathcal{P})}{Dt} = -g h \rho \mathbf{u} \cdot \nabla \eta - \mathcal{K} \nabla \cdot \mathbf{u} + g h \rho (w_b - \eta \nabla \cdot \mathbf{u}). \quad (32.90)$$

Note there is no internal energy for the shallow water fluid, so the mechanical energy equals to the total energy.

Expanding the material time derivative in equation (32.90) into its Eulerian components leads to

$$\frac{\partial}{\partial t} (\mathcal{K} + \mathcal{P}) + \mathbf{u} \cdot \nabla (\mathcal{K} + \mathcal{P}) = -g h \rho \mathbf{u} \cdot \nabla \eta - \mathcal{K} \nabla \cdot \mathbf{u} + g h \rho (w_b - \eta \nabla \cdot \mathbf{u}), \quad (32.91)$$

with rearrangement rendering

$$\frac{\partial}{\partial t} (\mathcal{K} + \mathcal{P}) + \nabla \cdot [\mathbf{u} (\mathcal{K} + \mathcal{P})] = \mathcal{P} \nabla \cdot \mathbf{u} - g h \rho \mathbf{u} \cdot \nabla \eta + g h \rho (w_b - \eta \nabla \cdot \mathbf{u}) \quad (32.92a)$$

$$= (\mathcal{P} - g h \rho \eta) \nabla \cdot \mathbf{u} + g h \rho (w_b - \mathbf{u} \cdot \nabla \eta) \quad (32.92b)$$

$$= -(g \rho h^2 / 2) \nabla \cdot \mathbf{u} + g h \rho (w_b - \mathbf{u} \cdot \nabla \eta) \quad (32.92c)$$

$$= -\nabla \cdot (\mathbf{u} g \rho h^2 / 2) + g \rho h \mathbf{u} \cdot \nabla h + g h \rho (w_b - \mathbf{u} \cdot \nabla \eta) \quad (32.92d)$$

$$= -\nabla \cdot (\mathbf{u} g \rho h^2 / 2) + g \rho h (\mathbf{u} \cdot \nabla (h - \eta) + w_b) \quad (32.92e)$$

$$= -\nabla \cdot (\mathbf{u} g \rho h^2 / 2) + g \rho h (-\mathbf{u} \cdot \nabla \eta_b + w_b) \quad (32.92f)$$

$$= -\nabla \cdot (\mathbf{u} g \rho h^2 / 2), \quad (32.92g)$$

where we used the identity

$$w_b = \frac{D\eta_b}{Dt} = \mathbf{u} \cdot \nabla \eta_b, \quad (32.93)$$

which follows since $\partial \eta_b / \partial t = 0$. We are thus left with the conservation law

$$\frac{\partial \mathcal{E}}{\partial t} + \nabla \cdot \mathbf{F} = 0, \quad (32.94)$$

which has the specific expression

$$\frac{\partial}{\partial t}(\mathcal{K} + \mathcal{P}) + \nabla \cdot [\mathbf{u} (\mathcal{K} + \mathcal{P} + g \rho h^2/2)] = 0, \quad (32.95)$$

where the total mechanical energy is

$$\mathcal{E} = \mathcal{K} + \mathcal{P} = \frac{1}{2} \rho h \mathbf{u} \cdot \mathbf{u} + \frac{1}{2} \rho g (\eta^2 - \eta_b^2), \quad (32.96)$$

and the flux of mechanical energy is

$$\mathbf{F} = \mathbf{u} (\mathcal{K} + \mathcal{P} + g \rho h^2/2) = \mathbf{u} (\mathcal{K} + \rho g \eta h). \quad (32.97)$$

Recall in our discussion of total energy for a continuously stratified fluid in Section 22.6.1, where we identified the the mechanical injection work term that appears in the energy flux in addition to the advection of energy. Here, that term is given by $g \rho h^2/2$ for the shallow water fluid.

32.5.4 Available potential energy

As discussed in Section 24.8, a huge portion of the gravitational potential energy is not realizable as kinetic energy, merely because the minimum potential energy state is when the fluid is at rest with some fluid parcels sitting above others. Available potential energy measures that amount of the gravitational potential energy that can be converted to kinetic energy through a reversible rearrangement of the fluid. We here display the available potential energy for a single shallow water layer, thus specializing the more general discussion given in Section 24.8.

The gravitational potential energy for a single shallow water layer is given by

$$P = g \rho \int dA \int_0^\eta z dz = \frac{g \rho}{2} \int \eta^2 dA, \quad (32.98)$$

where $\int dA$ is the horizontal integral over the full domain of the fluid and we choose to measure the potential energy relative to $z = 0$. The background or reference potential energy is realized by relaxing the free surface interface to its uniform area average value, $z = \bar{\eta}$ (see Figure 31.1), so that

$$P_{\text{ref}} = \frac{g \rho}{2} \int \bar{\eta}^2 dA = \frac{g \rho}{2} \bar{\eta}^2 A. \quad (32.99)$$

The available potential energy is the difference

$$E_{\text{APE}} = P - P_{\text{ref}} = \frac{g \rho}{2} \int (\eta^2 - \bar{\eta}^2) dA = \frac{g \rho}{2} \int (\eta')^2 dA \geq 0, \quad (32.100)$$

where

$$\eta' = \eta - \bar{\eta} \quad (32.101)$$

is the anomalous free surface. Equation (32.100) shows that the APE is non-negative for the shallow water layer. That is, any slope to the shallow water layer represents a store of positive APE. We derive the APE for a single shallow water layer with bottom topography in Exercise 32.8 and for two layers in Exercise 32.9. In both cases we find the APE is non-negative.

32.6 Steady momentum balance in a zonal channel

In this section we study the depth integrated steady momentum budget for a shallow water layer in a zonally re-entrant channel as depicted in Figure 32.7. The results hold for a single layer and can be generalized to multiple layers. In theoretical oceanography, such shallow water models have been used to garner insights into the Southern Ocean circulation ($f < 0$), and we keep this application in mind for the following. We are particularly interested in the force balances required to reach a steady state in the presence of a prescribed wind stress acceleration, τ/ρ . The channel has arbitrary topography, including northern and southern bounds with sloping shelves and shorelines. Applying a zonal surface stress inserts zonal momentum to the fluid through the ocean surface. For simplicity we set the atmospheric pressure to zero, $p_a = 0$, so that the bottom pressure is $p_b = \rho g h$, and there is only a form stress acting at the bottom.

A similar analysis was presented in Section 20.4 for the depth integrated axial angular momentum budget in a continuously stratified fluid. Following from that analysis, we ignore the role of internal viscous friction. However, we consider bottom frictional stresses written as a quadratic bottom drag

$$\mathbf{F}^{\text{drag}} = -C_d \mathbf{u} |\mathbf{u}|, \quad (32.102)$$

where $C_d > 0$ is a dimensionless bottom drag coefficient.

The horizontal areal extent of the domain is a function of space and time since the shallow water layer rises up and down the northern and southern shorelines as motion occurs. Even so, since layer thickness vanishes at the shoreline edge, the horizontal boundary conditions for the shallow water layer are easy to apply when working with the thickness weighted equations. That is, all thickness weighted fields vanish at the shoreline edge merely since the thickness vanishes at the edge. Indeed, working with thickness weighted fields allows us to handle any degree of vanishing layer thicknesses, including if the topography in the center of the channel becomes an island rather than a submerged seamount. We also made use of this property of thickness weighted budgets in Section 20.4 where we also considered sloping sides rather than the commonly considered, yet less realistic, vertical sides.

32.6.1 Volume transport for steady flow

Before considering the steady force balance, we establish a constraint based on volume conservation by considering the steady thickness equation

$$\nabla \cdot (h \mathbf{u}) = 0. \quad (32.103)$$

As discussed for incompressible flow in Chapter 17, this non-divergence condition means that there is zero net steady transport crossing any simply connected closed contour in the fluid. A particularly interesting closed contour is one that is periodic and extends across the full zonal extent of the channel (see Figure 32.7), in which case

$$\oint h \mathbf{u} \cdot \hat{\mathbf{n}} \, ds = 0, \quad (32.104)$$

where \oint denotes a periodic line integral across the zonal extent of the channel, $\hat{\mathbf{n}}$ is a unit vector normal to the contour, and ds is the arc-length line element along the contour. The constraint (32.104) reflects the inability of the steady flow to build up or deplete the fluid on one region of the channel at the expense of another. In particular, if the contour follows a constant latitude line,

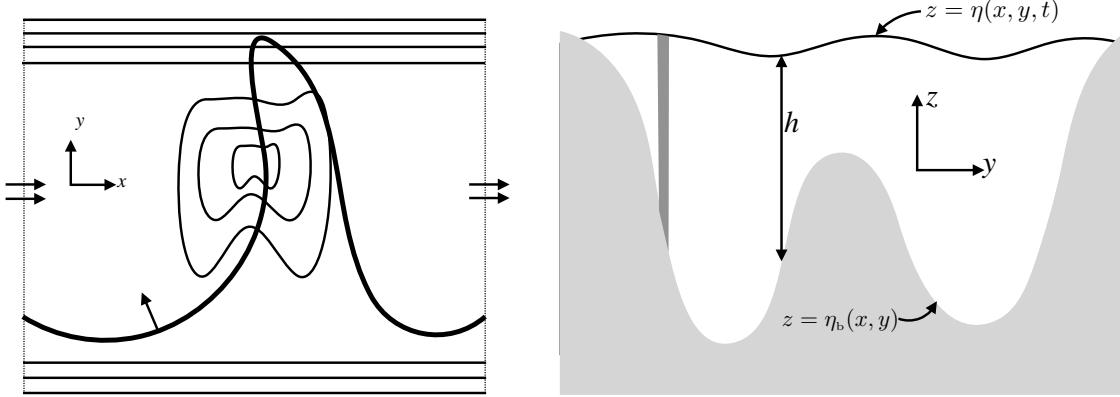


FIGURE 32.7: A zonally periodic/re-entrant southern hemisphere channel with northern and southern shelves and arbitrary seamount topography. Flow that leaves either the east or west boundary is assumed to re-enter the other side, so that the topology is periodic zonally. Left panel: horizontal (plan) view, showing the contours of the shelves and the topography, with flow leaving one of the zonal ends re-entering the other. The arbitrary dark solid contour extends across the zonal extent of the channel and is periodic, with a unit vector, $\hat{\mathbf{n}}$, depicted normal to a point along the contour. In a steady state, the net fluid transport crossing this contour vanishes: $\oint h \mathbf{u} \cdot \hat{\mathbf{n}} ds = 0$, meaning that there is no accumulation of fluid within any region of the channel. Right panel: meridional-depth view through an arbitrary longitude, along with a sample vertical column of water extending from the bottom to the surface. The shoreline edges occur where the layer thickness vanishes on the northern and southern shelves. So although the position of the shoreline edge is a function of space and time (since the fluid moves up and down the shoreline slope), the vanishing layer thickness found at the edge renders a simple treatment of boundary conditions for the thickness weighted equations.

then we see that for a steady state there is no net meridional transport across a latitude circle

$$\oint h \mathbf{u} \cdot \hat{\mathbf{y}} dx = \oint h v dx = 0 \quad \text{steady flow.} \quad (32.105)$$

In particular, this result means that the depth weighted Coriolis acceleration appearing in the zonal momentum equation vanishes when integrated zonally

$$\oint f h v dx = f \oint h v dx = 0 \quad \text{steady flow.} \quad (32.106)$$

We make use of this result in Section 32.6.3.

32.6.2 Steady meridional balances

Consider the meridional thickness weighted momentum equation (32.36b) in the presence of a wind stress and bottom drag

$$\frac{\partial(h v)}{\partial t} + \partial_x(h u v) + \partial_y(h v^2 + g h^2/2) + u h f = -(p_b/\rho) \partial_y \eta_b + \tau^y/\rho - C_d v |\mathbf{u}|. \quad (32.107)$$

Integrating zonally over the channel removes the zonal transport term, $\partial_x(h u v)$, due to periodicity

$$\oint \left[\frac{\partial(h v)}{\partial t} + \partial_y(h v^2 + g h^2/2) + u h f \right] dx = \oint [-(p_b/\rho) \partial_y \eta_b + \tau^y/\rho - C_d v |\mathbf{u}|] dx. \quad (32.108)$$

We can pull the time derivative outside of the zonal integral since the domain is zonally periodic, in which case the time changes to the net meridional transport across a latitude circle are given by

$$\frac{d}{dt} \oint v h dx = \oint [-\partial_y(h v^2 + g h^2/2) - u h f - (p_b/\rho) \partial_y \eta_b + \tau^y/\rho - C_d v |\mathbf{u}|] dx. \quad (32.109)$$

Correspondingly, a steady state along a latitude circle is realized by the balance

$$\oint [\partial_y(h v^2 + g h^2/2) + u h f] dx = \oint [-(p_b/\rho) \partial_y \eta_b + \tau^y/\rho - C_d v |\mathbf{u}|] dx. \quad (32.110)$$

The right hand side represents forcing by the topographic form stress, winds, and bottom drag. That forcing, integrated over a latitude circle, balances the left hand side, which is the Coriolis acceleration arising from zonal motion, plus the meridional divergence of meridional momentum advection plus layer integrated pressure.

32.6.3 Steady zonal balance

Now consider the zonal momentum equation written in its flux-form (see equation (32.36a))

$$\frac{\partial(h u)}{\partial t} + \partial_x(h u^2 + g h^2/2) + \partial_y(h u v) - v h f = -(p_b/\rho) \partial_x \eta_b + \tau^x/\rho - C_d u |\mathbf{u}|. \quad (32.111)$$

Assuming a steady state and integrating along a latitude circle leads to

$$\rho \oint \partial_y(h u v) dx = \oint [-p_b \partial_x \eta_b + \tau^x - C_d \rho u |\mathbf{u}|] dx, \quad (32.112)$$

where we dropped the Coriolis acceleration as per volume conservation in equation (32.106). For flows that are quasi-linear, the nonlinear term $\partial_y(h u v)$ will be subdominant to the wind stress and topographic form stress, thus leading to the approximate balance along each latitude circle

$$\oint p_b \partial_x \eta_b dx \approx \oint [\tau^x - C_d \rho u |\mathbf{u}|] dx \quad \text{nonlinear term small.} \quad (32.113)$$

We realize an exact balance over the full channel domain by integrating the latitude balance (32.112) meridionally, which allows us to drop the nonlinear term $\partial_y(h u v)$ since $h = 0$ at the northern and southern shoreline edges, thus leading to

$$\int \left[\oint \tau^x dx \right] dy = \int \left[\oint [p_b \partial_x \eta_b + C_d \rho u |\mathbf{u}|] dx \right] dy. \quad (32.114)$$

This is an exact steady state balance realized by integrating the zonal thickness weighted momentum equation over the full domain. It is a balance between the momentum input from the winds (left hand side) with the bottom form stress plus bottom frictional drag (right hand side).

32.6.4 The role of frictional bottom drag

Consider a flat bottom channel, in which case the area integrated balance (32.114) is between winds and bottom drag

$$\int \left[\oint \tau^x dx \right] dy = \int \left[\oint C_d \rho u |\mathbf{u}| dx \right] dy \implies \rho^{-1} \langle \tau^x \rangle = C_d \langle u |\mathbf{u}| \rangle, \quad (32.115)$$

where the angle brackets denote a channel area mean. Typical empirical values for the dimensionless bottom drag coefficient are

$$C_d \approx 2 \times 10^{-3}. \quad (32.116)$$

An area mean eastward zonal wind stress of $\langle \tau^x \rangle = 0.1 \text{ N m}^{-2}$ leads to a root-mean-square zonal velocity of

$$\sqrt{\langle u |\mathbf{u}| \rangle} \approx \sqrt{\langle u^2 \rangle} \approx 0.2 \text{ m s}^{-1}, \quad (32.117)$$

where the first approximation follows from assuming the zonal velocity dominates over the meridional velocity. How realistic is this number for the Southern Ocean? Field measurements from the Southern Ocean suggest that depth and area averaged velocities are far smaller than this value. Furthermore, if this value occurred in a channel 4000 m deep and 2000 km wide (a rough idealization of the Antarctic Circumpolar Current), then this depth averaged velocity would yield a zonal volume transport of $\approx 1500 \times 10^6 \text{ m}^3 \text{ s}^{-1}$, which is about ten times larger than the measured transport through the Drake Passage.

Munk and Palmén (1951) identified the problematic aspect of assuming a bottom frictional stress balance for the Southern Ocean. In brief, the field measurements do not support a frictional balance, either from bottom drag or from internal turbulent viscous friction. By inference, they proposed that topographic form stress is the dominant term that balances wind stress in the Southern Ocean. They supported that inference through estimates based on topographic features encountered by the Antarctic Circumpolar Current in its transit of the Southern Ocean. Subsequent studies, using theory, field measurements, and numerical models, support their conclusion. Indeed, in numerical models one finds that so long as there is only a modest degree of bottom slope, the bottom topographic form stress dominates over bottom drag. Given these considerations, we dispense with bottom drag for the remainder of this section.

32.6.5 Correlation between surface height and topographic slope

Given the minor role for bottom drag in establishing a steady channel flow in the Southern Ocean, the balance (32.114) says that an eastward wind stress must be balanced, in the area integral over the channel, by a westward topographic form stress

$$\int \left[\oint \tau^x \, dx \right] dy = \int \left[\oint p_b \partial_x \eta_b \, dx \right] dy. \quad (32.118)$$

What is required to establish a westward topographic form stress? Quite simply, in the area mean, there must be an anomalously large bottom pressure in regions where $\partial_x \eta_b > 0$ and an anomalously small bottom pressure in regions where $\partial_x \eta_b < 0$. Bottom pressure in a shallow water layer is determined by the column thickness. Hence, to establish the anomalous bottom pressures requires an anomalously thick fluid column upstream of topographic bumps and thin fluid column downstream. This situation is illustrated in Figure 32.8 described in Section 32.6.7.

To further reveal the correlation between surface height and bottom topography, write $p_b = \rho g h$ and use $\eta = h + \eta_b$ so that

$$p_{\text{bot}} \partial_x \eta_b = \rho g (\eta - \eta_b) \partial_x \eta_b = \rho g \eta \partial_x \eta_b - (g/2) \partial_x \eta_b^2. \quad (32.119)$$

The balance (32.118) thus becomes

$$\langle \tau^x \rangle = \rho g \langle \eta \partial_x \eta_b \rangle. \quad (32.120)$$

Furthermore, due to zonal periodicity, it is only the zonal anomalies in η and η_b that contribute so that

$$\langle \tau^x \rangle = \rho g \langle \eta' \partial_x \eta'_b \rangle, \quad (32.121)$$

where

$$\eta = \eta' + L^{-1} \oint \eta \, dx \quad \text{and} \quad \eta_b = \eta'_b + L^{-1} \oint \eta_b \, dx, \quad (32.122)$$

with $L = \oint dx$ the zonal length of the channel. With $\langle \tau^x \rangle > 0$, we see that surface height anomalies must be positively correlated with the bottom slope, $\int \eta' \partial_x \eta'_b \, dx \, dy > 0$. That is, the surface height is high where topography slopes are positive and low where topography slopes are negative.

As noted above, we must have a positive correlation between surface height anomalies and topographic slope, as in equation (32.121). It follows that a nonzero zonal integrated topographic form stress requires a nonzero phase shift between surface height anomalies and the bottom topography anomalies. That is, if the surface height and bottom topography were perfectly aligned along a latitude circle, then $\oint \eta' \partial_x \eta'_b dx = 0$, in which case there is a zero zonal integrated topographic form stress along that latitude. The required phase shift between the free surface and bottom topography has the free surface shifted ahead (i.e., to the west) of the topography. We consider an explicit example in Section 32.6.7 to help in our understanding.

32.6.6 Connection to meridional geostrophic transport

Zonal periodicity allows us to write the balance (32.121) as

$$\langle \tau^x \rangle = -\rho g \langle \partial_x \eta' \eta'_b \rangle \quad (32.123)$$

We assume that $g \partial_x \eta'$ is associated with an anomalous meridional geostrophic velocity

$$g \partial_x \eta' = f v'_g \quad (32.124)$$

so that the balance (32.123) is

$$\langle \tau^x \rangle = -\rho \langle f v'_g \eta'_b \rangle. \quad (32.125)$$

Hence, the steady balance is realized with anomalous meridional geostrophic transport correlated with topographic anomalies. Note that periodicity means that the steady meridional geostrophic transport has a zero zonal integral

$$\oint v'_g dx = (g/f) \oint \partial_x \eta' dx = 0, \quad (32.126)$$

which follows from the steady volume balance discussed in Section 32.6.1.

32.6.7 Sinusoidal example

To help further our understanding of the balance (32.121), and the phase shift required to develop nonzero zonal integrated topographic form stress, consider a sinusoidal topography that is a function only of the zonal direction. Also assume that the free surface has a sinusoidal shape (though we do not specify the dynamical mechanism for it to reach this shape). With these assumptions the anomalous surface height and bottom topography can be written

$$\eta' = \eta_o \sin(\kappa x + \varphi) \quad \text{and} \quad \eta'_b = D \sin(\kappa x) \quad (32.127)$$

where $\eta_o > 0$ is a constant amplitude for the free surface undulations, $D > 0$ is the constant amplitude for the bottom undulations, $\kappa = 2\pi n/L$ is the wavenumber for the undulations, $n > 0$ is an integer, L is the size of the zonal channel, and φ is a phase shift between the topography and the free surface. The corresponding meridional geostrophic flow is given by

$$f v'_g = g \partial_x \eta' = g \kappa \eta_o \cos(\kappa x + \varphi). \quad (32.128)$$

Plugging into the balance (32.121) leads to

$$\sin \varphi = \left[\frac{L \langle \tau^x \rangle}{D \rho g \eta_o \pi n} \right]. \quad (32.129)$$

For $\langle \tau^x \rangle > 0$ we see that the free surface undulations are, as expected, shifted to the west of the bottom topography undulations. As an explicit example from an idealized channel configuration, set

$$L = 10^7 \text{ m} \quad D = 10^3 \text{ m} \quad \eta_o = 1 \text{ m} \quad \tau^x = 1 \text{ N m}^{-2}, \quad (32.130)$$

in which case

$$\sin \varphi \approx (\pi n)^{-1}. \quad (32.131)$$

For $n = 1$, which corresponds to just one topographic bump, then we have a phase shift of $\varphi \approx 18^\circ$, and this case is depicted in Figure 32.8 for the southern hemisphere. If there are two bumps, then the phase shift is reduced to $\varphi \approx 9^\circ$ since each bump shares half the burden of balancing the wind stress.

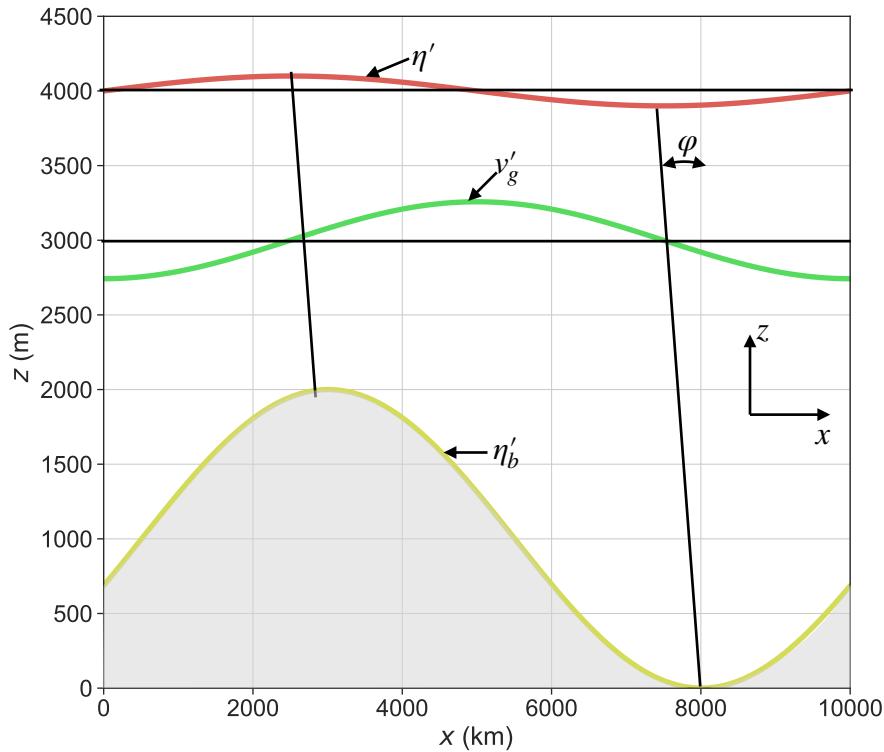


FIGURE 32.8: Zonal-depth view of the wind stress and bottom pressure acting on a steady state layer of shallow water fluid flowing over a sinusoidal bump (north is directed into the page and east is to the right). The eastward surface stress acts in the $+\hat{x}$ direction and it is balanced by a westward topographic form stress. To establish this form stress when integrated over the zonal extent of the channel requires the sea surface undulations to be phase shifted to the west of the bottom topography undulations, with the phase shift, φ , given by equation (32.129). To better visualize the sea surface, the surface undulations are in units of cm whereas the bottom topography undulations are in units of m. In general, for the topographic form stress to balance the wind stress, the bottom pressure must be anomalously large where $\partial \eta_b / \partial x > 0$ and small where $\partial \eta_b / \partial x < 0$, thus leading to the anomalously thick fluid column upstream of the bump and thin column downstream. This correlation also leads to a corresponding meridional geostrophic flow as shown here by the green curve for the southern hemisphere where $f < 0$, with $v_g > 0$ (northward) when shown above the horizontal line and $v_g < 0$ (southward) when below the line. Compare this figure to the analogous schematic in Figure 20.8.

Summarizing the dependencies

We here highlight the various dependencies in the phase equation (32.129).

- The phase shift increases with both larger wind stress and larger zonal extent to the domain. This dependency arises since with enhanced wind stress and an enhanced zonal *fetch* (distance over which the winds blow), there is an increase in zonal momentum inserted to the ocean that must be absorbed by the bottom. The larger phase shift increases this topographic form stress, thus enabling the balance.
- Conversely, the phase shift decreases for larger topography D , and larger undulations in the free surface, η_o , in which case the free surface becomes more aligned with the topographic ridges. This result follows since the topographic form stress is larger for larger topography, thus requiring less phase shift in the surface wave patterns to affect a bottom pressure anomaly.
- Phase shifts decrease when there are more topographic bumps in the channel, with n the parameter setting the number of bumps. For the Southern Ocean, [Munk and Palmén \(1951\)](#) identified around four or five large-scale topographic features that provide the dominant balance for the zonal wind stress.
- What if the parameters are such that the right hand side of equation (32.129) has a magnitude larger than unity (e.g., huge winds, very long fetch, small topography, small surface height amplitude)? This situation signals that topographic form stress is insufficient to balance the winds. In a numerical model, one can merely increase the topographic wavenumber, n , to increase the topographic form stress. Yet where topography is fixed, such as in Nature, then bottom frictional stresses come into play to help reach a force balance (see Section 32.6.4).

Distinguishing steady motion from zero motion

Why is the phase shift (32.129) independent of the fluid depth? One might suspect that to reach a force balance would require more form stress if there is more fluid. Instead, the force balance, as reflected in the phase shift, depends on the zonal anomalies of the surface height and bottom topography. The depth of the fluid is absent.

The answer to this question is that we are seeking a force balance. When forces are balanced there is no acceleration and thus, as per Newton's second law, the fluid maintains a constant velocity relative to the laboratory reference frame. If we instead wished to stop the fluid, then we would need to decelerate all fluid elements to zero velocity. Determining the forces needed to stop the fluid requires the total fluid mass and thus its depth (as well as the time over which the fluid is to stop). If the fluid is in motion, then halting the motion requires a net force, and that is a very different consideration than the case of zero net force. So in brief, a steady state refers to the absence of time dependence in the fluid from the perspective of an Eulerian (laboratory) observer, with a steady state *not* necessarily a static state.

32.6.8 Comments

The analysis of a single shallow water layer has direct relevance to a stratified flow. The reason is that when integrating over the full depth of the fluid, internal interfacial form stresses cancel pairwise. The resulting net balance for contact forces is concerned with just those acting on the boundaries at the surface and the bottom. We discussed this property of interfacial form stresses in Sections 19.1, 20.3, and 32.4. The direct connection between undulations in the sea surface and bottom pressure provides a particularly simple means to study the phase shift between bottom pressure anomalies and topography required to realize a nonzero zonally integrated form

stress. However, this feature of the single layer shallow water system is questionable for a realistic stratified fluid, since stratification breaks the direct connection between sea level undulations and topographic form stress.

The dominance of topographic form stress for the steady Southern Ocean balance contrasts to many of the frictional theories for gyre circulations, such as the classic solutions discussed in Section 35.9. In those theories, based on flat bottom and vertical side configurations, the curl of the wind stress is balanced by torques created by friction.⁷ One is thus led to conclude that zonally re-entrant channels exhibit a fundamentally distinct steady force balance from gyres. However, as discussed in Section 20.4.12, *Hughes and de Cueves* (2001) showed how topographic form stresses associated with sloping sides can lead to an inviscid balance for gyres. That is, friction is not needed so long as the bottom can support topographic torques. In this manner, gyres and channels share much in common so long as they both contain topography and sloping sides.

32.7 Angular momentum in a rotating tank

As our second case study for this chapter, we study angular momentum for a layer of inviscid shallow water fluid in a rotating cylindrical tank. This system was first discussed in Section 23.4, where we developed the horizontal equation of motion

$$\frac{D\mathbf{u}}{Dt} + f \hat{\mathbf{z}} \wedge \mathbf{u} = -\nabla(p/\rho + g_e z - \Omega^2 r^2/2), \quad (32.132)$$

where $r^2 = x^2 + y^2$ is the radial distance from the rotational axis,

$$\Omega = f/2 \quad (32.133)$$

is the constant angular rotation rate, and the vertical component to the right hand side is the hydrostatic balance, $\partial p/\partial z = -\rho g_e$. Where convenient, we make use of the polar coordinates (see Chapter 8.3) in the following, whereby

$$x = r \cos \vartheta \quad (32.134a)$$

$$y = r \sin \vartheta, \quad (32.134b)$$

with the polar angle ϑ measured counter-clockwise from the positive x -axis.

32.7.1 Angular momentum for a column of shallow water fluid

The angular momentum for a column of shallow water fluid, computed with respect to the vertical rotational axis, is given by (see Sections 11.7 and 18.4)

$$L^z = \delta M [\mathbf{x} \wedge (\mathbf{u} + \mathbf{U}_{\text{solid}})] \cdot \hat{\mathbf{z}}, \quad (32.135)$$

where $\mathbf{x} = x \hat{\mathbf{x}} + y \hat{\mathbf{y}} = r \hat{\mathbf{r}}$ is the position vector relative to the rotational axis, $\delta M = \rho h \delta A$ is the constant mass for the fluid column, and the solid-body rotation velocity is

$$\mathbf{U}_{\text{solid}} = (f/2) \hat{\mathbf{z}} \wedge \mathbf{x} = r \Omega \hat{\boldsymbol{\theta}}, \quad (32.136)$$

where $\hat{\mathbf{z}} \wedge \hat{\mathbf{r}} = \hat{\boldsymbol{\theta}}$ is the azimuthal unit vector pointing counter-clockwise around the origin.

⁷We will encounter such balances in Chapter 35 when studying vorticity.

We can further massage the expression for the angular momentum by writing

$$\mathbf{x} \wedge \mathbf{u} = (x v - y u) \hat{\mathbf{z}} = r^2 \dot{\vartheta} \hat{\mathbf{z}}, \quad (32.137)$$

where $\dot{\vartheta} = D\vartheta/Dt$ is the angular velocity. Likewise, we have

$$\mathbf{x} \wedge \mathbf{U}_{\text{solid}} = r^2 \Omega \hat{\mathbf{z}}, \quad (32.138)$$

so that the angular momentum can be written

$$L^z = \delta M [\mathbf{x} \wedge (\mathbf{u} + \mathbf{U}_{\text{solid}})] \cdot \hat{\mathbf{z}} = \delta M r^2 (\dot{\vartheta} + \Omega). \quad (32.139)$$

32.7.2 Material time evolution of the angular momentum

The material time evolution for the angular momentum is given by

$$\frac{DL^z}{Dt} = \delta M [\mathbf{u} \wedge (\mathbf{u} + \mathbf{U}_{\text{solid}})] \cdot \hat{\mathbf{z}} + \delta M \left[\mathbf{x} \wedge \left(\frac{D\mathbf{u}}{Dt} + \frac{D\mathbf{U}_{\text{solid}}}{Dt} \right) \right] \cdot \hat{\mathbf{z}}. \quad (32.140)$$

Note that we set

$$\frac{D(\delta M)}{Dt} = \rho D(h \delta A) Dt = 0 \quad (32.141)$$

since the shallow water fluid columns each have constant volume as they move with the horizontal flow (Section 31.2.4). Using the solid-body rotation velocity given by equation (32.136), and with a constant rotation rate, yields

$$\mathbf{u} \wedge \mathbf{U}_{\text{solid}} + \mathbf{x} \wedge \frac{D\mathbf{U}_{\text{solid}}}{Dt} = \mathbf{u} \wedge (\boldsymbol{\Omega} \wedge \mathbf{x}) + \mathbf{x} \wedge (\boldsymbol{\Omega} \wedge \mathbf{u}) \quad (32.142a)$$

$$= (\mathbf{x} \cdot \mathbf{u}) f \hat{\mathbf{z}}. \quad (32.142b)$$

Making use of the material evolution of the horizontal velocity given by equation (32.132) renders

$$\left[\mathbf{x} \wedge \frac{D\mathbf{u}}{Dt} \right] \cdot \hat{\mathbf{z}} = (\mathbf{x} \wedge [-f \hat{\mathbf{z}} \wedge \mathbf{u} - \nabla(p/\rho + g_e z - \Omega^2 r^2/2)]) \cdot \hat{\mathbf{z}} \quad (32.143a)$$

$$= -f (\mathbf{x} \cdot \mathbf{u}) - (\mathbf{x} \wedge g \nabla \eta) \cdot \hat{\mathbf{z}}. \quad (32.143b)$$

The centrifugal term dropped out since

$$\mathbf{x} \wedge \nabla r^2 = 2 \mathbf{x} \wedge r \hat{\mathbf{r}} = 2 \mathbf{x} \wedge \mathbf{x} = 0. \quad (32.144)$$

The gravitational term dropped out since

$$(\mathbf{x} \wedge \nabla z) \cdot \hat{\mathbf{z}} = (\mathbf{x} \wedge \hat{\mathbf{z}}) \cdot \hat{\mathbf{z}} = 0, \quad (32.145)$$

as does the vertical component to the pressure gradient. We are thus left with

$$\frac{1}{\delta M} \frac{DL^z}{Dt} = -g (\mathbf{x} \wedge \nabla \eta) \cdot \hat{\mathbf{z}}. \quad (32.146)$$

Consequently, the axial angular momentum for a fluid column is modified by the torque from the horizontal pressure gradient caused by undulations in the free surface height. Note how there is no contribution from the Coriolis acceleration, so that the evolution of angular momentum is the same whether viewed in the laboratory frame or rotating frame.

We can bring the expression (32.146) into a more transparent form by switching to polar coordinates

$$\boldsymbol{x} \wedge \nabla \eta = r \hat{\boldsymbol{r}} \wedge \left[\hat{\boldsymbol{r}} \frac{\partial \eta}{\partial r} + \hat{\boldsymbol{\vartheta}} \frac{1}{r} \frac{\partial \eta}{\partial \vartheta} \right] = \frac{\partial \eta}{\partial \vartheta} \hat{\boldsymbol{z}}, \quad (32.147)$$

so that

$$\frac{1}{\delta M} \frac{DL^z}{Dt} = -g \frac{\partial \eta}{\partial \vartheta}. \quad (32.148)$$

This result is directly analogous to the angular momentum evolution for a fluid moving around a sphere as derived in Section 18.4. Namely, in the presence of angular pressure gradients, the fluid experiences a torque that in turn leads to a change in the angular momentum relative to the vertical rotation axis.

32.7.3 Materially invariant angular momentum

The angular momentum for a fluid column is materially invariant (i.e., a constant on a material fluid parcel) if

$$\frac{DL^z}{Dt} = 0 \iff \frac{\partial \eta}{\partial \vartheta} = 0. \quad (32.149)$$

For a flat bottom, equation (23.80) says that the free surface takes on a radial parabolic shape when the fluid is in solid-body rotation. In this case, $\nabla \eta$ is in the radial direction, in which case $\boldsymbol{x} \wedge \nabla \eta = 0$. Consequently, when the fluid is in solid-body rotation, the angular momentum for each fluid column remains materially constant.

32.7.4 Comments

The material evolution equation (32.146) also holds for a fluid on the f -plane tangent to a sphere. The f -plane formulation is slightly simpler than the tank since the centrifugal term is absorbed into the geopotential (see Section 10.12.4). However, the tank is arguably more pedagogical as it is simpler to visualize and to conduct laboratory experiments. See Section 6.6.4 of [Marshall and Plumb \(2008\)](#) for more discussion of rotating tank experiments.

32.8 Inertia-gravity waves in a single shallow water layer

We here consider linear waves in a single layer of shallow water fluid with a flat bottom on an f -plane. Waves in this system arise from the gravitational restoring force.⁸ To develop the mathematical equations for these shallow water gravity waves, we linearize the equations of motion and then develop constraints that must be satisfied for the existence of nontrivial solutions.

32.8.1 Linearized shallow water equations

Recall the shallow water equations of motion are given by the momentum and continuity equations, written here in their Eulerian form

$$\frac{\partial \boldsymbol{u}}{\partial t} + (\boldsymbol{u} \cdot \nabla) \boldsymbol{u} + f \hat{\boldsymbol{z}} \wedge \boldsymbol{u} = -g \nabla \eta \quad (32.150a)$$

$$\frac{\partial h}{\partial t} + h \nabla \cdot \boldsymbol{u} + \boldsymbol{u} \cdot \nabla h = 0. \quad (32.150b)$$

⁸Rossby waves also arise when considering a non-constant Coriolis parameter. We study Rossby waves for the non-divergent barotropic model in Section 34.7.

With a flat bottom we assume $\eta_b = 0$ (see Figure 31.1) so that the surface height equals to the thickness,

$$\eta = h \quad \text{flat bottom with } \eta_b = 0. \quad (32.151)$$

We linearize the momentum and thickness equation relative to a state of rest in which

$$\eta(x, y, t) = H + \eta'(x, y, t) \quad (32.152a)$$

$$\mathbf{u}(x, y, t) = 0 + \mathbf{u}'(x, y, t). \quad (32.152b)$$

Substitution into the thickness equation (32.150b) leads to

$$\frac{\partial \eta'}{\partial t} + (H + \eta') \nabla \cdot \mathbf{u}' + \mathbf{u}' \cdot \nabla \eta' = 0. \quad (32.153)$$

The products $\eta' \nabla \cdot \mathbf{u}'$ and $\mathbf{u}' \cdot \nabla \eta'$ are second order (and nonlinear) in fluctuating quantities. Dropping these terms leads to the linearized surface height (or thickness) equation

$$\frac{\partial \eta'}{\partial t} + H \nabla \cdot \mathbf{u}' = 0 \quad \text{linearized surface height equation.} \quad (32.154)$$

Similarly, the linearized momentum equation takes the form (dropping the nonlinear advection term $(\mathbf{u}' \cdot \nabla) \mathbf{u}'$)

$$\frac{\partial \mathbf{u}'}{\partial t} + f \hat{z} \wedge \mathbf{u}' = -g \nabla \eta' \quad \text{linearized momentum equation.} \quad (32.155)$$

32.8.2 Relative vorticity of linear fluctuations

We here consider how the vertical component of relative vorticity⁹

$$\zeta = \hat{z} \cdot (\nabla \wedge \mathbf{u}) \quad (32.156)$$

evolves for the small amplitude shallow water fluctuations determined by equations (32.154) and (32.155). Taking the curl of the linearized momentum equation (32.155) leads to

$$\frac{\partial \zeta'}{\partial t} = -\nabla \cdot (f \mathbf{u}'). \quad (32.157)$$

On the f -plane and with the non-divergence condition $\nabla \cdot \mathbf{u}' + \partial_z w' = 0$, we have

$$\frac{\partial \zeta'}{\partial t} = f \frac{\partial w'}{\partial z}. \quad (32.158)$$

We can use the linearized thickness equation (32.154) to render an equivalent result

$$\frac{\partial \zeta'}{\partial t} = \frac{f}{H} \frac{\partial \eta'}{\partial t}. \quad (32.159)$$

In either case, relative vorticity of the linearized shallow water system is modified by vertical stretching in the presence of planetary rotation ($f \neq 0$). In the absence of planetary rotation, the relative vorticity remains static at each point in space. Consequently, if the relative vorticity for a non-rotating system starts with a zero value, it will remain so throughout the linearized evolution.

⁹In Chapter 35 we study vorticity within the shallow water system.

32.8.3 Non-rotating gravity waves

For the non-rotating case ($f = 0$), the linear velocity and thickness equations are

$$\frac{\partial \eta'}{\partial t} = -H \nabla \cdot \mathbf{u}' \quad (32.160\text{a})$$

$$\frac{\partial \mathbf{u}'}{\partial t} = -g \nabla \eta'. \quad (32.160\text{b})$$

Equation (32.160a) reveals that horizontal convergence drives temporal changes in the free surface height. Hence, a nonzero horizontal flow convergence is required for gravity waves to exist.

Take the time derivative of the thickness equation and the divergence of the momentum equation

$$\frac{\partial^2 \eta'}{\partial t^2} = -H \frac{\partial(\nabla \cdot \mathbf{u}')}{\partial t} \quad (32.161\text{a})$$

$$\frac{\partial(\nabla \cdot \mathbf{u}')}{\partial t} = -g \nabla^2 \eta'. \quad (32.161\text{b})$$

Time changes in the horizontal divergence are thus driven by curvature in the free surface. Substitution then reveals that the perturbation surface height satisfies the linear wave equation

$$\frac{\partial^2 \eta'}{\partial t^2} - g H \nabla^2 \eta' = 0. \quad (32.162)$$

The complement substitution leads to the slightly more general wave equation for the perturbation horizontal velocity

$$\frac{\partial^2 \mathbf{u}'}{\partial t^2} - g H \nabla(\nabla \cdot \mathbf{u}') = 0. \quad (32.163)$$

We discussed the linear wave equation in Section 3.7, revealing it to be the canonical hyperbolic partial differential equation. An appreciation of that discussion is useful though not necessary for the following examination of gravity wave properties.

Wave ansatz and dispersion relation

The linearized shallow water system (32.160a)-(32.160b) exhibits horizontal spatial symmetry. We are thus motivated to consider the linear wave ansatz¹⁰

$$(u', v', \eta') = (\tilde{u}, \tilde{v}, \tilde{\eta}) e^{i(\mathbf{k} \cdot \mathbf{x} - \omega t)}, \quad (32.164)$$

where

$$i = \sqrt{-1} \quad (32.165)$$

and where the real part of the right hand side is assumed. The amplitudes $(\tilde{u}, \tilde{v}, \tilde{\eta})$ are generally *complex* numbers that are independent of space and time. The vector

$$\mathbf{k} = \hat{\mathbf{x}} k + \hat{\mathbf{y}} l \quad (32.166)$$

is the horizontal vector wave-number and ω is the radial frequency with

$$\tau_{\text{period}} = 2\pi/\omega \quad (32.167)$$

¹⁰ *Ansatz* is a German word meaning “educated guess”. For the linear waves considered here, the wave ansatz (32.164) is a valid solution of the linearized shallow water equations (32.160a)-(32.160b) if the dispersion relation (32.170) is satisfied.

the wave period.¹¹

Making use of the linearized momentum equation (32.160b) allows us to connect the amplitude of the free surface undulation to the velocity

$$\tilde{\mathbf{u}} = \mathbf{k} \frac{g \tilde{\eta}}{\omega} \implies \tilde{\mathbf{u}} \wedge \mathbf{k} = 0 \quad \text{and} \quad \tilde{\mathbf{u}} \cdot \mathbf{k} = \frac{g \tilde{\eta} |\mathbf{k}|^2}{\omega}. \quad (32.168)$$

These relations mean that the horizontal velocity of a fluid particle within the wave is oriented parallel to the horizontal wave vector. If fluid particle motion is parallel to the wave, then the wave is said to be *longitudinal*. However, we recall that a shallow water fluid has a vertical velocity that is a linear function of depth as given by equation (31.35)

$$w(z) = w(\eta_b) - (z - \eta_b) \nabla \cdot \mathbf{u}. \quad (32.169)$$

The presence of vertical motion within the wave means that shallow water gravity waves, which have horizontal wave vectors, are not purely longitudinal; it is just the particle's horizontal motion that is longitudinal.

Substitution of the surface height fluctuation (32.164) into the wave equation (32.162) leads to the relation between frequency and wave-number

$$\omega = \pm c |\mathbf{k}|, \quad (32.170)$$

where

$$c = \sqrt{g H} \quad (32.171)$$

is the shallow water gravity wave speed. We emphasize that the wave ansatz (32.164) is a valid solution of the linearized shallow water equations (32.160a)-(32.160b) only if the *dispersion relation* (32.170) is satisfied.

The dispersion relation (32.170) indicates that each wave vector corresponds to a single wave frequency (mod the sign ambiguity, which corresponds to waves moving in opposite directions). Hence, there is no mixing, or dispersion, between waves of different wavenumber or frequency. Non-rotating shallow water gravity waves are thus a type of *non-dispersive* wave. Furthermore, use of the dispersion relation brings the amplitude relations (32.168) into the form

$$\tilde{\mathbf{u}} = \tilde{\eta} \frac{\mathbf{k}}{|\mathbf{k}|} \sqrt{\frac{g}{H}} \implies \tilde{\mathbf{u}} \cdot \mathbf{k} = \tilde{\eta} \frac{\omega}{H}. \quad (32.172)$$

Since the wave vector is real, this relation means that the velocity and the free surface move in phase as part of the linear wave.

Vanishing relative vorticity

Equation (32.161b) indicates that the linear fluctuations have a horizontal divergence driven by curvature in the surface height. In contrast, because the linearized velocity fluctuation is driven by the gradient of the surface height (see equation (32.160b)), the associated vorticity has a zero time tendency

$$\frac{\partial (\nabla \wedge \mathbf{u}')}{\partial t} = 0. \quad (32.173)$$

¹¹It is important to distinguish the radial frequency, ω , from the vector vorticity, $\boldsymbol{\omega}$, used elsewhere in this book.

This result follows from the discussion of vorticity for the shallow water waves in Section 32.8.2. Hence, if the initial flow configuration has zero vorticity, the linear gravity waves retain zero vorticity as they propagate. The velocity fluctuation (32.168) indeed has zero vorticity since

$$\hat{z} \cdot (\nabla \wedge \mathbf{u}') = i \hat{z} \cdot (\mathbf{k} \wedge \tilde{\mathbf{u}}) e^{i(\mathbf{k} \cdot \mathbf{x} - \omega t)} = 0, \quad (32.174)$$

where we used equation (32.168) to set $\mathbf{k} \wedge \tilde{\mathbf{u}} = 0$.

32.8.4 Potential vorticity of linear fluctuations

Returning now to the $f \neq 0$ case, we rearrange the linearized vorticity equation (32.159) to reach the linear conservation law

$$\frac{\partial q}{\partial t} = 0, \quad (32.175)$$

where the linearized shallow water potential vorticity is given by¹²

$$q = \zeta' - \frac{f\eta'}{H}. \quad (32.176)$$

For the rotating case, we see that potential vorticity is locally constant in the presence of linear fluctuations, such as the inertia-gravity waves discussed in Section 32.8.5. We make use of this constraint in Section 35.7 when studying geostrophic adjustment.

32.8.5 Inertia-gravity waves

Now let the Coriolis parameter, f , be a nonzero constant, so that the linearized thickness equation (32.154) and momentum equation (32.155) take the form

$$\frac{\partial \eta'}{\partial t} = -H \nabla \cdot \mathbf{u}' \quad (32.177a)$$

$$\frac{\partial \mathbf{u}'}{\partial t} + f \hat{z} \wedge \mathbf{u}' = -g \nabla \eta'. \quad (32.177b)$$

The resulting linear fluctuations are known as *inertia-gravity* or *Poincaré waves*. The name “inertia-gravity” is due to the presence of both the Coriolis frequency f and gravity g . Recall our discussion of inertial oscillations in Section 11.5, which describe free particle motion in a rotating reference frame. Inertia-gravity waves arise from the combination of inertial oscillations and non-rotating gravity wave oscillations from Section 32.8.3. That is, both f and g play a role as restoring forces for the waves.

The continuity equation, as a kinematic constraint, has the same form for both rotating and non-rotating systems. Hence, its linearized version, equation (32.177a), remains the same as for the non-rotating case in equation (32.160a). Like the non-rotating case, convergence in the horizontal flow drives surface height tendencies. Hence, horizontal flow convergence is required to support gravity waves in both the rotating and non-rotating systems.

¹²The linearized potential vorticity (32.176) can be obtained by taking the limit of $|f| \gg |\zeta|$ and $H \gg \Delta\eta$ in the shallow water potential vorticity, $Q = (f + \zeta)/h$, studied in Section 35.4.

32.8.6 Inertia-gravity wave dispersion relation

The linearized shallow water system (32.177a)-(32.177b) exhibits horizontal spatial symmetry given that f is assumed constant. We are thus motivated to consider the linear wave ansatz

$$(u', v', \eta') = (\tilde{u}, \tilde{v}, \tilde{\eta}) e^{i(\mathbf{k} \cdot \mathbf{x} - \omega t)}, \quad (32.178)$$

which is the same as assumed for the non-rotating case in Section 32.8.3. Although the same ansatz, the dispersion relation and amplitude relations differ since $f \neq 0$. In particular, we here need to make use of the complex nature of the wave amplitudes $(\tilde{u}, \tilde{v}, \tilde{\eta})$ to realize nontrivial solutions.

Substituting the wave ansatz (32.178) into equations (32.177a)-(32.177b) renders the homogeneous matrix-vector equation

$$\begin{bmatrix} -i\omega & -f & igk \\ f & -i\omega & igl \\ iHk & iHl & -i\omega \end{bmatrix} \begin{bmatrix} \tilde{u} \\ \tilde{v} \\ \tilde{\eta} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}. \quad (32.179)$$

This equation has a nontrivial solution only when the determinant of the matrix vanishes. The real part of the determinant cancels exactly, thus leaving just the imaginary part. Setting the imaginary part to zero yields the dispersion relation

$$\omega [\omega^2 - f^2 - gH(k^2 + l^2)] = 0. \quad (32.180)$$

There are three solutions to this cubic equation to which we now describe.

Geostrophic motion

The zero frequency solution, $\omega = 0$, corresponds to f -plane geostrophic motion. Such motion has zero horizontal divergence since the fluid is time dependent, in which case the linearized continuity equation (32.177a) has $\nabla \cdot \mathbf{u}' = 0$. We studied shallow water geostrophic motion in Section 32.2. We thus move onto the main focus of this chapter, which is the ageostrophic motion associated with inertia-gravity waves in which the horizontal flow is divergent.

Inertia-gravity wave solutions

The two $\omega \neq 0$ solutions to the dispersion relation (32.180) satisfy the quadratic dispersion relation

$$\omega^2 = f^2 [1 + L_d^2 (k^2 + l^2)], \quad (32.181)$$

where

$$L_d = \frac{c}{f} = \frac{\sqrt{gH}}{f} \quad (32.182)$$

is the *deformation radius* for the shallow water layer. As seen in the following, scales smaller than the deformation radius are well approximated by non-rotating physics, whereas rotation is central to the dynamics at scales larger than the deformation radius.

The non-zero frequency modes satisfying the dispersion relation (32.181) are inertia-gravity waves and Figure 32.9 depicts this relation. These waves have a frequency, ω , whose magnitude is greater than the inertial frequency

$$\omega^2 \geq f^2, \quad (32.183)$$

and are thus said to be *super-inertial waves*.

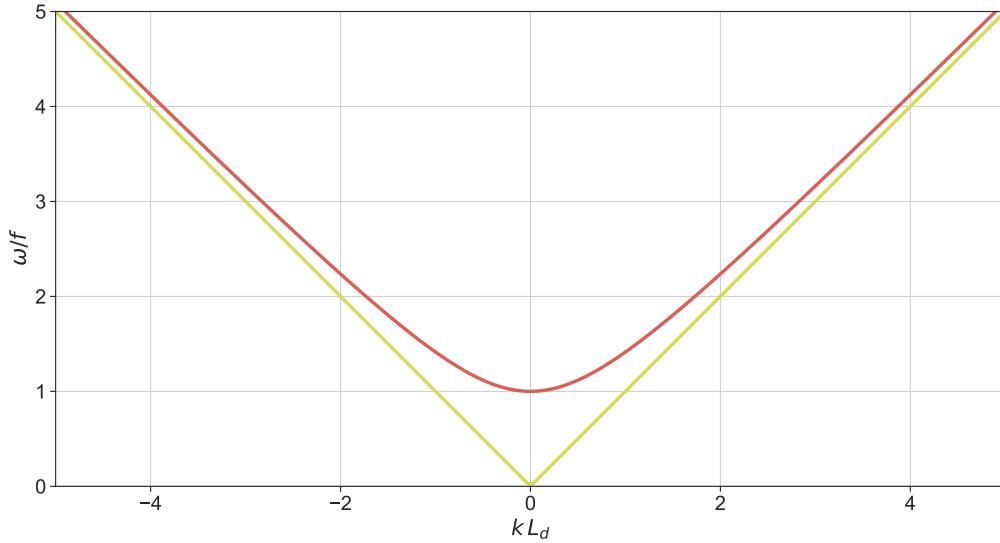


FIGURE 32.9: Illustrating the dispersion relation (32.181) for shallow water inertia-gravity waves (hyperbolic curve). Frequency is scaled by the Coriolis frequency, f , and horizontal wavenumber magnitude is scaled by the deformation radius $L_d = \sqrt{gH}/f$. For small wave numbers ($|\mathbf{k}| L_d \ll 1$ or $\lambda \gg 2\pi L_d$), the inertial-gravity wave frequency approaches the inertia frequency, f , with this behavior seen at the minimum of the dispersion curve. We expect this result since waves large relative to the deformation radius feel the Coriolis acceleration. At the opposite extreme of high wave numbers ($\lambda \ll 2\pi L_d$), the wave frequency approaches the non-rotating gravity wave frequency, shown here by the linear dispersion relation $\omega = k \sqrt{gH}$. Waves small relative to the deformation radius do not feel the Coriolis acceleration and thus converge to non-rotating gravity waves. Since all inertia-gravity waves satisfy $\omega^2 \geq f^2$, they are said to be *super-inertial* waves; i.e., waves whose frequency is larger in magnitude than the inertial frequency.

Shortwave limit

The short wave limit is in the regime where

$$k^2 + l^2 \gg \frac{f^2}{gH} = \frac{1}{L_d^2}, \quad (32.184)$$

so that the shortwave limit occurs when the wavelength is much shorter than the deformation radius. For a wave moving in the $\hat{\mathbf{x}}$ direction with wavelength $\lambda = 2\pi/k$, the shortwave limit occurs when

$$\lambda \ll 2\pi L_d. \quad (32.185)$$

Note that to remain consistent with the shallow water limit with a small vertical to horizontal aspect ratio (i.e., a hydrostatic fluid layer), the wavelength must be longer than the layer thickness, H . Finally, for the shortwave limit, the dispersion relation (32.181) reduces to the non-rotating dispersion relation (32.170)

$$\omega \approx \pm |\mathbf{k}| c = \pm |\mathbf{k}| \sqrt{gH}. \quad (32.186)$$

We see that waves smaller than the deformation radius are only weakly affected by the Coriolis acceleration so that their dispersion relation reduces to linear non-rotating gravity waves of Section 32.181.

Longwave limit

The opposite limit occurs when

$$k^2 + l^2 \ll \frac{1}{L_d^2}, \quad (32.187)$$

so the waves are much longer than the deformation radius. In this limit the dispersion relation is

$$\omega = \pm f. \quad (32.188)$$

This dispersion relation is satisfied by inertial oscillations. As discussed in Section 11.5, inertial oscillations are unaffected by pressure forces otherwise arising from surface height undulations. Instead, they are determined only by the Coriolis frequency. Hence, in this longwave limit the fluid particles move as inertially oscillating particles.

32.8.7 Horizontally longitudinal and transverse components

As discussed in Section 32.8.3, non-rotating gravity waves are horizontally longitudinal, which means that horizontal fluid particle motion induced by the waves is in the same direction as the wave vector. In contrast, we will see that in the presence of rotation, the waves develop a horizontally transverse wave component in addition to the longitudinal component.

Substituting the wave ansatz (32.178) into the linearized continuity equation (32.177a) leads to

$$\mathbf{k} \cdot \tilde{\mathbf{u}} = \tilde{\eta} \frac{\omega}{H}, \quad (32.189)$$

which is the same as for the non-rotating case given by equation (32.172). The wave ansatz (32.178) in the velocity equation (32.177b) gives

$$\hat{\mathbf{z}} \cdot (\mathbf{k} \wedge \tilde{\mathbf{u}}) = -\tilde{\eta} \frac{i f}{H}, \quad (32.190)$$

which results from the vanishing linearized potential vorticity as per equation (32.176). In reading this equation, it is important to recall that the wave amplitudes, $(\tilde{\mathbf{u}}, \tilde{\eta})$, are generally complex numbers.

Making use of equations (32.189) and (32.190) leads to the following decomposition of the wave velocity amplitude

$$\tilde{\mathbf{u}} = \frac{\tilde{\eta}}{H |\mathbf{k}|^2} [\omega \mathbf{k} - i f (\mathbf{k} \wedge \hat{\mathbf{z}})], \quad (32.191)$$

which in turn leads to the wave velocity

$$\mathbf{u}' = \frac{\tilde{\eta}}{H |\mathbf{k}|^2} \left[\underbrace{\omega \mathbf{k} \cos(\mathbf{k} \cdot \mathbf{x} - \omega t)}_{\text{horizontally longitudinal}} + \underbrace{f (\mathbf{k} \wedge \hat{\mathbf{z}}) \sin(\mathbf{k} \cdot \mathbf{x} - \omega t)}_{\text{horizontally transverse}} \right]. \quad (32.192)$$

For the northern hemisphere, where $f > 0$, the wave rotates clockwise relative to the direction of the wave, when looking from above, with the wave vector tracing out an ellipse. The horizontally transverse component of the velocity vector is $\pi/2$ out of phase with the horizontally longitudinal component. Furthermore, the transverse component corresponds to fluid particle motion that is perpendicular to the wave vector, in which case we say that the transverse component is *polarized* perpendicular to the horizontal wave vector.

32.8.8 Comments

There is much more to be said about gravity waves, particularly when having more than a single shallow water layer and then when allowing for continuous stratification. Chapter 7 of [Vallis \(2017\)](#) provides a thorough and pedagogical introduction to the subject. For our purposes, we only wished to whet the appetite. Also, we make use of the basic ideas of gravity waves when discussing geostrophic adjustment in Section 35.7.

32.9 Inertia-gravity waves generated by topography

In this section we consider a geophysically relevant means to force gravity waves via the flow of fluid over topography.

32.9.1 Linearized equations

The physical system consists of flow in a zonally reentrant channel with small amplitude topography, $\eta_b = \eta'_b$. We also introduce a constant zonal mean flow, $U \hat{\mathbf{x}}$. Linear inertia-gravity wave fluctuations are generated as the flow moves over the topography. Our goal is to examine the properties of these waves.

The free surface and velocity take on the assumed form

$$\eta = H + \eta'_b + \eta' \quad \text{and} \quad u = U + u' \quad \text{and} \quad v = 0 + v', \quad (32.193)$$

where primed variables are small so that the product of any two primed variables is neglected. We are unconcerned with how the zonal mean flow is generated, instead focusing on the small amplitude perturbations (u', v', η') . Substituting the expressions (32.193) into the shallow water equations (32.150a) and (32.150b) leads to the linearized equations

$$\frac{\partial \mathbf{u}'}{\partial t} + U \frac{\partial \mathbf{u}'}{\partial x} + f \hat{\mathbf{z}} \wedge \mathbf{u}' = -g \nabla \eta' \quad (32.194a)$$

$$\frac{\partial(\eta' - \eta'_b)}{\partial t} + U \frac{\partial(\eta' - \eta'_b)}{\partial x} = -H \nabla \cdot \mathbf{u}'. \quad (32.194b)$$

We chose to retain η'_b as part of the time derivative even though the topography is time-independent.

32.9.2 Galilean transformation

The background zonal flow is constant in time and zonally symmetric. In this situation it is convenient to introduce a Galilean transformation according to

$$\bar{t} = t \quad \text{and} \quad \bar{x} = x - Ut \quad \text{and} \quad \bar{y} = y. \quad (32.195)$$

Following our discussion of Galilean transformations in Section 13.6, we know that the derivative operators transform according to¹³

$$\frac{\partial}{\partial \bar{t}} = \frac{\partial}{\partial t} + U \frac{\partial}{\partial x} \quad \text{and} \quad \frac{\partial}{\partial \bar{x}} = \frac{\partial}{\partial x} \quad \text{and} \quad \frac{\partial}{\partial \bar{y}} = \frac{\partial}{\partial y}. \quad (32.196)$$

With this coordinate transformation the linear equations (32.194a)-(32.194b) take the form

$$\frac{\partial \mathbf{u}'}{\partial \bar{t}} + f \hat{\mathbf{z}} \wedge \mathbf{u}' = -g \nabla \eta' \quad (32.197a)$$

$$\frac{\partial(\eta' - \eta'_b)}{\partial \bar{t}} = -H \nabla \cdot \mathbf{u}'. \quad (32.197b)$$

Note how the $U \partial_x$ operator has been absorbed into the time derivative, $\partial_{\bar{t}}$. The price to pay for that simplification is that the topography that is independent of the time t , is now a function of the time \bar{t} . That is, in the reference frame moving with the mean flow with velocity $U \hat{\mathbf{x}}$, the topography moves in the opposite direction since

$$\eta'_b(x, y) = \eta'_b(\bar{x} + U \bar{t}, \bar{y}). \quad (32.198)$$

¹³Note that the Galilean transformation considered in Section 13.6 uses the opposite sign for the flow relative to that considered here. There is nothing fundamental about the sign; one merely needs to be careful when translating the earlier results to the current case.

32.9.3 Wave equation and relation between amplitudes

Taking $\partial_{\bar{t}}$ of the continuity equation (32.197b) and then substituting the time derivative of the velocity equation (32.197a) leads to the forced wave equation

$$\frac{\partial^2 \eta'}{\partial \bar{t}^2} - g H \nabla^2 \eta' + f H \zeta' = \frac{\partial^2 \eta'_b}{\partial \bar{t}^2}, \quad (32.199)$$

where

$$\zeta' = \hat{z} \cdot (\nabla \wedge \mathbf{u}') = \partial_x v' - \partial_y u' \quad (32.200)$$

is the relative vorticity of the waves.

Assumptions

To proceed, we make the following simplifying assumptions.

- The topography is a function only of the zonal position

$$\eta'_b = \eta'_b(x) = \eta'_b(\bar{x} + U \bar{t}). \quad (32.201)$$

- Seek a steady solution whereby $\partial_t = 0$ so that $\partial_{\bar{t}} = U \partial_x$.
- Assume zero meridional dependence so that $\partial_y = 0$. This assumption in-effect means that we are working on a domain that is either periodic in y or infinite in y ; i.e., there are no northern or southern walls to the zonal channel.
- Assume the topography has a wave structure

$$\eta'_b(x) = \tilde{\eta}_b e^{ik_b x} = \tilde{\eta}_b e^{ik_b (\bar{x} + U \bar{t})} = \tilde{\eta}_b e^{i(k_b \bar{x} - \omega_b \bar{t})}, \quad (32.202)$$

where $\tilde{\eta}_b$ is the specified amplitude of the topography, k_b is the specified wave number of the topography, and

$$\omega_b = -U k_b \quad (32.203)$$

is the intrinsic frequency of the topography as viewed in the moving reference frame. This monochromatic wave assumption is a strong idealization since realistic topography has many Fourier wave components. Nonetheless, once we obtain the solution for a single wave we can use the principle of superposition available for linear equations to develop a more general solution.

Wave ansatz

What sort of waves do we expect to excite? Without any prior notion, we might expect to excite a variety of waves. However, to simplify the analysis we seek wave solutions with the same wave number, k_b , and intrinsic frequency, ω_b , of the topography. These considerations lead us to study the wave ansatz

$$(u', v', \eta') = (\tilde{u}, \tilde{v}, \tilde{\eta}) e^{i(k_b \bar{x} - \omega_b \bar{t})}. \quad (32.204)$$

We are thus not in search of a dispersion relation, since the wave number and frequency are specified. Instead, we wish to determine the relation between the wave amplitudes. In particular, we will see that knowledge of $\tilde{\eta}$ is sufficient to specify the velocity amplitudes, (\tilde{u}, \tilde{v}) . Furthermore, we express the free surface wave amplitude, $\tilde{\eta}$, in terms of the topography amplitude, $\tilde{\eta}_b$.

Relating the amplitudes

Use of the wave ansatz (32.204) in the wave equation (32.199) leads to

$$(-\omega_b^2 + g H k_b^2) \tilde{\eta} = -\omega_b^2 \tilde{\eta}_b - i f H k_b \tilde{v}. \quad (32.205)$$

To simplify this expression, make use of the velocity equation (32.197a) and continuity equation (32.197b), which take on the modal form

$$\omega_b \hat{z} \cdot (\mathbf{k}_b \wedge \tilde{\mathbf{u}}) = -i f \mathbf{k}_b \cdot \tilde{\mathbf{u}} \quad (32.206a)$$

$$H \mathbf{k}_b \cdot \tilde{\mathbf{u}} = \omega_b (\tilde{\eta} - \tilde{\eta}_b), \quad (32.206b)$$

which, for $\mathbf{k}_b = \hat{x} k_b$, leads to

$$i k_b H \tilde{v} = f (\tilde{\eta} - \tilde{\eta}_b). \quad (32.207)$$

Substitution of equation (32.207) into the dispersion relation (32.205) leads to

$$\tilde{\eta} = \left[\frac{\omega_b^2 - f^2}{\omega_b^2 - f^2 - g H k_b^2} \right] \tilde{\eta}_b = \left[\frac{(f/k_b)^2 - U^2}{(f/k_b)^2 + (g H - U^2)} \right] \tilde{\eta}_b, \quad (32.208)$$

where the second equality follows by setting $\omega_b = -U k_b$ for the intrinsic frequency. Since the ratio on the right hand side is a real number, we see that the free surface is in phase with the bottom topography. Furthermore, note that for most cases of interest, $g H \gg U^2$ since non-rotating gravity wave speeds are typically much larger than the mean flow speeds (except perhaps in very shallow waters). In general, when $g H \leq U^2$ then the flow can be subject to hydraulic jumps since the particle speed overcomes the wave speed (see Exercise 31.2), with such jumps well outside the linear regime considered here. We thus assume $g H \gg U^2$, in which case the amplitude of the free surface undulations are much smaller than the topography undulations

$$|\tilde{\eta}| \ll |\tilde{\eta}_b|. \quad (32.209)$$

From the discussion of topographic form stress in Section 32.6, we know that a net topographic form stress over the extent of the channel requires there to be a nonzero phase shift between the topography and the free surface height. Given the absence of such a phase shift in this example, we see that the waves do not generate a net topographic form stress.



32.10 Exercises

EXERCISE 32.1: POTENTIAL TEMPERATURE SLOPES IN ATMOSPHERE AND OCEAN

Use the two-layer thermal wind relations from Section 32.2.2, also known as Margules' relation, to estimate the slope of the potential temperature surfaces in the atmosphere and ocean. This question is based on exercise 3.2 of [Vallis \(2006\)](#).

- (a) Model the atmosphere as two immiscible shallow water layers of different density stacked one above the other. Using reasonable values for any required physical parameters, estimate the vertical displacement of the interfacial surface associated with a pole-to-equator temperature difference of 40K. You may wish to consult [Wallace and Hobbs \(2006\)](#) or [Marshall and Plumb \(2008\)](#) for physical scales.

- (b) Estimate a vertical interfacial displacement in the ocean thermocline associated with a temperature difference of 20K over a horizontal distance of 4000 km. The interface between the two shallow water layers offers a crude representation of the main oceanic thermocline. Ignore salinity effects so that temperature and density are directly proportional.

Double-check your results by examining some atmosphere and ocean latitude-height profiles for potential temperature (e.g., Figure 5.8 of *Marshall and Plumb (2008)*).

EXERCISE 32.2: CIRCULAR STEADY GEOSTROPHIC FLOW

Consider a single layer of shallow water fluid in steady geostrophic balance on a f -plane so that

$$f \hat{\mathbf{z}} \wedge \mathbf{u}_g = -g \nabla \eta, \quad (32.210)$$

where $f > 0$ (northern hemisphere). Assume the free surface has a circular Gaussian shape

$$\eta = \eta_0 e^{-r^2/(2\sigma^2)} \quad (32.211)$$

where $r^2 = x^2 + y^2$ is the square of the radial position and σ is the standard deviation of the Gaussian.

- Determine the horizontal geostrophic velocity components corresponding to this free surface undulation.
- Determine the streamlines for the flow. Hint: recall the discussion in Section 13.8.2.

EXERCISE 32.3: STEADY STATE MOMENTUM AND GEOSTROPHY

Consider a single layer of shallow water fluid with zero boundary mass fluxes through the surface. Assume the lateral boundaries are solid. All boundaries are thus material. The domain integrated horizontal momentum (within the rotating reference frame) is defined by

$$\mathbf{P} = \int \rho \mathbf{u} dV = \int \rho h \mathbf{u} dA. \quad (32.212)$$

Show that

$$\frac{d\mathbf{P}}{dt} = 0 \quad (32.213)$$

can be realized either by (A) zero flow everywhere, (B) flow that is in geostrophic balance at each point, or (C) flow that is in geostrophic balance as a global integral.

EXERCISE 32.4: THICKNESS WEIGHTED MOMENTUM FOR TWO LAYERS

Following the methods from Section 32.3, derive the thickness weighted momentum equation for an inviscid two-layer stacked shallow water fluid. That is, derive the evolution equation for $h_1 \mathbf{u}_1 + h_2 \mathbf{u}_2$, thus providing the two-layer version of equation (32.23).

EXERCISE 32.5: CONTACT PRESSURE FORCE ON A SINGLE LAYER

As a check on our calculation of the contact pressure force (32.60b), consider a single shallow water layer under a massless atmosphere. Show that the contact pressure force per mass is given by

$$\frac{\mathbf{F}_{\text{net}}^{\text{press}}}{M} = g \hat{\mathbf{z}} - g \nabla \eta. \quad (32.214)$$

As expected, the horizontal component of this force equals to the pressure gradient body force per mass detailed in Section 31.2.1. The vertical pressure force balances the weight of the fluid as per the hydrostatic balance.

EXERCISE 32.6: TOPOGRAPHIC FORM STRESS FOR A RIDGE

As in Section 32.6, apply a constant eastward zonal wind to a zonally reentrant channel with a single shallow water layer. Let the layer flow over a topographic ridge of height H above the surrounding flat bottom, and let the ridge be a function just of zonal position, $\eta'_b(x)$. Furthermore, assume the ridge has a constant slope on both the upstream (west) side, S_{up} , and downstream (east) side, S_{dn} . An example is depicted in Figure 32.10. Following the force balance (32.121), derive an expression for the free surface height zonally averaged over the upstream side of the ridge, minus the free surface height zonally averaged over the downstream sides of the ridge,

$$\Delta\eta' = (\bar{\eta}')_{\text{up}} - (\bar{\eta}')_{\text{dn}}, \quad (32.215)$$

where

$$(\bar{\eta}')_{\text{up}} = \frac{\int_{x_{\text{up}}}^{x_0} \eta' dx}{L_{\text{up}}} \quad \text{and} \quad (\bar{\eta}')_{\text{dn}} = \frac{\int_{x_0}^{x_{\text{dn}}} \eta' dx}{L_{\text{dn}}}. \quad (32.216)$$

Show that the expression for $\Delta\eta'$ is independent of the two slopes. Instead, the only geometric property that determines $\Delta\eta'$ is the ridge height, H . Discuss this result.

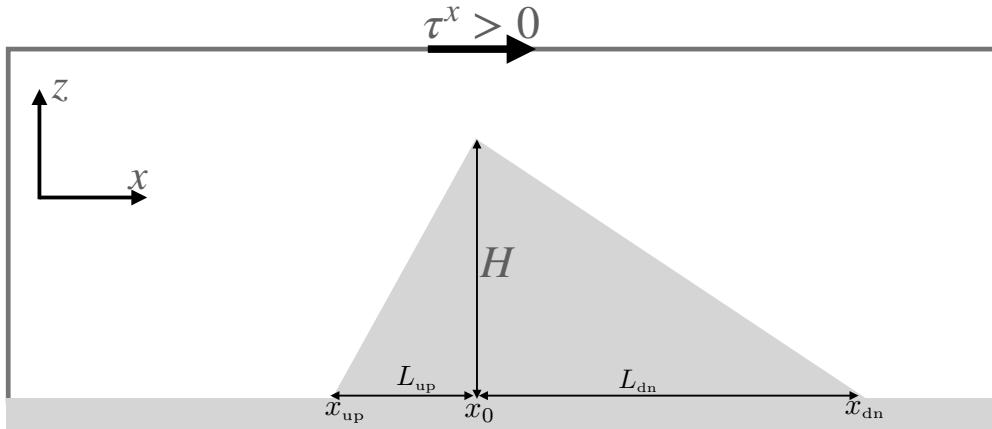


FIGURE 32.10: Zonal-vertical view of a single layer of shallow water fluid moving in a zonally re-entrant channel for use in Exercise 32.6. The domain has a topographic ridge that is a function just of the zonal direction and has constant slopes on its western and eastern sides. There is a constant eastward zonal wind stress.

EXERCISE 32.7: GEOSTROPHIC TRANSPORT

Consider a zonal-vertical section of shallow water flow in the middle latitude northern hemisphere. Let the section be 1000 m deep and away from side and bottom boundaries. Assume the sea level is 1 cm higher at the eastern end of the section than the western end. Estimate the mass transport (kg/sec) of constant density seawater going through the section. What direction is the transport? Hint: Assume geostrophic balance; choose a representative constant seawater density; and note that the zonal width of the section cancels out so it is not needed.

EXERCISE 32.8: APE FOR A SINGLE SHALLOW WATER LAYER WITH BOTTOM TOPOGRAPHY

Generalize the APE discussion in Section 32.5.4 to allow for a nonzero bottom topography, $z = \eta_b(x, y)$. Show that the APE is non-negative, just as for the flat bottom case. Assume the domain is simply connected.

EXERCISE 32.9: APE FOR TWO SHALLOW WATER LAYERS

Compute the APE for two shallow water layers using the notation from Figure 31.6 with nontrivial

bottom topography, $z = \eta_b(x, y)$. Show that the APE is non-negative. Assume the domain is simply connected.

EXERCISE 32.10: RATIO OF KE TO APE FOR SINGLE LAYER f -PLANE GEOSTROPHY

Consider a single layer of shallow water fluid in geostrophic balance on an f -plane with a flat bottom. Show that the ratio of kinetic energy to available potential energy scales like

$$\frac{E_{\text{KE}}}{E_{\text{APE}}} \sim \left[\frac{L_d^{\text{ext}}}{L} \right]^2. \quad (32.217)$$

In this equation, L is the horizontal length scale for the fluctuation of the free surface η (i.e., $\nabla \eta \sim \eta'/L$), and

$$L_d^{\text{ext}} = \frac{\sqrt{g H}}{f} \quad (32.218)$$

is the external deformation radius. The scaling (32.217) means that for scales larger than the external deformation radius, L_d , the available potential energy is larger than the kinetic energy. The converse holds for scales smaller than L_d .

EXERCISE 32.11: RATIO OF KE TO APE FOR 1.5 LAYER f -PLANE GEOSTROPHY

Consider a reduced gravity system (Section 31.3) in geostrophic balance on an f -plane. Show that the ratio of kinetic energy to available potential energy scales like

$$\frac{E_{\text{KE}}}{E_{\text{APE}}} \sim \left[\frac{L_d^{\text{int}}}{L} \right]^2. \quad (32.219)$$

In this equation, L is the horizontal length scale for the fluctuation of the internal interface $\eta_{3/2}$ (i.e., $\nabla \eta_{3/2} \sim \eta'_{3/2}/L$), and

$$L_d^{\text{int}} = \frac{\sqrt{g_{3/2}^r \bar{h}}}{f} \quad (32.220)$$

is the internal deformation radius with $g_{3/2}^r = g(\rho_2 - \rho_1)/\rho_1$ the reduced gravity and $\bar{h} = A^{-1} \int (\eta_{1/2} - \eta_{3/2}) dA$ the area averaged layer thickness (see Figure 31.4). The scaling (32.219) means that for scales larger than the internal deformation radius, L_d^{int} , the available potential energy is larger than the kinetic energy, and conversely for scales smaller than L_d^{int} .

To solve this exercise you must make use of the following.

- Derive the APE for two layers with a flat bottom.
- Assume the contribution to the APE from free surface undulations, is much smaller than from the interior interface. So that the APE is roughly due just to undulations of the interior interface. This assumption follows from Figure 31.5.
- The scaling (32.219) is identical to that found for the quasi-geostrophic system in Section 42.6.4. However, to solve this exercise it is not sufficient to merely reproduce the scaling discussed in Section 42.6.4. Instead, use here the expressions for APE and KE appropriate for the shallow water system.

EXERCISE 32.12: NON-DIMENSIONALIZED LINEAR SHALLOW WATER EQUATIONS

For the linear equations (32.177a)-(32.177b), introduce

$$\mathbf{x} = L \hat{\mathbf{x}}, \quad \mathbf{u}' = U \hat{\mathbf{u}}, \quad t = \frac{L \hat{t}}{U}, \quad f = \hat{f} T^{-1}, \quad \eta' = H \hat{\eta}, \quad (32.221)$$

where L is a horizontal length scale, $T = L/U$ is an advective time scale, U is a velocity scale, and $H \ll L$ is the resting layer thickness. All variables with hats are non-dimensional and not to be confused with unit vectors. Substitute into equations (32.177a)-(32.177b) and identify the non-dimensional ratio of the advective velocity scale to the non-rotating gravity wave speed

$$\text{Fr} \equiv \frac{U}{\sqrt{g H}}. \quad (32.222)$$

This non-dimensional number is known as the *Froude number*. When the Froude number is larger than unity, the fluid can experience an instability known as an *hydraulic jump*, with some elements of jumps discussed in Exercise 31.2.

EXERCISE 32.13: SHALLOW WATER KELVIN WAVES

This exercise works through the basics of Kelvin waves as presented by Section 3.8.3 of [Vallis \(2017\)](#). The Kelvin wave is an inertia-gravity wave that arises from the presence of a boundary¹⁴ and rotation. To develop the basic Kelvin-wave solution, it is convenient to orient the f plane with a solid boundary along a line of constant latitude, $y = y_0$. The meridional velocity component must vanish at this boundary to satisfy the no-normal flow condition. We are thus motivated to consider an ansatz with $v' = 0$ everywhere and see whether there are nontrivial solutions maintaining this structure.

1. Derive the linearized equations of motion for the f -plane shallow water Kelvin wave.
2. Derive the one-dimensional wave equation for the zonal velocity fluctuation.
3. What is the wave speed?
4. Following the discussion in Section 3.7, write the linear wave equation solution in the form

$$u'(x, y, t) = F_1(x + ct, y) + F_2(x - ct, y), \quad (32.223)$$

Verify that this form satisfies the wave equation derived above for any general functions F_1 and F_2 .

5. What is the expression for the surface height displacement, η' ?

From the above we can determine the y -dependence to the Kelvin wave

$$L_d \frac{\partial F_1}{\partial y} = F_1 \quad \text{and} \quad L_d \frac{\partial F_2}{\partial y} = -F_2, \quad (32.224)$$

where $L_d = c/f = \sqrt{g H}/f$ is the shallow water deformation radius given by equation (32.182). Solutions to equations (32.224) are given by

$$F_1 = F(x + ct) e^{(y-y_0)/L_d} \quad \text{and} \quad F_2 = G(x - ct) e^{-(y-y_0)/L_d}. \quad (32.225)$$

To ensure boundedness in the region $y > y_0$ where the fluid is assumed to exist, we drop the F_1 piece of the general solution, thus leaving

$$u' = \hat{x} e^{-(y-y_0)/L_d} G(x - ct) \quad \text{and} \quad \eta' = (H/g)^{1/2} e^{-(y-y_0)/L_d} G(x - ct). \quad (32.226)$$

¹⁴Kelvin waves also occur at the equator, which acts as a boundary due to the change in sign of f . We only consider Kelvin waves that arise from a vertical side wall.

The specific form of the wave signal, $G(x - ct)$, is determined by the initial conditions of the pulse. Regardless the form of this wave pulse, these signals are non-dispersive waves that propagate in the positive \hat{x} direction, in which case the boundary $y = y_0$ is on the right when facing in the direction of propagation. This orientation holds for any boundary orientation in the northern hemisphere, whereby Kelvin waves propagate with the solid boundary on the right when looking in the direction of wave movement. For the southern hemisphere Kelvin waves propagate with the boundary to the left of the wave motion. Hence, Kelvin waves propagate in a cyclonic direction. This result extends to the equator as well, where $f = 0$ at the equator acts as a “wall” that supports the equatorial Kelvin waves.

Kelvin waves have an exponential decay in the direction perpendicular to the boundary with an e-folding length scale given by the deformation radius, L_d . We thus refer to the waves as “trapped waves”, with exponential boundary trapping over a length scale determined by the deformation radius.

EXERCISE 32.14: DEFORMATION RADIUS

The deformation radius appears in many contexts within rotating fluid dynamics. Here, we compute this length scale for selective geophysical flow regimes.

- (a) Compute the shallow water deformation radius for an ocean continental shelf of depth 500 m.
- (b) Compute the shallow water deformation radius for the deep ocean with depth 5000 m.
- (c) The deformation radius defined in this chapter is sometimes called the *external deformation radius* as it makes use of the full depth of the fluid and the gravitational acceleration. In contrast, the deformation radius defined in terms of internal layer thickness and reduced gravity g' leads to the internal deformation radius. The internal deformation radius, $L_d^{\text{int}} = \sqrt{g' h} / f$ is the appropriate rotational length scale for density layers in the interior of the ocean or isentropic layers in the interior of the atmosphere. Compute the deformation radius for a density layer of thickness $h = 200$ m and reduced gravity of $g' = g/1000$.



Part VI

Vorticity and potential vorticity

Vorticity locally measures the spin of a fluid element. It is the natural extension to continuum kinematics of the angular velocity and angular momentum used in rigid body kinematics. For geophysical flows, external forces, ultimately due to differential heating over the planet, resupply vorticity in the face of dissipation. The addition of planetary vorticity, arising from motion on a rotating spherical planet, also renders a nonzero vorticity to geophysical fluids even when the fluid is at rest in the rotating earth reference frame. Vorticity thus plays a central role in geophysical fluid motions, even for relatively sluggish and laminar flow, since motion on a rotating planet always involves vorticity. Furthermore, potential vorticity is a strategically chosen component of vorticity whose conservation properties are simpler than the full vorticity vector. Indeed, under certain assumptions of balance (considered in Part VII), knowledge of potential vorticity offers the means to deduce all prognostic information about the flow. For these reasons, potential vorticity has found great use for understanding and predicting geophysical fluid flows.

OUTLINE OF THIS PART

We start this part of the book by introducing vorticity and circulation in Chapter 33, making use of Stokes' Theorem to show that the area integral of vorticity over a finite region yields the circulation around the region's boundary. Chapter 34 considers the case where the horizontal flow is non-divergent, thus leading to the study of non-divergent barotropic flow. This flow is fully described by the vorticity field, and it offers many insights into large-scale vortical flows in the atmosphere and ocean. Chapter 35 then introduces the mechanics of vorticity and potential vorticity within a shallow water fluid, which builds on the discussion of the simpler non-divergent barotropic model. It was for the shallow water system that [Rossby \(1940\)](#) showed how powerful the notion of potential vorticity conservation is for understanding geophysical fluid flow patterns.

Chapter 36 then dives into the fundamentals of vorticity and circulation. It is here that we encounter Kelvin's Circulation Theorem, which identifies the materially conserved nature of circulation around an arbitrary loop in a perfect barotropic flow. In Chapter 37 we take a brief diversion to compare and contrast vorticity and angular momentum. Chapter 38 presents the foundations of potential vorticity, with Chapter 39 closing this part of the book by studying the remarkable impermeability property of the potential vorticity flux and its implications for potential vorticity budgets.

Vorticity and circulation

Vorticity measures the spin of a flow at each point within the fluid. Vorticity is the generalization to continuum kinematics of the angular velocity relevant to rigid body kinematics. Circulation also measures spin, but as defined over a closed line integral (circuit) within the fluid. Helmholtz was an early proponent of vorticity whereas Kelvin was a proponent of circulation. These two properties of fluid motion are connected through Stokes' theorem. In this chapter we introduce the basic kinematic notions of vorticity and circulation with a focus on developing intuition.

READER'S GUIDE FOR THIS CHAPTER

This chapter makes use of vector calculus identities for Cartesian coordinates as detailed in Chapter 2. The ideas introduced in this brief chapter are fundamental to the remaining chapters in this part of the book.

33.1	Vorticity	746
33.1.1	Rotation of line elements	746
33.1.2	Rotating reference frame	747
33.1.3	There are no vorticity sources	747
33.1.4	Further study	748
33.2	Irrational flows	748
33.2.1	Characterizing irrational flows	748
33.2.2	Comments	748
33.3	Circulation of the velocity field	749
33.4	The free vortex	749
33.4.1	Motion of a fluid particle	751
33.4.2	Circulation	751
33.5	Translation and solid-body rotation	752
33.5.1	Absolute vorticity	752
33.5.2	Solid-body rotation on a plane	752
33.5.3	Circulation for solid-body rotation	753
33.5.4	Comments	753
33.6	Relative vorticity from curvature and shear	753
33.6.1	Circular flow	754
33.6.2	Generalization to natural coordinates	754
33.6.3	Example vorticities	755
33.7	Exercises	757

33.1 Vorticity

Vorticity is the curl of the velocity field

$$\boldsymbol{\omega} = \nabla \wedge \mathbf{v}. \quad (33.1)$$

Vorticity measures the rotation or spin of fluid motion at each point and, unlike angular momentum, it does so without reference to an origin. In addition to writing vorticity as the curl of the velocity, we may choose to use the equivalent expression

$$\boldsymbol{\omega} = [\nabla \cdot (\mathbf{v} \wedge \hat{\mathbf{x}})] \hat{\mathbf{x}} + [\nabla \cdot (\mathbf{v} \wedge \hat{\mathbf{y}})] \hat{\mathbf{y}} + [\nabla \cdot (\mathbf{v} \wedge \hat{\mathbf{z}})] \hat{\mathbf{z}}. \quad (33.2)$$

That is, a vorticity component in a particular coordinate direction is the divergence of the velocity field after being rotated by $-\pi/2$ around the coordinate axis direction. For example, the vector $\mathbf{v} \wedge \hat{\mathbf{z}}$ is the result of rotating the velocity by $-\pi/2$ radians around the $\hat{\mathbf{z}}$ axis, with the identity

$$\hat{\mathbf{z}} \cdot (\nabla \wedge \mathbf{v}) = \nabla \cdot (\mathbf{v} \wedge \hat{\mathbf{z}}) \quad (33.3)$$

leading to the vertical component of the vorticity in equation (33.2).

Being the curl of a vector, the vorticity transforms as a vector under coordinate rotations. However, vorticity changes sign under mirror symmetry, thus making it a pseudo-vector (Section 1.4.2). A simple means to understand this property is to note that the spinning earth rotates counter clockwise when viewed from above the north pole and clockwise when viewed from below the south pole (see Figure 10.3).

Figure 33.1 provides an example zonal flow with a meridional strain (shear). The vertical component to the vorticity is negative for this flow, as per the right hand rule. Furthermore, an imaginary test “paddle wheel” placed anywhere within this flow spins clockwise about its axis. The nonzero spin of test paddle wheels is a fundamental character of fluid flow with nonzero vorticity.

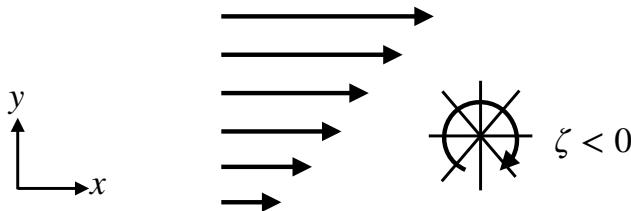


FIGURE 33.1: An example zonal flow with a meridional shear, $\mathbf{v} = u(y) \hat{\mathbf{x}}$, and a corresponding vertical vorticity component that is negative: $\zeta = \hat{\mathbf{z}} \cdot (\nabla \wedge \mathbf{v}) = \partial v / \partial x - \partial u / \partial y = -\partial u / \partial y < 0$. The clockwise arrow surrounds a test “paddle wheel” that exhibits a clockwise spin about its axis when placed in this flow. Such test paddle wheels only spin when there is nonzero vorticity. The right hand rule determines the sign of the vorticity, which for this example is into the page (negative $\hat{\mathbf{z}}$).

33.1.1 Rotation of line elements

In Section 14.2 we considered the kinematic evolution of a line element, $\delta\mathbf{x}$, separating two fluid particles, with that evolution provided by equation (14.31)

$$\frac{D(\delta x_m)}{Dt} = \delta x_n \frac{\partial v_m}{\partial x_n} \implies \frac{D(\delta \mathbf{x})}{Dt} = (\delta \mathbf{x} \cdot \nabla) \mathbf{v}. \quad (33.4)$$

This equation says that the material line element evolves according to the velocity gradient tensor $\partial_n v_m$. The symmetric portion of this tensor is the rate of strain tensor,

$$\mathbb{S}_{mn} = (\partial_n v_m + \partial_m v_n)/2, \quad (33.5)$$

whose action generates changes in the distance between the fluid particles (Section 14.2.7). The anti-symmetric portion to the velocity gradient tensor is known as the rotation tensor,

$$\mathbb{A}_{mn} = (\partial_n v_m - \partial_m v_n)/2. \quad (33.6)$$

The rotation tensor is related to vorticity via equation (14.44)

$$\mathbb{A}_{mn} = -\epsilon_{mnp} \omega_p / 2 \iff \mathbb{A} = \frac{1}{2} \begin{bmatrix} 0 & -\omega_3 & \omega_2 \\ \omega_3 & 0 & -\omega_1 \\ -\omega_2 & \omega_1 & 0 \end{bmatrix}, \quad (33.7)$$

so that

$$2 \mathbb{A}_{mn} \delta x_n = -\epsilon_{mnp} \omega_p \delta x_n \implies 2 \mathbb{A} \cdot \delta \mathbf{x} = \boldsymbol{\omega} \wedge \delta \mathbf{x}. \quad (33.8)$$

From our discussion of rotation in Section 10.4, this equation means that vorticity in a fluid generates a rigid rotation of a material line element (Section 14.2.8). This result accords with Figure 33.1, whereby vorticity leads to the spin of a test paddle wheel; i.e., the rotation of line elements.

33.1.2 Rotating reference frame

As another means to understand the kinematics of vorticity, let us view the flow field from a reference frame that rotates with a constant angular velocity, $\boldsymbol{\Gamma}$. Following equation (10.60e), we know that the velocity observed in the non-rotating reference frame, \mathbf{v} , is related to the rotating reference frame velocity, \mathbf{v}_{rot} , via

$$\mathbf{v} = \mathbf{v}_{\text{rot}} + \boldsymbol{\Gamma} \wedge \mathbf{x}. \quad (33.9)$$

The vorticity in the two reference frames is thus related by

$$\boldsymbol{\omega} = \nabla \wedge \mathbf{v} = \nabla \wedge \mathbf{v}_{\text{rot}} + \nabla \wedge (\boldsymbol{\Gamma} \wedge \mathbf{x}) = \boldsymbol{\omega}_{\text{rot}} + 2 \boldsymbol{\Gamma}. \quad (33.10)$$

If there is a point in the fluid whereby the rotating reference frame's angular velocity equals to one-half the vorticity at that point, $\boldsymbol{\Gamma} = \boldsymbol{\omega}/2$, then the rotating reference frame's vorticity vanishes at that point

$$\boldsymbol{\Gamma} = \boldsymbol{\omega}/2 \implies \boldsymbol{\omega}_{\text{rot}} = 0. \quad (33.11)$$

Hence, we may interpret $\boldsymbol{\omega}/2$ as the angular velocity of the fluid circulating around any point in the fluid. That is, the vorticity is twice the local angular velocity of the fluid. Furthermore, if the vorticity, $\boldsymbol{\omega}$, is spatially constant, then we can move to a rotating reference frame in which the vorticity vanishes everywhere, with such flow referred to as *irrotational*.

33.1.3 There are no vorticity sources

Vorticity has zero divergence

$$\nabla \cdot \boldsymbol{\omega} = \nabla \cdot (\nabla \wedge \mathbf{v}) = 0. \quad (33.12)$$

This property is akin to the non-divergent nature of the velocity vector in an incompressible flow (see Chapter 17). However, vorticity is non-divergent for both compressible and incompressible flow. Consequently, there are no interior sources or sinks of vorticity for any fluid. This very basic property plays an important role in developing some further kinematic properties of vorticity in Chapter 36.

33.1.4 Further study

This video from [3Blue1Brown](#) provides some compelling graphics to help develop intuition for the divergence and curl of a vector, with particular emphasis on fluid flow. In so doing, one learns much about vorticity.

33.2 Irrotational flows

Most geophysical flows have nonzero vorticity. Indeed, even when at rest on the earth, a geophysical fluid carries the vorticity of the rotating planet. However, if we can ignore the planetary vorticity component, as when focused on motions too small to feel the Coriolis acceleration, we can find some geophysically relevant flows with vanishing vorticity. Linear gravity waves in the absence of planetary rotation provide a particularly relevant example (Section 32.8.2).

33.2.1 Characterizing irrotational flows

Irrotational fluid flow is characterized by a zero vorticity

$$\boldsymbol{\omega} = 0 = \text{irrotational flow.} \quad (33.13)$$

Since the curl of a gradient vanishes, irrotational flow has a velocity field equal to the gradient of a velocity potential

$$\nabla \wedge \mathbf{v} = 0 \Rightarrow \mathbf{v} = \nabla \Psi. \quad (33.14)$$

Irrotational flow is therefore sometimes called *potential flow*. Figure 33.2 illustrates a two-dimensional flow field generated by taking the gradient of a scalar potential so that the flow has zero vorticity. In this case, the vertical component of the vorticity vanishes at each point since $\partial v / \partial x = \partial u / \partial y$.

If the flow is incompressible, as in a Boussinesq fluid (Section 24.1), then the velocity potential is a harmonic function since it satisfies Laplace's equation

$$\nabla \cdot \mathbf{v} = 0 \implies \nabla^2 \Psi = 0. \quad (33.15)$$

The study of harmonic functions is a very mature area of mathematical physics, thus providing a great deal of analytic power towards the study of potential flow for incompressible fluids.

33.2.2 Comments

This book does not discuss turbulence in any depth. Nevertheless, we here note that three dimensional turbulence fundamentally relies on vorticity. Hence, irrotational flows, though they may exhibit chaotic motions, are not turbulent since they do not allow for the nonlinear cascade of energy to small spatial scales, with this cascade a fundamental characteristic of three dimensional turbulence. As we will see in Section 36.4, vorticity evolves from sources that tilt and stretch vortex tubes. Vortex stretching is the key source for the turbulent cascade in three dimensional turbulence. Section 3.3 of [Tennekes and Lumley \(1972\)](#) provides a pedagogical discussion of vorticity in the context of three-dimensional turbulence.

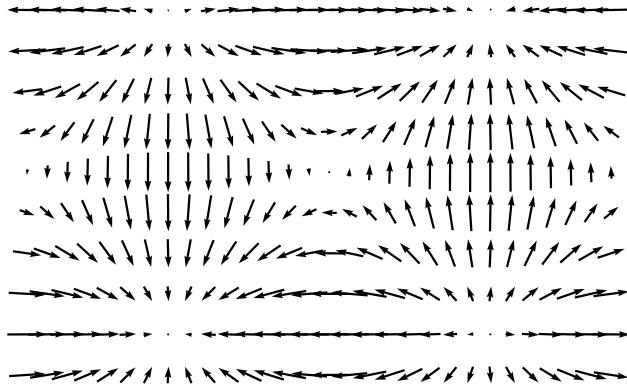


FIGURE 33.2: An example horizontal flow based on a potential, $\Psi = \sin(x/5) \sin(y/5)$. The flow has zero vorticity, $\omega \cdot \hat{z} = \zeta = \partial v / \partial x - \partial u / \partial y = 0$, since the flow is based on a scalar potential: $\omega = \nabla \wedge \mathbf{v} = \nabla \wedge \nabla \Psi = 0$. This example illustrates how irrotational flow may have nontrivial structure even though a test paddle wheel will not spin since there is zero vorticity.

33.3 Circulation of the velocity field

The velocity circulation, or more briefly the *circulation*, is defined as the oriented closed loop line integral of velocity as projected onto the unit tangent of the path

$$C \equiv \oint_{\partial S} \mathbf{v} \cdot d\mathbf{r}, \quad (33.16)$$

with Figure 33.3 offering a schematic. The line element, $d\mathbf{r}$, is oriented in the counter-clockwise direction around the circuit ∂S . More precisely, let $\mathbf{r}(\varphi)$ be an expression for the position of a point on the circuit, with $\varphi(x, y, z, t)$ a parameter that measures the distance along the closed circuit (see Section 2.4). The difference between two very close positions along the circuit defines the increment

$$d\mathbf{r} = \mathbf{r}(\varphi + \delta\varphi) - \mathbf{r}(\varphi). \quad (33.17)$$

By construction, $d\mathbf{r}$ is tangent to the circuit so that $\mathbf{v} \cdot d\mathbf{r}$ picks out the component of the velocity that is tangent to the path.

Stokes' theorem (Section 2.6) renders the very important identity

$$C = \oint_{\partial S} \mathbf{v} \cdot d\mathbf{r} = \int_S (\nabla \wedge \mathbf{v}) \cdot \hat{\mathbf{n}} dS = \int_S \boldsymbol{\omega} \cdot \hat{\mathbf{n}} dS, \quad (33.18)$$

where $\hat{\mathbf{n}}$ is the outward normal vector orienting the area according to the right-hand rule applied to the bounding circuit. The area integral expression motivates interpreting velocity circulation as the “flux of vorticity” penetrating the area. Stokes' theorem provides the means to connect the vorticity theories promoted by Helmholtz to the circulation theories of Kelvin.

33.4 The free vortex

Consider a two-dimensional rotating fluid in the x - y plane with angular velocity given by

$$\boldsymbol{\Omega} = \frac{\mathbf{x} \wedge \mathbf{v}}{r^2} = \frac{K \hat{z}}{r^2}. \quad (33.19)$$

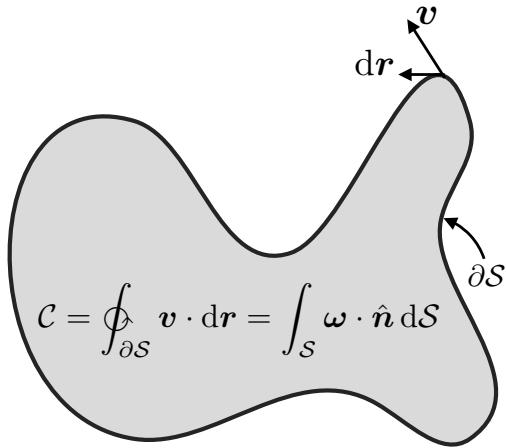


FIGURE 33.3: The velocity circulation around the boundary of a region, $\partial\mathcal{S}$, is determined by the line integral of the velocity projected into the direction of the line integral, $\mathbf{v} \cdot d\mathbf{r}$. Stokes' theorem shows that the velocity circulation computed as a line integral is identical to the normal projection of the vorticity integrated over the area of the region, $\mathcal{C} = \oint_{\partial\mathcal{S}} \mathbf{v} \cdot d\mathbf{r} = \int_S \boldsymbol{\omega} \cdot \hat{\mathbf{n}} dS$.

The constant K has dimensions $L^2 T^{-1}$, and $r^2 = x^2 + y^2$ is the squared distance from the axis of rotation with $\hat{\mathbf{z}}$ the unit vector normal to the x - y plane. The angular velocity falls off as the squared distance from the center, whereas it is singular at the origin. As shown in this section, the fluid flow associated with this *free vortex* has zero vorticity and zero circulation for all points except the origin. Yet the same points with zero vorticity and zero circulation have a constant angular momentum relative to the origin. As shown by Exercise 33.4, and pursued in more detail in Chapter 37, nonzero angular momentum can arise in a fluid with zero circulation so long as there is a nonzero strain within the fluid, such as in the free vortex.

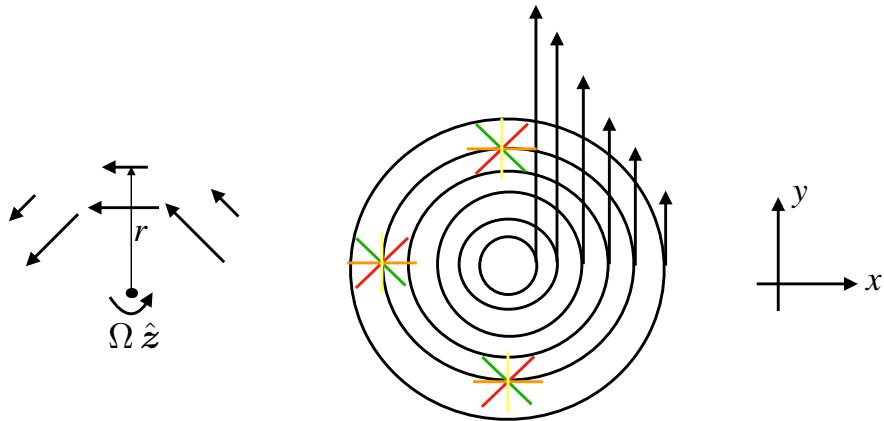


FIGURE 33.4: Irrotational counter-clockwise planar flow in the presence of a free vortex with velocity $\mathbf{v} = (K/r)\hat{\theta}$. The tangential velocity decays as $1/r$ from the origin and the vorticity, $\nabla \wedge \mathbf{v}$, vanishes for all points except the origin. Also, test paddle wheels (colored line segments) do not spin when removed from the center since there is zero vorticity. Notice also that the free vortex has constant angular momentum per mass (computed relative to the origin), since the tangential velocity falls off as $1/r$ thus canceling the moment-arm distance r .

33.4.1 Motion of a fluid particle

A fluid particle moves in a circular orbit when in the free vortex flow field. Hence, the particle velocity is perpendicular to its position vector, $\mathbf{x} = x \hat{\mathbf{x}} + y \hat{\mathbf{y}}$, with respect to the origin

$$\mathbf{v} \cdot \mathbf{x} = 0. \quad (33.20)$$

The velocity for this pure rotational flow is given by (see Section 10.4)

$$\mathbf{v} = \boldsymbol{\Omega} \wedge \mathbf{x} = \frac{K(-y \hat{\mathbf{x}} + x \hat{\mathbf{y}})}{r^2} = \frac{K \hat{\boldsymbol{\vartheta}}}{r}, \quad (33.21)$$

where $\hat{\boldsymbol{\vartheta}}$ is the polar angle unit vector oriented in the counter-clockwise direction (see Section 8.3). Each component of the velocity falls off as $1/r$ when moving away from the origin. Away from the origin, the vorticity vector vanishes

$$\boldsymbol{\omega} = \nabla \wedge \mathbf{v} = 0, \quad (33.22)$$

whereas it is singular at the origin.

As noted above, although vorticity is zero everywhere (except at the origin), the angular momentum is nonzero, as expected since the fluid is rotating around the vortex center. The angular momentum for this system arises just from the nonzero strain in the flow field (see Exercise 33.4), with the strain causing fluid particles to move relative to one another. The angular momentum per unit mass, relative to the center of the vortex, is constant and pointed vertically

$$\mathbf{r} \wedge \mathbf{v} = r \hat{\mathbf{r}} \wedge (K/r) \hat{\boldsymbol{\vartheta}} = K \hat{\mathbf{z}}. \quad (33.23)$$

This result follows since the velocity falls off as $1/r$ to cancel the moment-arm distance, r . Hence, the angular momentum per mass is the same for all fluid particles in the presence of a free vortex, no matter what radial distance the particles have from the vortex center.

We illustrate the free vortex velocity field (33.21) in Figure 33.4 along with test paddle wheels. The paddle wheels remain stationary when placed away from the origin, in the region where the vorticity vanishes. As the paddle wheel centers move counter-clockwise with the flow, the marked paddle wheel blades remain oriented at the same angle. That is, the paddle wheels orbit around the vortex center but they do not spin. The free vortex thus illustrates a fluid flow with non-zero angular momentum yet with zero vorticity (except at the origin).

33.4.2 Circulation

The circulation vanishes for any circuit bounded away from the origin, and it does so trivially since vorticity vanishes away from the origin. However, the circulation is nonzero for any circuit enclosing the origin

$$\mathcal{C} = \oint_{\partial S} \mathbf{v} \cdot d\mathbf{r} = \int_0^{2\pi} \mathbf{v} \cdot \hat{\boldsymbol{\vartheta}} r d\vartheta = 2\pi K. \quad (33.24)$$

To reach this result, we set the line element to

$$d\mathbf{r} = \hat{\boldsymbol{\vartheta}} r d\vartheta \quad (33.25)$$

and inserted the velocity (33.21) represented in cylindrical polar coordinates, $\mathbf{v} \cdot \hat{\boldsymbol{\vartheta}} = K/r$. Hence, the singular point vortex at $r = 0$ induces a nonzero circulation for all circuits that enclose the vortex.

33.5 Translation and solid-body rotation

Rigid or solid-body fluid motion occurs when all fluid particles are rigidly locked into their relative positions, as if in a rigid solid body. There are two kinds of rigid body motion: translation and rotation. The velocity field for this motion is given by

$$\mathbf{v} = \mathbf{U} + \boldsymbol{\Omega} \wedge \mathbf{x}, \quad (33.26)$$

where \mathbf{x} is the position vector relative to the origin, \mathbf{U} is a translation velocity, and $\boldsymbol{\Omega}$ is an angular velocity. For rigid body motion, both \mathbf{U} and $\boldsymbol{\Omega}$ are spatially uniform, but can in general be time dependent. The rate of strain tensor vanishes for uniform translation or solid-body motion (see Exercise 14.1)

$$S_{mn} = \frac{1}{2}(\partial_m v_n + \partial_n v_m) = 0. \quad (33.27)$$

However, the vorticity is nonzero (see Exercise 33.1)

$$\boldsymbol{\omega} = \nabla \wedge (\boldsymbol{\Omega} \wedge \mathbf{x}) = 2\boldsymbol{\Omega}. \quad (33.28)$$

We encountered this vorticity in Section 33.1.2 when connecting vorticity and angular velocity.

33.5.1 Absolute vorticity

For planetary fluid dynamics, rotation of the planet imparts *planetary vorticity* to fluids. Hence, the total or *absolute* vorticity of a fluid is the vector sum of the *relative vorticity*, $\boldsymbol{\omega}$, plus the planetary vorticity

$$\boldsymbol{\omega}_{\text{absolute}} = \boldsymbol{\omega}_{\text{planet}} + \boldsymbol{\omega}. \quad (33.29)$$

In this equation,

$$\boldsymbol{\omega}_{\text{planet}} = 2\boldsymbol{\Omega}_{\text{planet}} \quad (33.30)$$

is the planetary vorticity associated with solid-body motion of a fluid particle stationary with respect to the planet, and

$$\boldsymbol{\omega} = \nabla \wedge \mathbf{v} \quad (33.31)$$

is the relative vorticity. The relative vorticity measures the vorticity of the fluid due to motion relative to the rotating sphere, with \mathbf{v} the velocity relative to the rotating sphere.

33.5.2 Solid-body rotation on a plane

Consider the circular solid-body rotation on a plane shown in Figure 33.5, in which the velocity is purely tangential and linearly proportional to the distance from the center

$$\mathbf{v} = \boldsymbol{\Omega} \wedge \mathbf{x} = |\boldsymbol{\Omega}|(-y \hat{\mathbf{x}} + x \hat{\mathbf{y}}) = |\boldsymbol{\Omega}|r \hat{\boldsymbol{\vartheta}}. \quad (33.32)$$

Assuming the center of mass to be at the circle center, the angular momentum for the flow is the same as that for a solid-body. Even though the motion of each fluid particle is rigidly fixed relative to all other particles, there is a nonzero vorticity in this flow as illustrated by the spin of colored test paddle wheels in Figure 33.5.

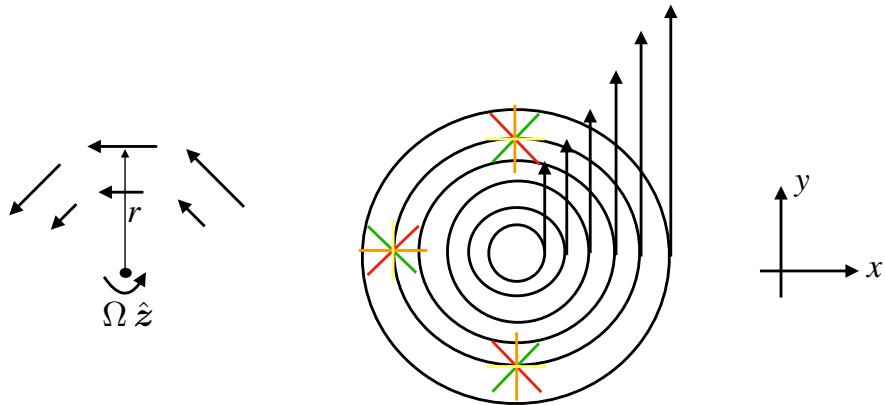


FIGURE 33.5: Rigid body fluid motion, whereby the fluid velocity is: (A) purely tangential and linearly proportional to the radial distance from the vortex center, $\mathbf{v} = |\Omega| r \hat{\vartheta}$; (B) fluid particles maintain a fixed relative position; (C) and vorticity is constant and points perpendicular to the page, $\boldsymbol{\omega} = 2\Omega = 2|\Omega| \hat{z}$. Test paddle wheels rigidly move around the center, and they exhibit a spin about their axis that manifests the nonzero vorticity.

33.5.3 Circulation for solid-body rotation

For solid-body rotation, the velocity circulation around a circular path of radius R is given by

$$\mathcal{C} = \oint \mathbf{v} \cdot d\mathbf{r} = \oint (\boldsymbol{\Omega} \wedge \mathbf{r}) \cdot d\mathbf{r} = R^2 |\Omega| \oint d\vartheta = 2\pi R^2 |\Omega| = 2A |\Omega|, \quad (33.33)$$

where $A = \pi R^2$ is the area of the circle. Hence, the velocity circulation per area for solid-body rotating fluid flow is twice the angular rotation rate, which is the magnitude of the vorticity

$$\mathcal{C}/A = |\boldsymbol{\omega}| = 2 |\Omega|. \quad (33.34)$$

33.5.4 Comments

As seen in Section 33.4, fluid flow in the presence of a free vortex has zero vorticity for all points except the origin of the vortex. However, the same points also have a constant angular momentum relative to the origin, and they experience a nonzero strain. In contrast, constant solid-body rotating fluid flow has a nonzero vorticity, nonzero angular momentum, yet a zero strain. Chapter 37 details the connection between vorticity, strain, and angular momentum, where we see that angular momentum can be nonzero if either vorticity or strain are nonzero. These ideas are illustrated in [this 3-minute video](#) as well as in this [10 minute video from the Open University](#).

33.6 Relative vorticity from curvature and shear

In this section we decompose the vorticity into two terms: one arising from curvature in the flow and another arising from shears in the direction normal to the flow. This decomposition is formulated for horizontal flows, but can be generalized to arbitrary flow. It offers yet another means to understand the kinematic properties of vorticity.

33.6.1 Circular flow

Before treating the general case, it is useful to consider a two-dimensional velocity that locally takes the form of an angular flow

$$\mathbf{u} = u^\vartheta(r, \vartheta) \hat{\vartheta}, \quad (33.35)$$

where ϑ is the polar angle. Circulation around the circular wedge shown in Figure 33.6 has zero contributions from the two radial segments since these segments are perpendicular to the angular flow. The circulation is thus given by

$$\mathcal{C} = \oint_{\text{wedge}} \mathbf{u} \cdot d\mathbf{r} \quad (33.36a)$$

$$= \int_{\vartheta}^{\vartheta + \delta\vartheta} u^\vartheta(r + \delta r, \vartheta') (r + \delta r) d\vartheta' + \int_{\vartheta + \delta\vartheta}^{\vartheta} u^\vartheta(r, \vartheta') r d\vartheta' \quad (33.36b)$$

$$= \int_{\vartheta}^{\vartheta + \delta\vartheta} [u^\vartheta(r + \delta r, \vartheta') (r + \delta r) - r u^\vartheta(r, \vartheta')] d\vartheta' \quad (33.36c)$$

$$\approx \int_{\vartheta}^{\vartheta + \delta\vartheta} \left[u^\vartheta(r, \vartheta') \delta r + \frac{\partial u^\vartheta}{\partial r} r \delta r \right] d\vartheta' \quad (33.36d)$$

$$= r \delta r \int_{\vartheta}^{\vartheta + \delta\vartheta} \left[\frac{u^\vartheta}{r} + \frac{\partial u^\vartheta}{\partial r} \right] d\vartheta', \quad (33.36e)$$

where the approximation holds when $\delta r \rightarrow 0$. Taking the further limit $\delta\vartheta \rightarrow 0$ renders

$$\mathcal{C} \approx \zeta \delta A = \zeta r \delta r \delta\vartheta = \left[\frac{u^\vartheta}{r} + \frac{\partial u^\vartheta}{\partial r} \right] r \delta r \delta\vartheta \implies \zeta = \frac{u^\vartheta}{r} + \frac{\partial u^\vartheta}{\partial r}. \quad (33.37)$$

The first term in the vorticity arises from the nonzero radius of curvature of the circular flow whereas the second term arises from radial shear.

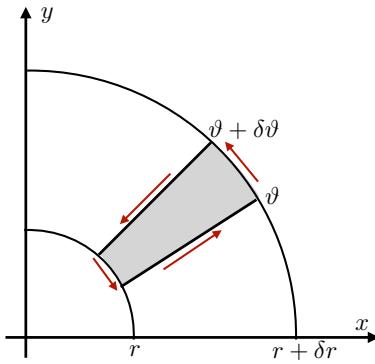


FIGURE 33.6: Circulation around the circular wedge $[r, r + \delta r] \otimes [\vartheta, \vartheta + \delta\vartheta]$.

33.6.2 Generalization to natural coordinates

The decomposition (33.37) can be generalized to arbitrary horizontal flow by making use of the natural coordinates from Section 27.2. Here, we introduce the locally orthogonal coordinates, (s, n) , with s the arc-length defined along the trajectory of a fluid element and n measuring the distance normal to the trajectory. We make the convention that the unit tangent direction, \hat{s} , is aligned

along the local flow direction whereas the unit normal direction, \hat{n} , is to the left facing downstream. Furthermore, the radius of curvature at a point along a trajectory (see Section 5.2.3) is positive if the flow turns into the positive \hat{n} direction (left turn) and negative for oppositely curved flow (right turn) (see Figures 27.2 and 27.3). Finally, the radius of curvature is infinite for straight flow.

For the counter-clockwise circuit in Figure 33.6, a left turn occurs with $\hat{n} = -\hat{r}$ so that equation (33.37) takes on the general form

$$\zeta = \underbrace{\frac{|\mathbf{u}|}{R}}_{\text{curvature}} - \underbrace{\frac{\partial |\mathbf{u}|}{\partial n}}_{\text{shear}} = \zeta_{\text{curv}} + \zeta_{\text{shear}}. \quad (33.38)$$

Again, the first term arises from curvature in the flow, with R the radius of curvature. This *curvature vorticity* is sometimes also called the *orbital vorticity*. A trajectory turning to the left has $R > 0$ and this curved trajectory contributes to a positive vorticity; conversely for a trajectory turning to the right. The second term in equation (33.38) arises from shears computed normal to the flow direction. If the flow speed decreases in the normal direction (e.g., towards the center of the circle in Figure 33.6), then that too contributes to a positive vorticity. Furthermore, flow with $\zeta = 0$ arises if there is an exact compensation between the curvature-induced vorticity with the shear-induced vorticity

$$\zeta = 0 \implies \frac{|\mathbf{u}|}{R} = \frac{\partial |\mathbf{u}|}{\partial n}. \quad (33.39)$$

33.6.3 Example vorticities

Rigid body vortex and free vortex

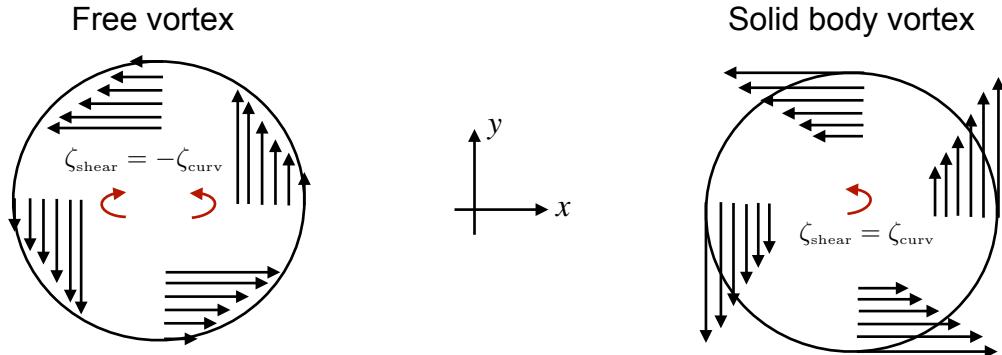


FIGURE 33.7: Decomposing the relative vorticity into its contributions from curvature and normal shear. Left panel: the free vortex from Figure 33.4 has $\zeta_{\text{curv}} = -\zeta_{\text{shear}} = K/r^2$ so that $\zeta = \zeta_{\text{curv}} + \zeta_{\text{shear}} = 0$. Right panel: the solid body vortex from Figure 33.5 has $\zeta_{\text{curv}} = \zeta_{\text{shear}} = \Omega$ so that $\zeta = 2\Omega$.

To help further understand the decomposition (33.38), consider the case of solid body rotation (Figure 33.5) where $\mathbf{v} = \Omega r \hat{\theta}$ and $\hat{n} = -\hat{r}$. For a circle the radius of curvature equals to the radius of the circle, so that

$$\zeta_{\text{curv}} = \frac{\Omega r}{r} = \Omega \quad \text{and} \quad \zeta_{\text{shear}} = -\partial_n |\mathbf{u}| = \partial_r (\Omega r) = \Omega, \quad (33.40)$$

in which case

$$\zeta = \zeta_{\text{curv}} + \zeta_{\text{shear}} = 2\Omega = \hat{z} \cdot (\nabla \wedge \mathbf{v}), \quad (33.41)$$

as depicted in Figure 33.7. Likewise, for the free vortex (Figure 33.4) we have $\mathbf{v} = (K/r)\hat{\vartheta}$ so that

$$\zeta_{\text{curv}} = K/r^2 \quad \text{and} \quad \zeta_{\text{shear}} = \partial|\mathbf{u}|/\partial r = -K/r^2, \quad (33.42)$$

which yields zero relative vorticity

$$\zeta = \zeta_{\text{curv}} + \zeta_{\text{shear}} = 0. \quad (33.43)$$

Gaussian jet moving along a line and around a circle

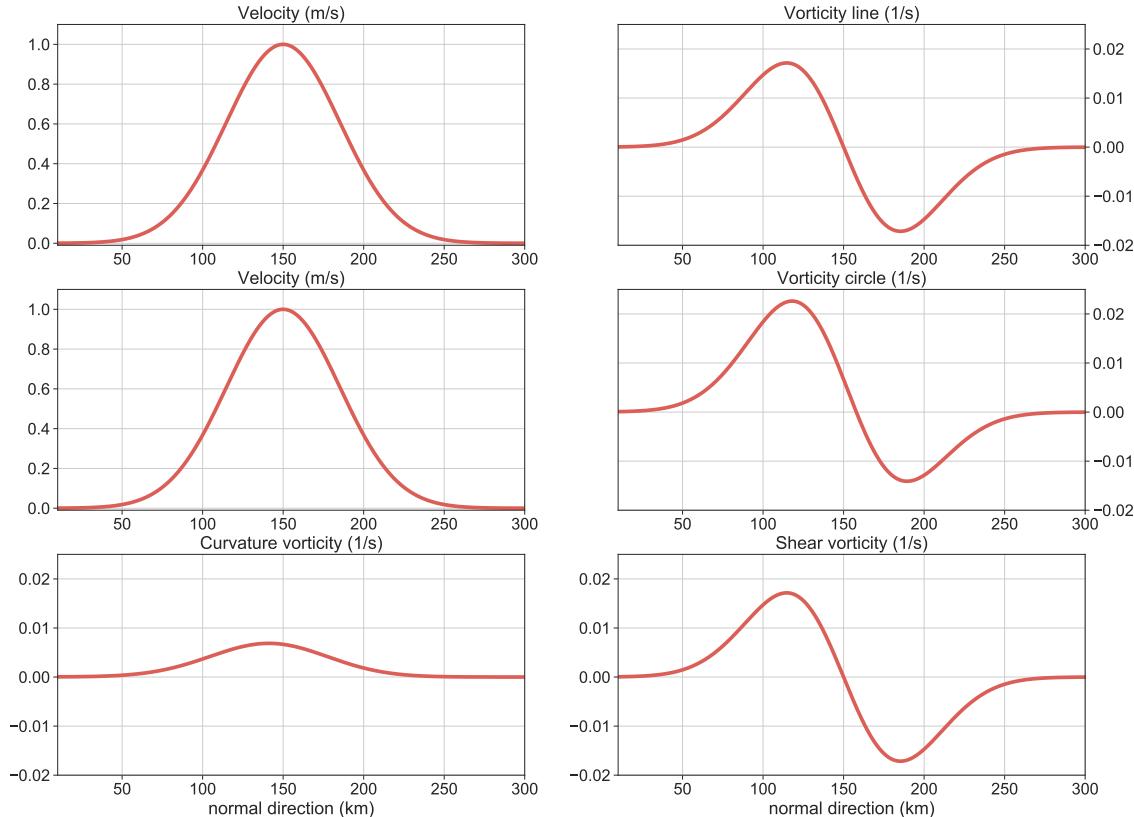


FIGURE 33.8: Top row: velocity and vorticity within a Gaussian jet that is moving along a straight line according to equations (33.44) and (33.45). We set the jet e-folding scale as $L = 50$ km, jet maximum at $x_{\text{max}} = 150$ km, and velocity scale as $v_o = 1 \text{ m s}^{-1}$. Note the symmetry of the vorticity relative to the jet maximum, with the vorticity extrema corresponding to the inflection points of the Gaussian velocity field. Middle left: the same Gaussian velocity profile only now for a jet that is moving counter-clockwise around a circle so that $\mathbf{v} = v(r)\hat{\vartheta}$ according to equation (33.46). To avoid a singularity of the curvature vorticity at $r = 0$ (infinite curvature) we assume the velocity field vanishes within a small distance from the origin. Middle right: vorticity for the Gaussian jet moving around the circle. Middle right: vorticity for the circular jet, $\zeta = r^{-1}\partial_r(rv)$. Notice how the vorticity is not symmetric relative to the jet maximum. Rather, the zero vorticity occurs to the right of the jet maximum, and the vorticity on the inside of the jet maximum is larger in magnitude than the outer vorticity. Lower left: vorticity due to the curvature of the jet as it moves around the circle, $\zeta_{\text{curve}} = v(r)/r$. Lower right: vorticity due to the radial shear in the Gaussian jet, $\zeta_{\text{shear}} = -2(r - r_{\text{max}})v(r)/L^2$.

Next consider a Gaussian jet moving along a straight line in the meridional direction with velocity field

$$\mathbf{v}(x) = v(x)\hat{\mathbf{y}} = v_o \exp[-(x - x_{\text{max}})^2/L^2]\hat{\mathbf{y}}, \quad (33.44)$$

where v_o is the velocity scale, L is the e-folding length scale for the jet, and x_{\max} is the position of the jet maximum. The corresponding vorticity of the jet is given by

$$\zeta_{\text{line}} = \partial_x v = -2[(x - x_{\max})/L^2]v(x). \quad (33.45)$$

We depict the velocity (33.44) and vorticity (33.45) in the top row of Figure 33.8. Note the symmetry of the vorticity around the jet maximum at $x = x_{\max}$, with the vorticity extrema corresponding to inflection points of the jet.¹

Now assume the same jet is moving counter-clockwise around a circle so that the velocity field is given by

$$\mathbf{v}(r) = v(r)\hat{\vartheta} = v_o \exp[-(r - r_{\max})^2/L^2]\hat{\vartheta}, \quad (33.46)$$

where we made use of the polar coordinates from Section 8.3 with $r^2 = x^2 + y^2$ the squared radial distance from the center, and $\hat{\vartheta}$ the unit vector pointing in the counter-clockwise direction from the \hat{x} axis. Making use of the curl operation in polar coordinates given by equation (8.85c) renders the vorticity

$$\zeta = r^{-1}\partial_r(rv) = \frac{v}{r} - \frac{2(r - r_{\max})v}{L^2} = \zeta_{\text{curve}} + \zeta_{\text{shear}}. \quad (33.47)$$

We depict these terms in the third row of Figure 33.8. Note how the vorticity, ζ , has the amplitude of its maximum increased, whereas the amplitude of its minimum is decreased. Correspondingly, the vorticity is not symmetric about the jet maximum, with its zero crossing to the outside of the jet maximum. Asymmetry of the vorticity arises from the curvature vorticity, $\zeta_{\text{curve}} = v/r$, which has its peak inside the jet maximum. So although the shear vorticity, ζ_{shear} , is symmetric around the jet axis, the curvature vorticity causes a movement of vorticity towards the center of the circle.



33.7 Exercises

EXERCISE 33.1: VORTICITY FOR SOLID-BODY ROTATION

Show that a fluid in solid-body rotation with angular velocity

$$\mathbf{v}_{\text{solid-body}} = \boldsymbol{\Omega} \wedge \mathbf{x}, \quad (33.48)$$

has a vorticity given by

$$\nabla \wedge \mathbf{v}_{\text{solid-body}} = 2\boldsymbol{\Omega}. \quad (33.49)$$

EXERCISE 33.2: PLANETARY ROTATION IS NON-DIVERGENT

Show that a fluid in solid-body rotation with angular velocity

$$\mathbf{v}_{\text{solid-body}} = \boldsymbol{\Omega} \wedge \mathbf{x}, \quad (33.50)$$

has zero divergence

$$\nabla \cdot \mathbf{v}_{\text{solid-body}} = 0. \quad (33.51)$$

Consequently, rotation of the planet imparts zero divergence to fluid motion. We make use of this result in part to justify our study of non-rotating fluid kinematics in Part III of this book.

¹Inflection points are where the curvature changes sign.

EXERCISE 33.3: VELOCITY POTENTIAL FOR THE FREE VORTEX

What is the velocity potential (33.14) for the free vortex whose velocity field is given by (33.21)?

Hint: The problem is two-dimensional and rotationally symmetric, so it is convenient to make use of polar coordinates $x = r \cos \vartheta$ and $y = r \sin \vartheta$ as in Appendix 8.3.

EXERCISE 33.4: STRAIN TENSOR FOR THE FREE VORTEX

Determine all components to the strain tensor

$$S_{pq} = \begin{bmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{bmatrix} = \begin{bmatrix} \frac{\partial u}{\partial x} & \frac{1}{2} \left[\frac{\partial u}{\partial y} + \frac{\partial v}{\partial x} \right] \\ \frac{1}{2} \left[\frac{\partial v}{\partial x} + \frac{\partial u}{\partial y} \right] & \frac{\partial v}{\partial y} \end{bmatrix} \quad (33.52)$$

for the free vortex as specified by the velocity field (33.21). Present the answer in the form of a 2×2 matrix.

EXERCISE 33.5: VANISHING VISCOS FRICITION FOR SOLID-BODY MOTION

As discussed in Section 19.7.5, viscous effects from molecular viscosity in an incompressible fluid appear in the momentum equation as a Laplacian weighted by a constant molecular viscosity

$$\text{viscous force per mass} = \nu \nabla^2 \mathbf{v}, \quad (33.53)$$

where $\nu > 0$ is the molecular kinematic viscosity, which is a constant that is a property of the fluid. Show that the viscous operator vanishes for a fluid in solid-body rotation. That is, solid-body motion engenders no frictional dissipation. This result reflects the lack of frictional interaction in a fluid absent shears and strains.

EXERCISE 33.6: VORTICITY FOR A C-GRID NUMERICAL MODEL

Vorticity is commonly diagnosed in numerical model simulations. In this exercise we consider how one might determine a discrete equation for the vertical vorticity when the horizontal velocity is arranged according to the Arakawa C-grid ([Arakawa and Lamb, 1977](#)) commonly used in ocean models, and as depicted in Figure 33.9. Derive an expression for the area averaged vorticity over the shaded grid cell region centered at the vorticity point, $q_{i,j}$. Make use of Stokes' theorem with the surrounding C-grid velocity components and the corresponding grid distances.



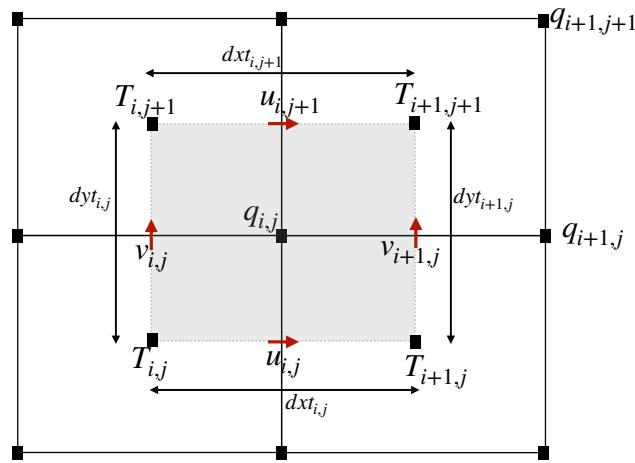


FIGURE 33.9: Layout for velocity on a discrete Arakawa C-grid for use in Exercise 33.6. The central T-point is labeled $T_{i,j}$ and its corresponding vorticity point, $q_{i,j}$, is located to its northeast. This exercise aims to determine the area averaged vorticity for the shaded region. The zonal velocity, $u_{i,j}$, is arranged on the east face of the T-cell, whereas the meridional velocity, $v_{i,j}$, is on the north face. The zonal and meridional grid distances are indicated, thus measuring distances between adjacent tracer points and so measuring the sides of the shaded region.

Non-divergent barotropic flow

A single layer of homogeneous shallow water fluid is among the simplest models available for the study of fluid motion. In the language of vorticity as described in Chapter 36, a shallow water layer is a barotropic fluid since it has zero baroclinicity (Section 36.5). Consequently, the shallow water model is often referred to as a divergent barotropic model, with divergence referring to the nonzero divergence in the horizontal velocity.

For many purposes, the horizontal flow divergence supported by a shallow water layer can be a distraction. That is, for many geophysical studies one is interested in the low frequency vortical (Rossby wave) motions rather than the divergent and higher frequency gravity wave motions (Section 32.8). The *non-divergent barotropic model* focuses on vortical motion since its horizontal velocity has zero divergence. Relative vorticity is the primary dynamical field for the non-divergent barotropic model whereby knowledge of the vorticity is sufficient to fully determine the velocity. Furthermore, the absolute vorticity is materially invariant in the absence of irreversible processes such as friction. That is, the absolute vorticity moves with the horizontal flow without any sources in a perfect fluid non-divergent barotropic model. Hence, meridional motion of the fluid parcel reveals an exact exchange of vorticity between the fluid and the planet. This exchange is beautifully manifest in the physical mechanism for Rossby waves, and it offers a key reason to study this model in the vorticity part of the book.

READER'S GUIDE FOR THIS CHAPTER

We develop some of the analytic and semi-analytic properties available for the non-divergent barotropic model, thus providing an entry-point into the vast literature making use of this model. Besides offering a frequently used theoretical model for large-scale dynamics, the non-divergent barotropic model is useful for studies of coherent vortex structures, with Chapter 3 of [McWilliams \(2006\)](#) exploring analytical vortex solutions. The model also served as the basis for pioneering numerical weather prediction model described by [Charney et al. \(1950\)](#). Since all fields in this chapter are a spatial function only of the horizontal position, $\nabla\psi = \nabla_z\psi$.

34.1	Loose threads	762
34.2	Basic equations and their properties	762
34.2.1	Velocity equation	762
34.2.2	Kinematics of rigid fluid columns	763
34.2.3	Vanishing vertical velocity	764
34.2.4	Velocity self-advection	764
34.2.5	Further reading	765
34.3	Vorticity	765

34.3.1	Vorticity equation	765
34.3.2	Absolute vorticity is the potential vorticity	766
34.3.3	Jacobian form of vorticity advection	766
34.3.4	Taylor-Bretherton identities	767
34.3.5	Poisson equation for the streamfunction	768
34.3.6	Zonal flow as an exact geostrophic solution	768
34.4	Connection to equivalent barotropic flow	768
34.4.1	Vorticity equation for the depth averaged flow	769
34.4.2	Comments	769
34.5	The pressure field	769
34.5.1	Poisson equation for pressure	770
34.5.2	Self-advection source in the pressure equation	770
34.5.3	Gradient wind balance	771
34.5.4	Sample pressure gradients	772
34.5.5	Pressure source from Coriolis acceleration	774
34.5.6	Pressure source from friction	774
34.5.7	Comments and further study	774
34.6	Constraints from absolute vorticity invariance	775
34.6.1	Relative vorticity from curvature and planetary beta	775
34.6.2	Relative vorticity from curvature and normal shears	777
34.6.3	Curvature, shear, and planetary contributions	778
34.7	Barotropic Rossby waves	778
34.7.1	Dispersion relation	779
34.7.2	Phase speed	779
34.7.3	Group velocity	780
34.7.4	Importance of β for Rossby waves	780
34.8	Exercises	781

34.1 Loose threads

- Constraints from absolute vorticity invariance in Section 34.6 need more discussion and figures. Bring in some of the material from Chapter 18 of *Cushman-Roisin and Beckers* (2011).

34.2 Basic equations and their properties

The non-divergent barotropic model arises from the following assumptions.

- The flow occurs in a single homogeneous fluid layer, as for the shallow water model from Chapter 31.
- The horizontal velocity is non-divergent: $\nabla \cdot \mathbf{u} = \partial_x u + \partial_y v = 0$, which means that the thickness of the layer is fixed.

The second property distinguishes this model from a single layer of shallow water fluid. It provides some rather strict constraints on the flow relative to the shallow water layer. In this section we exhibit the governing equations and derive some of their properties.

34.2.1 Velocity equation

The velocity equation for the non-divergent barotropic model follows in a manner akin to the shallow water model from Chapter 31. Namely, the fluid is a homogeneous (uniform and constant

density) layer so that the horizontal velocity satisfies

$$\frac{D\mathbf{u}}{Dt} + f \hat{\mathbf{z}} \wedge \mathbf{u} = -\nabla\varphi \quad \text{and} \quad \nabla \cdot \mathbf{u} = 0, \quad (34.1)$$

where the pressure is normalized according to

$$\varphi = p/\rho \quad (34.2)$$

with ρ the constant layer density, and where material evolution occurs with the two-dimensional non-divergent flow

$$\frac{D}{Dt} = \frac{\partial}{\partial t} + \mathbf{u} \cdot \nabla. \quad (34.3)$$

All fields are depth independent and there is no vertical motion, so that the motion occurs in rigid fluid columns. Furthermore, the horizontal non-divergent flow can be described by a streamfunction

$$\mathbf{u} = \hat{\mathbf{z}} \wedge \nabla\psi \implies u = -\frac{\partial\psi}{\partial y} \quad \text{and} \quad v = \frac{\partial\psi}{\partial x}. \quad (34.4)$$

34.2.2 Kinematics of rigid fluid columns

Recall the thickness equation (31.17) for a shallow water layer

$$\frac{Dh}{Dt} = -h \nabla \cdot \mathbf{u}, \quad (34.5)$$

where $h = \eta - \eta_b$ is the column thickness, $z = \eta(x, y, t)$ is the upper layer interface (the free surface), and $z = \eta_b(x, y)$ is the lower interface (the bottom topography) (see Figure 31.1). With zero divergence in the horizontal velocity, the thickness of a fluid column is constant when moving with the horizontal flow

$$(\partial_t + \mathbf{u} \cdot \nabla) h = \partial_t \eta + \mathbf{u} \cdot \nabla(\eta - \eta_b) = 0. \quad (34.6)$$

We satisfy this equation by choosing the upper surface to be rigid and flat

$$\eta = 0 \quad (34.7)$$

and furthermore by constraining the horizontal flow to follow the bottom topography

$$\mathbf{u} \cdot \nabla\eta_b = (\nabla\psi \wedge \nabla\eta_b) \cdot \hat{\mathbf{z}} = 0. \quad (34.8)$$

In brief, we are concerned with horizontally non-divergent flow within a homogeneous fluid layer for which a fluid column does not expand or contract and so it only moves along lines of constant topography.

For the case of a flat bottom domain that is bounded on its side by a vertical wall, then the no-flow boundary condition takes the form

$$\mathbf{u} \cdot \hat{\mathbf{n}} = (\hat{\mathbf{z}} \wedge \nabla\psi) \cdot \hat{\mathbf{n}} = \hat{\mathbf{t}} \cdot \nabla\psi = 0, \quad (34.9)$$

where $\hat{\mathbf{n}}$ is the horizontal outward unit vector at the boundary, and $\hat{\mathbf{t}} = \hat{\mathbf{n}} \wedge \hat{\mathbf{z}}$ is the unit tangent vector along the boundary. The kinematic condition (34.9) means that the streamfunction is a constant along the solid vertical sidewall boundary. We encountered this boundary condition in Section 17.4.2 when studying two-dimensional non-divergent flow. A zonally reentrant channel is an example considered in Section 34.7, whereby ψ is a constant along the northern and southern walls.

34.2.3 Vanishing vertical velocity

Without horizontal divergence and with a flat free surface, there is identically zero vertical motion within the layer

$$w = 0. \quad (34.10)$$

Another manner to deduce this property by noting that the surface kinematic boundary condition for a static and flat free surface leads to

$$w(\eta) = \frac{\partial \eta}{\partial t} + \mathbf{u} \cdot \nabla \eta = 0 + 0, \quad (34.11)$$

so that the vertical velocity at $z = \eta$ vanishes. With $w(\eta) = 0$, and with $\nabla \cdot \mathbf{u} = 0$ for the horizontal velocity, then $w = 0$ throughout the layer. Correspondingly, this constraint means that the no-normal flow bottom kinematic condition (Section 15.4.1) renders a horizontal velocity that is aligned with the topography, $\mathbf{u} \cdot \nabla \eta_b = 0$, which is a property we already encountered in Section 34.2.2. Since the vertical velocity is zero, the gravitational potential energy is a uniform constant, so that the mechanical energy arises just from kinetic energy.

34.2.4 Velocity self-advection

The velocity self-advection appearing in the velocity equation (34.1) can be written as the divergence of a 2×2 symmetric tensor

$$-(\mathbf{u} \cdot \nabla) \mathbf{u} = \nabla \cdot \mathbb{E} \iff -u_m \partial_m u_n = \partial_m \mathbb{E}_{mn}, \quad (34.12)$$

where

$$\mathbb{E} = \begin{bmatrix} -u^2 & -uv \\ -uv & -v^2 \end{bmatrix}, \quad (34.13)$$

thus bringing the momentum equation (34.1) to the Eulerian form

$$\frac{\partial \mathbf{u}}{\partial t} + f \hat{\mathbf{z}} \wedge \mathbf{u} = -\nabla \varphi + \nabla \cdot \mathbb{E}. \quad (34.14)$$

The tensor,

$$\rho \mathbb{E} = -\rho \mathbf{u} \otimes \mathbf{u} \quad (34.15)$$

is the *kinetic stress tensor*, whose three-dimensional form was introduced in Section 19.5.

In anticipation of our study of vorticity in Section 34.3, we find it useful to decompose the kinetic tensor into its horizontally isotropic and horizontally anisotropic components¹

$$\mathbb{E} = \begin{bmatrix} -u^2 & -uv \\ -uv & -v^2 \end{bmatrix} = -\mathcal{K} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} + \begin{bmatrix} -(u^2 - v^2)/2 & -uv \\ -uv & (u^2 - v^2)/2 \end{bmatrix} \equiv -\mathcal{K} \mathbb{I} + \mathcal{E}, \quad (34.16)$$

where we introduced the kinetic energy per mass

$$\mathcal{K} = \mathbf{u} \cdot \mathbf{u}/2 = (u^2 + v^2)/2, \quad (34.17)$$

and the trace-free anisotropic portion of the kinetic stress tensor

$$\mathcal{E} = \mathcal{K} \mathbb{I} + \mathbb{E} = \begin{bmatrix} -(u^2 - v^2)/2 & -uv \\ -uv & (u^2 - v^2)/2 \end{bmatrix} \iff \mathcal{E}_{mn} = \mathcal{K} \delta_{mn} - u_m u_n. \quad (34.18)$$

Making use of the decomposition (34.16) brings the velocity equation (34.19) to the form

$$\frac{\partial \mathbf{u}}{\partial t} + f \hat{\mathbf{z}} \wedge \mathbf{u} = \nabla \cdot [\mathcal{E} - \mathbb{I}(\mathcal{K} + \varphi)]. \quad (34.19)$$

¹Recall our discussion of isotropy in Section 1.8.

34.2.5 Further reading

As discussed by [Hoskins et al. \(1983\)](#), [Waterman and Hoskins \(2013\)](#) and [Waterman and Lilly \(2015\)](#), the decomposition (34.16) leads to a very insightful kinematic geometric interpretation of eddying flow features.

34.3 Vorticity

For the two-dimensional non-divergent fluid, the vertical component of the relative vorticity is given by the Laplacian of the streamfunction

$$\zeta = \hat{z} \cdot (\nabla \wedge \mathbf{u}) = \frac{\partial v}{\partial x} - \frac{\partial u}{\partial y} = \nabla^2 \psi. \quad (34.20)$$

We here study the motion of this fluid on a tangent plane in which the Coriolis parameter is given by the β -plane expression

$$f = f_0 + \beta (y - y_0). \quad (34.21)$$

34.3.1 Vorticity equation

To form the dynamical equation for the vorticity, take the zonal derivative of the meridional momentum equation (see equation (34.1)) and meridional derivative of the zonal momentum equation

$$\frac{\partial}{\partial t} \left(\frac{\partial v}{\partial x} - \frac{\partial u}{\partial y} \right) + \frac{\partial}{\partial x} [\nabla \cdot (\mathbf{u} v)] - \frac{\partial}{\partial y} [\nabla \cdot (\mathbf{u} u)] + f \nabla \cdot \mathbf{u} + \beta v = 0. \quad (34.22)$$

Notice how the pressure gradient dropped out since there is zero baroclinicity for a barotropic flow (see Section 36.5 for a discussion of baroclinicity). We now make use of the identity

$$\frac{\partial}{\partial x} [\nabla \cdot (\mathbf{u} v)] - \frac{\partial}{\partial y} [\nabla \cdot (\mathbf{u} u)] = \frac{\partial \mathbf{u}}{\partial x} \cdot \nabla v - \frac{\partial \mathbf{u}}{\partial y} \cdot \nabla u + \mathbf{u} \cdot \nabla \zeta = \mathbf{u} \cdot \nabla \zeta, \quad (34.23)$$

where we used the non-divergence condition, $\partial_x u + \partial_y v = 0$, to write

$$\frac{\partial \mathbf{u}}{\partial x} \cdot \nabla v - \frac{\partial \mathbf{u}}{\partial y} \cdot \nabla u = \frac{\partial u}{\partial x} \frac{\partial v}{\partial x} + \frac{\partial v}{\partial x} \frac{\partial u}{\partial y} - \frac{\partial u}{\partial y} \frac{\partial u}{\partial x} - \frac{\partial v}{\partial y} \frac{\partial u}{\partial y} \quad (34.24a)$$

$$= -\frac{\partial v}{\partial y} \frac{\partial v}{\partial x} + \frac{\partial v}{\partial x} \frac{\partial v}{\partial y} + \frac{\partial u}{\partial y} \frac{\partial v}{\partial y} - \frac{\partial v}{\partial y} \frac{\partial u}{\partial y} \quad (34.24b)$$

$$= 0. \quad (34.24c)$$

We are thus led to the vorticity equation

$$\frac{\partial \zeta}{\partial t} + \nabla \cdot (\mathbf{u} \zeta) = -\beta v \quad \text{and} \quad \frac{D\zeta}{Dt} = -\beta v. \quad (34.25)$$

We see that the material evolution of relative vorticity in the non-divergent barotropic fluid is only affected by meridional advection of planetary vorticity. Since $\beta > 0$ over the globe, northward flow ($v > 0$) produces a negative source (clockwise tendency) for relative vorticity following a fluid parcel. This source term is the *beta effect* studied in Section 36.7.

34.3.2 Absolute vorticity is the potential vorticity

Since f is time independent, we can write the vorticity equation (34.25) in the form

$$\frac{\partial \zeta_a}{\partial t} + \mathbf{u} \cdot \nabla \zeta_a = 0, \quad (34.26)$$

where

$$\zeta_a = \zeta + f \quad (34.27)$$

is the vertical component of the absolute vorticity. Hence, there is no stretching or tilting of fluid columns in the non-divergent barotropic fluid. Again, we conceive of the flow as that of rigid fluid columns of constant thickness. Furthermore, the material conservation of ζ_a allows us to identify the potential vorticity with the absolute vorticity for the non-divergent barotropic fluid

$$q = \zeta_a = \zeta + f \implies \frac{\partial q}{\partial t} + \mathbf{u} \cdot \nabla q = 0 \implies \frac{Dq}{Dt} = 0. \quad (34.28)$$

To maintain a materially constant absolute vorticity requires the relative vorticity to change oppositely to that of the planetary vorticity. For example, in the northern hemisphere, the relative vorticity must decrease ($\zeta \downarrow$) for fluid particles moving northward. This change in the relative vorticity is needed to counteract the increasing planetary vorticity ($f \uparrow$) when moving northward. This result accords with the $-\beta v$ source found in the relative vorticity equation (34.25). Furthermore, it is a reflection of the beta effect studied in Section 36.7.2.

The presence of topography adds further richness to the flow. According to the kinematic boundary condition (34.8), horizontal flow is aligned with isobaths (lines of constant bathymetry or topography). Flow moving along constant isobaths generally crosses latitude lines, and in so doing the relative vorticity must change precisely to maintain $f + \zeta$ materially constant.

34.3.3 Jacobian form of vorticity advection

In some contexts it is convenient to write the advection operator acting on relative vorticity as

$$\mathbf{u} \cdot \nabla \zeta = u \partial_x \zeta + v \partial_y \zeta \quad (34.29a)$$

$$= -\partial_y \psi \partial_x \zeta + \partial_x \psi \partial_y \zeta \quad (34.29b)$$

$$= \hat{\mathbf{z}} \cdot (\nabla \psi \wedge \nabla \zeta) \quad (34.29c)$$

$$\equiv J(\psi, \zeta), \quad (34.29d)$$

where J is the Jacobian operator

$$J(A, B) = \hat{\mathbf{z}} \cdot (\nabla A \wedge \nabla B) = \frac{\partial A}{\partial x} \frac{\partial B}{\partial y} - \frac{\partial A}{\partial y} \frac{\partial B}{\partial x}. \quad (34.30)$$

We can also make use of the Jacobian for the advection of absolute vorticity

$$\mathbf{u} \cdot \nabla \zeta_a = \hat{\mathbf{z}} \cdot (\nabla \psi \wedge \nabla \zeta_a) \equiv J(\psi, \zeta + f). \quad (34.31)$$

The Jacobian operator corresponds to the Poisson bracket used in Hamiltonian mechanics.

34.3.4 Taylor-Bretherton identities

An equivalent means to write the vorticity equation is to start from the velocity equation in the form (34.19). Taking the curl and projecting onto the vertical direction then leads to

$$\frac{\partial \zeta}{\partial t} = -\beta v + \hat{z} \cdot [\nabla \wedge \nabla \cdot \mathcal{E}]. \quad (34.32)$$

The nonlinear forcing from the anisotropic kinetic stress can be written

$$\hat{z} \cdot [\nabla \wedge \nabla \cdot \mathcal{E}] = \hat{z}_n \epsilon_{npq} \partial_p (\partial_m \mathcal{E}_{mq}) \quad (34.33a)$$

$$= \partial_m \partial_p (\epsilon_{npq} \hat{z}_n \mathcal{E}_{mq}) \quad (34.33b)$$

$$= -\partial_m \partial_p (\epsilon_{pnq} \hat{z}_n \mathcal{E}_{qm}) \quad (34.33c)$$

$$= -\partial_m \partial_p (\hat{z} \wedge \mathcal{E})_{pm} \quad (34.33d)$$

$$= -\nabla \cdot [\nabla \cdot (\hat{z} \wedge \mathcal{E})], \quad (34.33e)$$

where we used symmetry of the anisotropic kinetic tensor, $\mathcal{E}_{mq} = \mathcal{E}_{qm}$.

Taylor-Bretherton identity for relative vorticity

Comparing to the vorticity equation in the form (34.25) reveals the identity

$$\nabla \cdot (\mathbf{u} \zeta) = \nabla \cdot [\nabla \cdot (\hat{z} \wedge \mathcal{E})] \implies \mathbf{u} \zeta = \nabla \cdot (\hat{z} \wedge \mathcal{E}). \quad (34.34)$$

This equation says that the advective vorticity flux equals to the divergence of the counter-clockwise rotated anisotropic kinetic stress tensor. Equation (34.34) is a special form of the *Taylor-Bretherton identity* that provides a connection between the vorticity flux and the momentum flux. We encounter the shallow water form of this identity Chapter 47 when studying the decomposition of eddy and mean flows.

Verifying the Taylor-Bretherton identity

The divergence expression on the left hand side of equation (34.34) can be satisfied by $\hat{z} \wedge \nabla \Upsilon + \nabla \cdot (\hat{z} \wedge \mathcal{E})$, with Υ an arbitrary gauge function. However, $\Upsilon = 0$ is zero for the anisotropic kinetic tensor (34.18), as seen by

$$[\nabla \cdot (\hat{z} \wedge \mathcal{E})]_1 = \partial_m (\epsilon_{mnp} \hat{z}_n \mathcal{E}_{p1}) \quad (34.35a)$$

$$= \epsilon_{m3p} \partial_m \mathcal{E}_{p1} \quad (34.35b)$$

$$= -\epsilon_{3mp} \partial_m \mathcal{E}_{p1} \quad (34.35c)$$

$$= -\epsilon_{312} \partial_1 \mathcal{E}_{21} - \epsilon_{321} \partial_2 \mathcal{E}_{11} \quad (34.35d)$$

$$= -\partial_x \mathcal{E}_{21} + \partial_y \mathcal{E}_{11} \quad (34.35e)$$

$$= \partial_x (u v) + \partial_y (-u^2 + v^2)/2 \quad (34.35f)$$

$$= v \partial_x u + u \partial_x v - u \partial_y u + v \partial_y v \quad (34.35g)$$

$$= u \zeta, \quad (34.35h)$$

and likewise

$$[\nabla \cdot (\hat{z} \wedge \mathcal{E})]_2 = \partial_y \mathcal{E}_{12} - \partial_x \mathcal{E}_{22} = v \zeta. \quad (34.36)$$

Taylor-Bretherton identity for potential vorticity

Building from the development for relative vorticity, we can readily connect the potential vorticity flux, $\mathbf{u} q$, to the anisotropic kinetic stress. We do so by considering the two equivalent forms for the potential vorticity equation

$$\frac{\partial q}{\partial t} = -\nabla \cdot (\mathbf{u} q) \quad \text{and} \quad \frac{\partial q}{\partial t} = -\nabla \cdot [\mathbf{u} f + \nabla \cdot (\hat{\mathbf{z}} \wedge \mathcal{E})]. \quad (34.37)$$

Hence, the Taylor-Bretherton identity for potential vorticity in the two dimensional non-divergent flow is given by

$$\mathbf{u} q = \mathbf{u} f + \nabla \cdot (\hat{\mathbf{z}} \wedge \mathcal{E}). \quad (34.38)$$

34.3.5 Poisson equation for the streamfunction

Given boundary conditions, the barotropic vorticity equation (34.26) allows us to determine the evolution of vorticity. We can in turn invert the Poisson equation (see Section 3.5 for discussion of the Poisson equation)

$$\nabla^2 \psi = \zeta \quad (34.39)$$

to determine the streamfunction and then the velocity field $\mathbf{u} = \hat{\mathbf{z}} \wedge \nabla \psi$. This inversion requires information about the boundary conditions for the streamfunction, as discussed in Section 34.2.2. By this method, time integration of the absolute vorticity equation is sufficient to fully specify time evolution of the horizontal velocity. Notably, we do not need to explicitly determine pressure to determine the flow.

34.3.6 Zonal flow as an exact geostrophic solution

An arbitrary zonal velocity with a meridional shear, $\mathbf{u} = U(y) \hat{\mathbf{x}}$, is an exact solution of the perfect fluid non-divergent barotropic model. We see this property by plugging into the velocity equation (34.1) and noting that $D\mathbf{u}/Dt = 0$. Hence, this flow is an exact geostrophic solution whose pressure field is itself also just a function of latitude and whose meridional gradient is determined by

$$f \hat{\mathbf{z}} \wedge \mathbf{u} = -\nabla \varphi \implies \partial_y \varphi = -f(y) U(y). \quad (34.40)$$

Furthermore, each term in the vorticity equation (34.25) identically vanishes, so that the vorticity

$$\zeta = -\partial_y U \quad (34.41)$$

remains constant in time at each point in space.

34.4 Connection to equivalent barotropic flow

In many cases, flows respecting the quasi-geostrophic assumptions possess a vertical profile that can be separated from the horizontal. In this case we write the horizontal velocity as

$$\mathbf{u}(x, y, z, t) = \Psi(z) \mathbf{u}^{eb}(x, y, t), \quad (34.42)$$

where $\Psi > 0$ is a single-signed non-dimensional structure function that has a unit average, $\langle \Psi \rangle = 1$, when computed over the fluid layer thickness, and where $\mathbf{u}^{eb}(x, y, t)$ carries the horizontal spatial dependence of the flow. With $\Psi > 0$, the horizontal flow remains in the same direction throughout the fluid column; i.e., eastward flow at the top of the column remains eastward at the bottom.

34.4.1 Vorticity equation for the depth averaged flow

To connect the very particular form (34.42) for the flow with the non-divergent barotropic model, we anticipate a discussion in Section 42.2.2 in which the quasi-geostrophic vorticity equation is shown to be

$$\frac{\partial \zeta_g}{\partial t} + \mathbf{u}_g \cdot \nabla \zeta_g = -\beta \zeta_g + f_0 \frac{\partial w}{\partial z}, \quad (34.43)$$

where \mathbf{u}_g is the horizontally non-divergent geostrophic velocity, and $\zeta_g = \hat{z} \cdot (\nabla \wedge \mathbf{u}_g)$ is the geostrophic relative vorticity. We see that the quasi-geostrophic vorticity is affected by both the beta-effect and by vertical stretching of fluid columns (the $\partial w / \partial z$ term), whereas vertical stretching is absent from the non-divergent barotropic vorticity equation (34.25). There are occasions in which it is sensible to assume the vertical velocity vanishes at the top and bottom of the fluid domain, such as when considering flow in the absence of topography. In this case, performing the decomposition (34.42) for the horizontal flow and then taking a vertical average of the quasi-geostrophic vorticity equation (34.43) leads to

$$\frac{\partial \zeta_g^{eb}}{\partial t} + \langle \Psi^2 \rangle \mathbf{u}_g^{eb} \cdot \nabla \zeta_g^{eb} = -\beta v_g^{eb}. \quad (34.44)$$

This equation motivates us to define

$$\mathbf{u}^* = \langle \Psi^2 \rangle \mathbf{u}_g^{eb} \quad \text{and} \quad \zeta^* = \langle \Psi^2 \rangle \zeta_g^{eb}, \quad (34.45)$$

which are the original geostrophic fields, $\mathbf{u}_g(x, y, z, t)$ and $\zeta_g(x, y, z, t)$, when evaluated at a depth where $\Psi(z^*) = \langle \Psi^2 \rangle$. The depth, z^* , is known as the *equivalent barotropic depth*. Introduction of the starred fields then brings the vorticity equation (34.44) into the form of the non-divergent barotropic vorticity equation

$$\frac{\partial \zeta^*}{\partial t} + \mathbf{u}^* \cdot \nabla \zeta^* = -\beta v^*. \quad (34.46)$$

34.4.2 Comments

Charney et al. (1950) justified their study of the non-divergent barotropic vorticity model by noting its connection to the commonly observed equivalent barotropic structure of the large-scale middle latitude atmosphere. The equivalent barotropic model has been a very useful analysis and prediction tool for meteorologists, and it formed the basis of many numerical weather prediction models into the 1980s. Section 7.1 of *Haltiner and Williams* (1980) offers further details on such numerical models.

One hypothesis for why quasi-geostrophic flows tend towards an equivalent barotropic profile relates to movement of energy in rotationally dominant turbulent flows, whereby energy cascades to the larger scales. As discussed in *Smith and Vallis* (2001) and *Venaille et al.* (2012), among others, this *inverse* energy cascade pumps mechanical energy into a vertically uniform or “barotropic” structure. In a realistic flow, including topography, stratification, and variable forcing, this cascade is never realized completely, thus rendering an approximate, or equivalent, barotropic flow. Conversely, moving to the highly idealized world, the case of $f = 0$ with flat bottom is commonly referred to as *two-dimensional turbulence* (*Kraichnan and Montgomery*, 1980). This model has a long history of key theoretical results that presaged their analog in quasi-geostrophic turbulence.

34.5 The pressure field

As noted in Section 34.3.5, we do not need to determine the pressure force to determine the evolution of the non-divergent barotropic flow. Instead, we can determine the flow by time stepping vorticity

and then inverting the elliptic problem to get the streamfunction. Furthermore, the free surface is absolutely flat even in the presence of topography. But in the shallow water model if the free surface is flat then there is no horizontal pressure gradient. So how does pressure act in the non-divergent barotropic model? In this model, the flow is constrained as if there was a flat *rigid lid* placed on the surface. To maintain the lid flat requires a corresponding *lid pressure*. It is the lid pressure that drives the flow.

The velocity field in a non-divergent barotropic flow is highly constrained, as seen by recalling that at each space and time point within the flow satisfies

$$\nabla \cdot \mathbf{u} = 0 \quad \text{and} \quad \mathbf{u} \cdot \nabla \eta_b = 0 \quad \text{and} \quad w = 0. \quad (34.47)$$

As we show in this section, the depth independent pressure force maintains these constraints. That is, the kinematics of the constrained flow induce a pressure that enforces the constraints. Examining that pressure field furthers our understanding of the dynamical forces acting in the fluid. This role for pressure as an enforcer of non-divergence is shared by the three-dimensional non-divergent flow found in a Boussinesq fluid (see Section 24.3).

34.5.1 Poisson equation for pressure

We derive the pressure equation by using the two-dimensional non-divergence property of the horizontal flow and then developing a diagnostic relation. We can eliminate the time derivative from equation (34.1) by taking $\partial/\partial x$ on the zonal equation and $\partial/\partial y$ on the meridional equation, then adding. The result is a diagnostic relation for the Laplacian of the pressure²

$$-\nabla^2 \varphi = \partial_x[\nabla \cdot (\mathbf{u} u)] + \partial_y[\nabla \cdot (\mathbf{u} v)] - f \zeta + \beta u, \quad (34.48)$$

where we set

$$\nabla \cdot \partial_t \mathbf{u} = \partial_t \nabla \cdot \mathbf{u} = 0 \implies \partial_x(\partial_t u) = -\partial_y(\partial_t v). \quad (34.49)$$

Making use of the boundary conditions discussed in Section 34.2.2, the elliptic partial differential equation (34.48) can be inverted to find the pressure field (Section 3.5). As for the three-dimensional incompressible case, we see that the pressure is instantaneously constrained to satisfy the non-divergence condition placed on the velocity.

Numerically inverting an elliptic operator in equation (34.48) is straightforward on simple domains, such as flat bottom rectangular regions or a smooth sphere. However, when the bottom is not flat, or when there are islands (i.e., the domain is not simply connected), then the elliptic inversion can be difficult to perform so that it is difficult to reach the optimal solution. This algorithmic complexity is one reason numerical barotropic models are less commonly used for realistic numerical experimentation than the more general shallow water models.

34.5.2 Self-advection source in the pressure equation

To help understand the pressure forces acting in the fluid, we examine the nature of the self-advection source appearing in the pressure equation (34.48), in which we write

$$\partial_x[\nabla \cdot (\mathbf{u} u)] + \partial_y[\nabla \cdot (\mathbf{u} v)] = \mathbb{S}_{mn} \mathbb{S}_{mn} - \mathbb{A}_{mn} \mathbb{A}_{mn}, \quad (34.50)$$

²We maintain the minus sign on the left hand side of equation (34.48) so that a positive source on the right hand side leads to a positive φ . We can readily understand the sign by taking a Fourier transform, whereby the Laplacian operator picks up a minus sign when converted to Fourier space. We follow the same sign convention when studying the Green's function for the Poisson equation in Chapter 4.

where the rate of strain tensor, \mathbb{S} , and rotation tensor, \mathbb{A} , have components given by equations (14.33a) and (14.33b)

$$\mathbb{S}_{mn} = \frac{1}{2} \left[\frac{\partial v_m}{\partial x_n} + \frac{\partial v_n}{\partial x_m} \right] = \mathbb{S}_{nm} \quad \text{rate of strain tensor} \quad (34.51)$$

$$\mathbb{A}_{mn} = \frac{1}{2} \left[\frac{\partial v_m}{\partial x_n} - \frac{\partial v_n}{\partial x_m} \right] = -\mathbb{A}_{nm} \quad \text{rotation tensor.} \quad (34.52)$$

For two-dimensional flow the rotation tensor is related to the relative vorticity via

$$\mathbb{A}_{mn} = -\epsilon_{mn} \zeta / 2 = -\epsilon_{mn} \nabla^2 \psi / 2, \quad (34.53)$$

where ϵ_{mn} is the anti-symmetric Levi-Civita permutation symbol (see Section 1.4.1)

$$\epsilon_{mn} = \begin{bmatrix} \epsilon_{11} & \epsilon_{12} \\ \epsilon_{21} & \epsilon_{22} \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}, \quad (34.54)$$

which leads to

$$\mathbb{A}_{mn} \mathbb{A}_{mn} = \zeta^2 / 2. \quad (34.55)$$

The identity (34.50) indicates that the squared strain provides a positive source to the Poisson equation (34.48) whereas squared vorticity provides a negative source. We support this result by considering two idealized flows in Section 34.5.4.

34.5.3 Gradient wind balance

For two-dimensional non-divergent flow we can write the self-advection in terms of the streamfunction, or equivalently to the Jacobian of the velocity field, through the following manipulations

$$\mathbb{S}_{mn} \mathbb{S}_{mn} - \mathbb{A}_{mn} \mathbb{A}_{mn} = \partial_x [\nabla \cdot (\mathbf{u} u)] + \partial_y [\nabla \cdot (\mathbf{u} v)] \quad (34.56a)$$

$$= \nabla \cdot (\partial_x \mathbf{u} u) + \nabla \cdot (\mathbf{u} \partial_x u) + \nabla \cdot (\partial_y \mathbf{u} v) + \nabla \cdot (\mathbf{u} \partial_y v) \quad (34.56b)$$

$$= \partial_x \mathbf{u} \cdot \nabla u + \partial_y \mathbf{u} \cdot \nabla v \quad (34.56c)$$

$$= (\partial_x u)^2 + (\partial_y v)^2 + 2 \partial_x v \partial_y u \quad (34.56d)$$

$$= 2 [(\partial_x u)^2 + \partial_x v \partial_y u] \quad (34.56e)$$

$$= 2 [(\partial_{xy} \psi)^2 - \partial_{xx} \psi \partial_{yy} \psi] \quad (34.56f)$$

$$= 2 \hat{\mathbf{z}} \cdot (\partial_x \nabla \psi \wedge \partial_y \nabla \psi) \quad (34.56g)$$

$$= -2 J(\partial_x \psi, \partial_y \psi), \quad (34.56h)$$

$$= 2 J(v, u), \quad (34.56i)$$

where we introduced the Jacobian operator from equation (34.25). The pressure equation (34.48) thus takes on the form

$$-\nabla^2 \varphi = -2 J(u, v) - \nabla \cdot (f \nabla \psi). \quad (34.57)$$

We consider this equation as a *gradient wind balance* in analog to the gradient wind balance discussed in Section 27.6. Here, the Jacobian term accounts for the centrifugal acceleration of the curved fluid motion, and the $\nabla \cdot (f \nabla \psi)$ term accounts for the Coriolis acceleration, both for the f -plane and β -plane. Equation (34.57) thus offers a more accurate diagnostic relation for the pressure field than provided by assuming a geostrophically balanced flow.

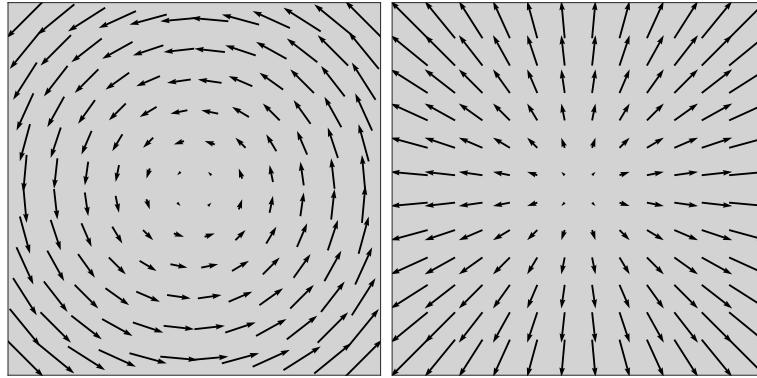


FIGURE 34.1: Left panel: example flow field from solid-body rotation, $\mathbf{u} = \boldsymbol{\Omega} \wedge \mathbf{x} = (-y, x)$, which has vorticity $\nabla \wedge \mathbf{u} = 2\boldsymbol{\Omega}$ and zero strain. Right panel: corresponding acceleration from self-advection, $-(\mathbf{u} \cdot \nabla) \mathbf{u} = \boldsymbol{\Omega}^2 \mathbf{x}$. The acceleration from self-advection is exactly compensated by the pressure gradient force: $-(\mathbf{u} \cdot \nabla) \mathbf{u} - \nabla \varphi = 0$, thus allowing for the solid-body motion to be an exact solution to two-dimensional non-divergent flow. The units are arbitrary.

34.5.4 Sample pressure gradients

We present two examples to help support understanding of how pressure gradients arise in a 2d-barotropic flow so as to maintain, at each time instance, a non-divergent two-dimensional velocity field. For simplicity we just consider the non-rotating case so that it is sufficient to examine how pressure responds to self-advection.

Solid-body flow

Consider a velocity field in a flat domain that is initialized in circular solid-body motion (Figure 34.1)

$$\mathbf{u} = \boldsymbol{\Omega} \wedge \mathbf{x} = \Omega(-y \hat{\mathbf{x}} + x \hat{\mathbf{y}}), \quad (34.58)$$

where $\boldsymbol{\Omega} = \Omega \hat{\mathbf{z}}$ is a constant angular velocity. This flow has zero strain and non-zero vorticity

$$\hat{\mathbf{z}} \cdot (\nabla \wedge \mathbf{u}) = \zeta = 2\Omega \implies -\mathbb{A}_{mn} \mathbb{A}_{mn} = -\Omega^2. \quad (34.59)$$

The velocity time tendency from the self-advection acceleration is

$$-(\mathbf{u} \cdot \nabla) \mathbf{u} = \Omega^2 \mathbf{x}, \quad (34.60)$$

which is the outward directed centrifugal acceleration associated with the circular solid-body motion.³ The centrifugal acceleration generates a velocity field that diverges from the origin. However, the velocity is constrained to remain non-divergent at each instance, so the centrifugal acceleration cannot be the full story. Indeed, to maintain the non-divergent constraint, a pressure gradient is instantaneously established that exactly counteracts the centrifugal acceleration. Hence, there is a low pressure established at the origin so that the pressure gradient force points inward. This example illustrates how vorticity provides a source of low pressure in an incompressible flow. Furthermore, since the pressure gradient exactly counteracts the centrifugal acceleration from the velocity self-advection, the solid-body flow is an exact steady solution for the non-divergent and non-rotating barotropic system.

³We considered the more general case in Section 27.2.2 when decomposing the material acceleration for two-dimensional flow into natural coordinates.

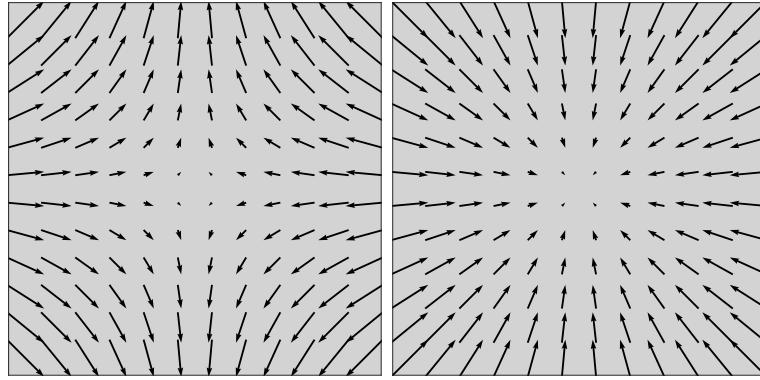


FIGURE 34.2: Left panel: example purely strained flow field with zero vorticity, $\mathbf{u} = \Omega(-x\hat{\mathbf{x}} + y\hat{\mathbf{y}})$. Right panel: corresponding converging acceleration from self-advection, $-(\mathbf{u} \cdot \nabla) \mathbf{u} = -\Omega^2 \mathbf{x}$. The units are arbitrary.

The pressure field is found by solving the Poisson equation

$$-\nabla^2 \varphi = -\Omega^2. \quad (34.61)$$

Assuming circular symmetry allows us to let φ be a function just of the radial distance from the origin, in which case the pressure field is parabolic and the pressure gradient acceleration is radial

$$\varphi = (\Omega r)^2/2 \quad \text{and} \quad -\nabla \varphi = -\Omega^2 \mathbf{x} = (\mathbf{u} \cdot \nabla) \mathbf{u}. \quad (34.62)$$

To reach this result, we used the cylindrical expression for the Laplacian operator (equation (8.83b)) and set $\varphi(r = 0) = 0$. To help understand the physics of this parabolic pressure field, recall the analysis in Section 23.4 of a solid-body rotating homogeneous fluid layer in a cylindrical tank. In contrast to the barotropic system, the horizontal velocity in a homogeneous fluid layer, such as in a shallow water fluid, is divergent so that the layer thickness is not constrained to remain flat. Hence, the centrifugal acceleration causes the velocity to diverge from the center so that the layer thickness increases radially outward, with the layer bounded by the tank wall. At steady state, the homogeneous fluid layer has a parabolic free surface with a minimum at the center (see equation (23.80)). The parabolic free surface creates a pressure field that precisely corresponds to the pressure field (34.62) in the non-divergent barotropic system. The adjustment of a homogeneous fluid layer, such as a shallow water fluid, contains linear fluctuations in the form of gravity waves such as discussed in Chapter 32. In contrast, the adjustment required to reach a steady state occurs instantaneously in the non-divergent barotropic fluid.

Irrational flow

Now consider the following pure strain flow (Figure 34.2)

$$\mathbf{u} = \Omega(-x\hat{\mathbf{x}} + y\hat{\mathbf{y}}), \quad (34.63)$$

whose vorticity vanishes and whose self-advection acceleration is given by

$$-(\mathbf{u} \cdot \nabla) \mathbf{u} = -\Omega^2 \mathbf{x} \implies \mathbb{S}_{mn} \mathbb{S}_{mn} = \Omega^2. \quad (34.64)$$

This acceleration is exactly the opposite of that produced by the solid-body motion studied in the previous example. Hence, to maintain the non-divergent nature of the barotropic flow, a pressure

gradient is established with a high pressure at the center that exactly counteracts the converging self-advection acceleration arising in the strained flow

$$-\nabla\varphi = \Omega^2 \mathbf{x} \implies -(\mathbf{u} \cdot \nabla) \mathbf{u} - \nabla\varphi = 0. \quad (34.65)$$

This example illustrates how strain provides a source for high pressure in an incompressible flow. Furthermore, we see that this flow is an exact steady solution for non-rotating non-divergent barotropic flow.

34.5.5 Pressure source from Coriolis acceleration

In addition to the self-advection source, pressure is affected by a source from the Coriolis acceleration

$$-\nabla^2\varphi_{\text{geostrophy}} \equiv \nabla \cdot (f \hat{\mathbf{z}} \wedge \mathbf{u}) = \beta u - \zeta f. \quad (34.66)$$

As such, we can write the pressure gradient as⁴

$$-\nabla\varphi_{\text{geostrophy}} = f \hat{\mathbf{z}} \wedge \mathbf{u}. \quad (34.67)$$

Consider cyclonic flow around an arbitrary point. What is the pressure field induced by this flow? As in our discussion in Section 34.5.4 of the solid body rotating flow, a cyclonic flow has an associated centrifugal acceleration that points outward. To counteract the centrifugal acceleration, and thus to maintain a non-divergent flow, requires an inward pointing pressure force; i.e., a low pressure center. Hence, cyclonic circulation induces, through the Coriolis acceleration, a negative pressure source.

34.5.6 Pressure source from friction

Consider flow with friction, in which case the velocity equation (34.1) takes the form

$$\frac{D\mathbf{u}}{Dt} + f \hat{\mathbf{z}} \wedge \mathbf{u} = -\nabla\varphi + \mathbf{F}, \quad (34.68)$$

with \mathbf{F} a frictional acceleration. In this case we have yet another source for pressure given by

$$-\nabla^2\varphi_{\text{friction}} \equiv -\nabla \cdot \mathbf{F} \implies -\nabla\varphi_{\text{friction}} = -\mathbf{F}. \quad (34.69)$$

As discussed in Section 19.7, viscous friction is generally associated with a nonzero rate of strain. We see that the frictional acceleration induces a high pressure source in regions where frictional accelerations converge, $-\nabla \cdot \mathbf{F} > 0$, with this source acting to maintain non-divergent flow in the presence of converging frictional acceleration.

34.5.7 Comments and further study

- The pressure equation (34.48) is elliptic, and elliptic equations need boundary conditions. In the presence of topography the boundary conditions are modified relative to the flat bottom case. Hence, pressure knows about topography through its boundary conditions. The resulting pressure force keeps the flow non-divergent and the flow aligned with topography as per the kinematic conditions in Section 34.2.2.

⁴Formally, we can add an arbitrary gauge function to the right hand side of equation (34.67), with this term of the form $\hat{\mathbf{z}} \wedge \nabla\chi$. However, since we derived the pressure Poisson equation from the velocity equation, then we know there is no gauge function arising in equation (34.67).

- The discussion of pressure induced by self-advection in Section 34.5.4 is based on a similar presentation in Appendix B of *Jeevanjee and Romps* (2015a).
- *Bryan* (1969) provided the first working numerical algorithm to simulate the ocean general circulation. Bryan's method made use of the rigid lid approximation of Section 34.5 so that the depth integrated velocity is non-divergent. However, the vorticity in Bryan's ocean model is affected by more than just the beta-effect. The reason is that the depth integrated velocity equation includes contributions from baroclinic processes, and such processes affect the barotropic vorticity in a baroclinic fluid. We detail such effects in Sections 41.5 and 41.6.

The rigid lid method was generally used for large-scale circulation modeling until the late 1990s. Free surface methods, allowing divergence in the depth integrated flow, have largely displaced the rigid lid as a practical method for time stepping ocean models (e.g., see chapter 12 of *Griffies* (2004)).

34.6 Constraints from absolute vorticity invariance

We here examine constraints on the flow imposed by material invariance of absolute vorticity, which holds for two-dimensional non-divergent flow when there is no dissipation

$$\frac{D(\zeta + f)}{Dt} = 0. \quad (34.70)$$

These constraints offer insights into the flow behavior and allow us to predict responses to perturbations. We frame the discussion in terms of the decomposition (33.38) of relative vorticity into a curvature (or orbital) term and normal shear term

$$\zeta = \frac{|\mathbf{u}|}{R} - \frac{\partial |\mathbf{u}|}{\partial n} \equiv \zeta_{\text{curv}} + \zeta_{\text{shear}}, \quad (34.71)$$

so that

$$\frac{D(\zeta_{\text{curv}} + \zeta_{\text{shear}} + f)}{Dt} = 0. \quad (34.72)$$

When the flow turns, the curvature term contributes with R the radius of curvature.⁵ If the flow develops shears in the direction normal to the flow, then the shear term contributes. In the presence of $\beta = \partial_y f$, meridional motion through the planetary vorticity field requires a compensating response from relative vorticity. In general each of the three terms contribute to the relative vorticity, though in the following we consider cases where one is subdominant.

34.6.1 Relative vorticity from curvature and planetary beta

According to the decomposition (34.71), fluid that curves to the left (facing downstream) picks up a positive relative vorticity from flow curvature, $R > 0 \Rightarrow \zeta > 0$, as depicted in Figure 34.3. The oppositely curved flow has a negative radius of curvature so flow curving to the right picks up a negative relative vorticity, $R < 0 \Rightarrow \zeta < 0$. We focus here on the case where the normal shear induced relative vorticity can be neglected so that we are only concerned with curvature induced vorticity plus planetary vorticity (beta effect).

⁵ As discussed in Section 27.2, we take the convention whereby the normal direction is to the left of flow when facing downstream. Flow turning into the normal direction (to the left) has a positive radius of curvature, $R > 0$, and flow turning opposite to the normal direction (to the right) has $R < 0$.

Consider a flow that is initially zonal with zero relative vorticity. If the flow turns meridionally then it experiences a change in relative vorticity both through the curvature term plus a change in planetary vorticity since f changes. To maintain constant absolute vorticity, a fluid column that moves meridionally requires the relative vorticity induced by the curved flow path to counteract the change in planetary vorticity. As we now discuss, the constraint of fixed absolute vorticity, in the absence of induced normal shears, means that eastward flow (westerly winds) cannot turn meridionally while maintaining fixed absolute vorticity, whereas westward flow (easterly winds) can turn (see Figure 34.3).

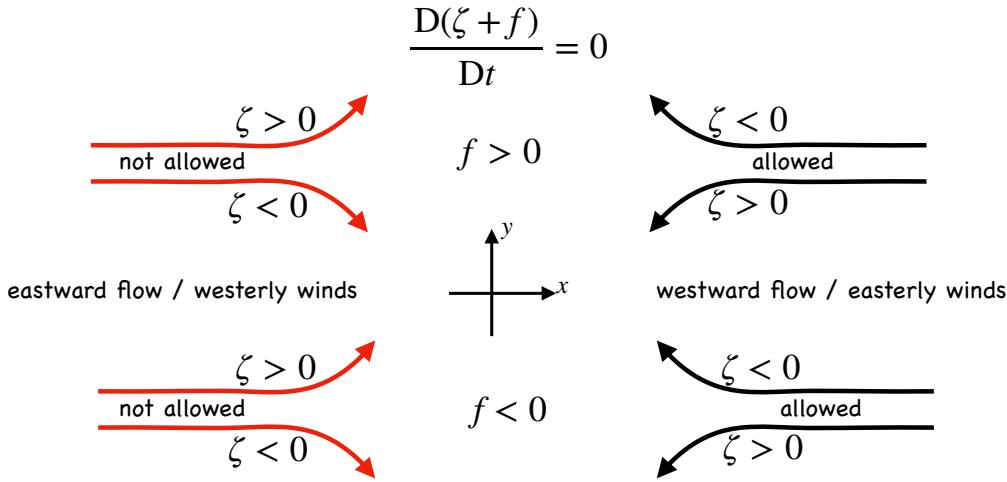


FIGURE 34.3: Illustrating the constraints on a two dimensional non-divergent flow imposed by material invariance of absolute vorticity: $\zeta + f = \text{constant}$. We assume flow is over a flat region and assume there is only curvature induced relative vorticity (no shear-induced relative vorticity; $\partial|\mathbf{u}|/\partial n = 0$) as per the decomposition in equation (34.71). In each of the four cases depicted, the entering flow has zero relative vorticity, which means that absolute vorticity must remain constant at the initial Coriolis parameter, $\zeta + f = f_{\text{initial}}$. The red eastward flow (westerly winds) that turns meridionally picks up a curvature vorticity that supports the change in planetary vorticity, thus precluding material invariance of absolute vorticity. Hence, the meridional turning of eastward flow is not allowed so that flow must remain zonal for absolute vorticity to remain invariant. In contrast, the oppositely directed westward flow (easterly winds) can deviate either to the north or south and still retain a constant absolute vorticity. We illustrate flows for both the northern and southern hemispheres. This figure is adapted from Figure 4.8 of Holton (1992).

Consider westward flow in the northern hemisphere ($f > 0$). If the flow turns to the north (to the right facing downstream) then this flow picks up a curvature-induced negative relative vorticity, $\zeta < 0$, and an increase in the planetary vorticity (f increases). Likewise, a westward flowing fluid column that turns equatorward (to the left) has a positive curvature-induced relative vorticity ($\zeta > 0$) and a reduction in planetary vorticity (f decreases). Hence, westward flow in the northern hemisphere can turn either poleward (to the north) or equatorward (to the south) and still maintain constant absolute vorticity, so long as the curved motion induces the proper relative vorticity to counteract the changes to f . The same arguments also hold in the southern hemisphere, so that the general scenarios are depicted in Figure 34.3.

The situation is different for eastward flow. Consider again flow in the northern hemisphere. A poleward (to the left) turning fluid column is associated with a positive curvature-induced relative vorticity, $\zeta > 0$, as well as an increase in the planetary vorticity. Hence, this motion changes the absolute vorticity and as such it is not allowed if the absolute vorticity is constrained to remain constant. Likewise, an equatorward (to the right) turning eastward fluid column induces a negative curvature-induced relative vorticity, $\zeta < 0$, and a decrease in planetary vorticity, again leading to

a change in absolute vorticity. Hence, eastward flow (westerly winds) in either hemisphere must remain zonal to maintain a constant absolute vorticity.

As an application of these results, consider the situation depicted in Figure 34.4, whereby inviscid flow in the interior of an ocean domain moves westward into a frictional western boundary layer. The constraints imposed by absolute vorticity invariance allow for this flow to occur, whereas the opposite is disallowed whereby eastward inviscid flow cannot enter an eastern boundary. We return to this example in Section 35.9 when discussing western intensification of ocean gyres.

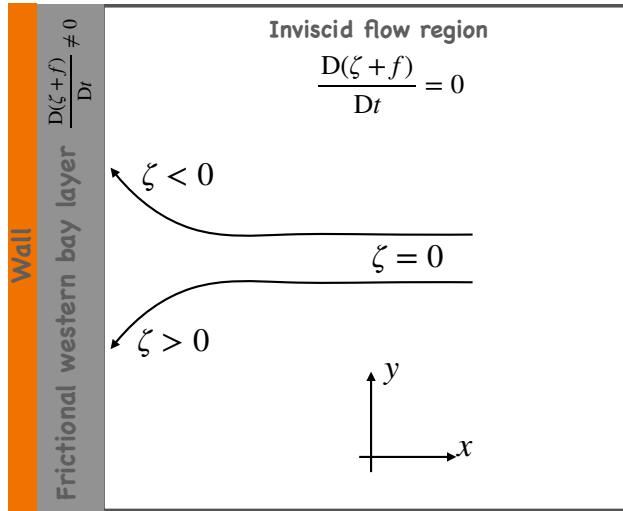


FIGURE 34.4: Illustrating the constraints on a homogeneous and constant thickness fluid layer imposed by material invariance of absolute vorticity: $\zeta + f = \text{constant}$. As per the results from Figure 34.3, inviscid flow with initially zero relative vorticity can enter a western boundary layer as depicted here, whereas it cannot enter an eastern boundary layer. This conclusion assumes that there is no shear-induced relative vorticity ($\partial|\mathbf{u}|/\partial n = 0$) that can overcome changes in the vorticity induced by changes to f and by curvature-induced relative vorticity (see Section 33.6.2). This figure is adapted from Figure 19.12 of [Vallis \(2017\)](#).

34.6.2 Relative vorticity from curvature and normal shears

Now consider the case in which the meridional displacements are small so that the beta effect can be neglected. In this case there is a material balance between relative vorticity arising from curvature and relative vorticity from normal shears

$$\frac{D(\zeta_{\text{curv}} + \zeta_{\text{shear}})}{Dt} = 0. \quad (34.73)$$

Hence, there is a direct exchange of relative vorticity between the contributions from normal shears and curvature.

We depict an example in Figure 34.5 whereby a vortex undergoes a left turn facing downstream. While on the curve, the relative vorticity of the vortex is in part due to the positive curvature vorticity, $\zeta_{\text{curv}} > 0$. If the relative vorticity is positive on the straight portion of the curve, then when on the curve the shear vorticity must lose some of its strength in order to compensate for the curvature vorticity. Conversely, if the relative vorticity is negative on the straight portion of the curve, then when on the curve the shear vorticity gains in strength to allow for the positive curvature vorticity.

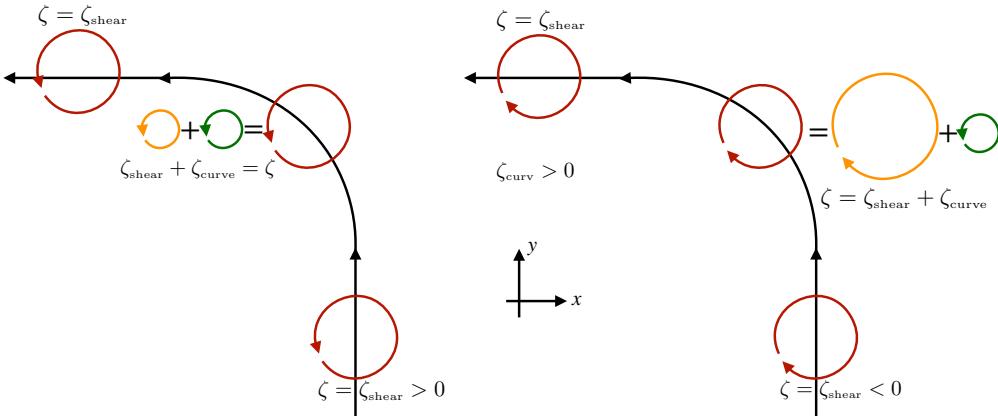


FIGURE 34.5: Material invariance of relative vorticity, $D\zeta/Dt = D(\zeta_{curv} + \zeta_{shear})/Dt = 0$, means that as a vortex moves around a curve its shear vorticity is modified to keep the total relative vorticity invariant. On the straight portion of these trajectories, the relative vorticity is due only to shear vorticity, $\zeta = \zeta_{shear}$ since $\zeta_{curv} = 0$. However, when the vortex enters the curve, maintaining a constant relative vorticity requires an exchange of shear vorticity with the curvature vorticity. In this example we illustrate a steady flow that turns to the left so that the vortex picks up a positive curvature vorticity, $\zeta_{curv} > 0$. Left panel: a vortex that enters the left turn with a positive relative vorticity must give some of its shear vorticity to the curvature vorticity in order to maintain ζ constant along the trajectory. Right panel: a vortex that enters the curve with a negative relative vorticity sees its shear vorticity increase in magnitude to compensate for the positive curvature vorticity.

34.6.3 Curvature, shear, and planetary contributions

We now consider all three terms appearing in the vorticity equation (34.72). Let us consider again the eastward flow that turns to the north in the northern hemisphere. Such flow is not allowed if the only source for relative vorticity is curvature. However, if the eastward flow, as it turns, picks up a shear that induces a nonzero negative relative vorticity, then such flow can turn so long as the shear-induced negative relative vorticity balances the positive absolute vorticity from increases in f and the curvature-induced vorticity. Writing this condition for the shear-induced relative vorticity yields

$$\zeta_{shear} = -\Delta f - \zeta_{curv} = -(f_{final} - f_{init}) - \zeta_{curv} < 0, \quad (34.74)$$

where f_{init} and f_{final} are the initial and final Coriolis parameters. Conversely, if the flow deviates towards the equator then it can do so only if there is a positive shear-induced relative vorticity

$$\zeta_{shear} = -(f_{final} - f_{init}) - \zeta_{curv} > 0. \quad (34.75)$$

34.7 Barotropic Rossby waves

The non-divergent barotropic model has no gravity waves since gravity waves are supported by a nonzero horizontal flow convergence (see discussion of shallow water gravity waves in Section 32.8). We can formally consider the gravity wave speed to be infinite.

However, the non-divergent barotropic model does admit linear wave fluctuations known as *Rossby waves*, and Rossby waves arise from the beta-effect (Section 36.7.2). In this section we briefly illustrate the basic physical attributes of barotropic Rossby waves. We do so by studying linear wave perturbations to uniform mean zonal flow in a zonally re-entrant channel on a beta-plane. This section is meant to be merely a teaser for a more thorough study of Rossby waves such as that provided in Chapter 6 of [Vallis \(2017\)](#). We assume some familiarity with wave kinematics as introduced for shallow water gravity waves in Section 32.8.

34.7.1 Dispersion relation

Consider a zonal channel with meridional extent $y_s \leq y \leq y_n$ where $y_n - y_s = L$, and allow for a basic state geostrophic flow to exist with a constant zonal velocity $U \hat{x}$ and corresponding streamfunction

$$\psi_{\text{basic}} = -U y. \quad (34.76)$$

Now perturb this flow with a wave pattern that leads to a streamfunction of the form

$$\psi = \psi_{\text{basic}} + \psi_o \sin(k x - \omega t) \sin[(y - y_s) l], \quad (34.77)$$

where ψ_o is the wave amplitude, k is the zonal wave number, and c is the wave phase speed. The meridional structure is chosen to satisfy the no-normal flow boundary condition at the northern and southern walls (Section 34.2.2), which leads to $\partial_x \psi = 0$ at $y = y_s$ and $y = y_n$. The southern boundary condition is trivially satisfied, yet the northern boundary is satisfied only if the meridional wavenumber, l , is quantized according to

$$(y_n - y_s) l = L l = 2\pi n \implies l = 2\pi/L \quad \text{with } n \text{ an integer.} \quad (34.78)$$

The corresponding velocity and relative vorticity are given by

$$u = -\partial_y \psi = U - \psi_o k \sin(k x - \omega t) \cos[(y - y_s) l] \quad (34.79a)$$

$$v = \partial_x \psi = \psi_o k \cos(k x - \omega t) \sin[(y - y_s) l] \quad (34.79b)$$

$$\zeta = \nabla^2 \psi = -\psi_o (k^2 + l^2) \sin(k x - \omega t) \sin[(y - y_s) l], \quad (34.79c)$$

so that terms in the barotropic vorticity equation (34.25) are

$$\partial_t \zeta = \psi_o \omega (k^2 + l^2) \cos(k x - \omega t) \sin[(y - y_s) l] \quad (34.80a)$$

$$\partial_x \zeta = -\psi_o k (k^2 + l^2) \cos(k x - \omega t) \sin[(y - y_s) l] \quad (34.80b)$$

$$\partial_y \zeta = -\psi_o l (k^2 + l^2) \sin(k x - \omega t) \cos[(y - y_s) l] \quad (34.80c)$$

$$\mathbf{u} \cdot \nabla \zeta = -\psi_o k U (k^2 + l^2) \cos(k x - \omega t) \sin[(y - y_s) l]. \quad (34.80d)$$

Plugging into the barotropic vorticity equation leads, without approximation, to the dispersion relation

$$\omega (k^2 + l^2) = k (k^2 + l^2) U - \beta k \implies \omega = U k - \frac{\beta k}{k^2 + l^2}. \quad (34.81)$$

The presence of a constant zonal velocity acts to *Doppler shift* the wave frequency by the amount $U k$. Furthermore, it is the presence of $\beta \neq 0$ that gives rise to these *Rossby waves*, with $\beta = 0$ rendering a frequency just due to the zonal background flow.

34.7.2 Phase speed

Components to the wave phase speed are

$$c_p^x = \frac{\omega}{k} = U - \frac{\beta}{k^2 + l^2} \quad \text{and} \quad c_p^y = \frac{\omega}{l} = \frac{U k}{l} - \frac{\beta k}{l (k^2 + l^2)}. \quad (34.82)$$

It is notable that the zonal phase speed is westward relative to the background zonal velocity. To get a sense for the scales of these waves, consider zero mean flow and set the zonal phase speed

equal to the shallow water gravity wave phase speed, $c_{\text{gravity}} = (g H)^{1/2}$ (Section 32.8), in which case the Rossby wave has a wavenumber

$$(k^2 + l^2)^{1/2} = \beta / \sqrt{g H}. \quad (34.83)$$

Writing $K = \sqrt{k^2 + l^2} = 2\pi/\lambda$ leads to a wavelength of $\lambda \approx 2 \times 10^4$ km for $\phi = 60^\circ\text{S}$ and $\beta = (2\Omega/R) \cos \phi \approx 1.14 \times 10^{-11} \text{ m}^{-1} \text{ s}^{-1}$.

Stationary Rossby waves are those whose zonal phase speed vanishes due to the Doppler shift from the zonal flow

$$c_p^x = 0 \implies U = \frac{\beta}{k^2 + l^2}, \quad (34.84)$$

in which case the zonal flow speed exactly matches the phase speed of the Rossby wave. Writing $K = \sqrt{k^2 + l^2} = 2\pi/L$ leads to a length scale for the stationary Rossby waves

$$L_{\text{stationary}} = 2\pi (U/\beta)^{1/2}. \quad (34.85)$$

With a eastward speed of $U = 1 \text{ m s}^{-1}$ at $\phi = 60^\circ\text{S}$ we find stationary ocean barotropic Rossby waves have a wavelength of $\lambda_{\text{stationary}} \approx 1800$ km, whereas for the atmosphere with $U = 25 \text{ m s}^{-1}$ we find $\lambda_{\text{stationary}} \approx 9300$ km.

34.7.3 Group velocity

Components to the phase speed are not components to a vector, whereas the group velocity is a vector and it measures the velocity that energy moves with the waves

$$\mathbf{c}_g = \hat{\mathbf{x}} \partial_k \omega + \hat{\mathbf{y}} \partial_l \omega \quad (34.86a)$$

$$= \hat{\mathbf{x}} \left[U + \frac{\beta (k^2 - l^2)}{(k^2 + l^2)^2} \right] + \hat{\mathbf{y}} \frac{2\beta k l}{(k^2 + l^2)^2} \quad (34.86b)$$

$$= \hat{\mathbf{x}} \left[c_p^x + \frac{2\beta k^2}{(k^2 + l^2)^2} \right] + \hat{\mathbf{y}} \frac{2\beta k l}{(k^2 + l^2)^2}. \quad (34.86c)$$

Note how the zonal component to the group velocity is always eastward relative to the phase speed. Furthermore, the zonal group velocity is westward if the zonal wavenumber is small enough (long zonal Rossby waves)

$$\frac{k^2 - l^2}{(k^2 + l^2)^2} < -U/\beta \iff \text{long zonal Rossby waves have westward group velocity}, \quad (34.87)$$

whereas it is eastward for short zonal Rossby waves

$$\frac{k^2 - l^2}{(k^2 + l^2)^2} > -U/\beta \iff \text{short zonal Rossby waves have eastward group velocity}. \quad (34.88)$$

34.7.4 Importance of β for Rossby waves

Gravity is the restoring force giving rise to gravity waves (Chapter 32). For Rossby waves, it is the planetary β that gives rise to their existence. We see such to be the case by noting that with $\beta = 0$ there is no wave disturbance as measured by the dispersion relation (34.81). In Figure 34.6 we illustrate the mechanism for the westward phase propagation of the Rossby waves (with $U = 0$ for simplicity). As detailed in the figure caption, the constraint of absolute momentum conservation

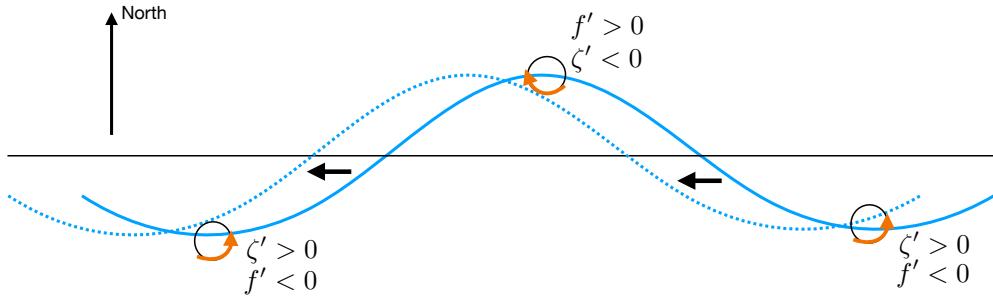


FIGURE 34.6: The westward phase speed of a Rossby wave arises from the presence of $\beta > 0$ and the constraint that potential vorticity ($q = f + \zeta$ for a non-divergent barotropic model) is materially constant in a perfect fluid. In this figure we consider a material line within the fluid in either the northern or southern hemispheres, with the solid line the initial position of the fluctuation and the dotted line the position later in time. For a northward perturbation relative to the initial latitude, a fluid parcel moves to a latitude with Coriolis parameter larger than its original value. Absolute vorticity conservation requires the parcel to pick up a negative relative vorticity perturbation, $\zeta' < 0$. The opposite occurs for a southward perturbation. The flow induced by the relative vorticity anomaly acts to move the fluctuation westward. In the absence of β , the meridional movement of parcels does not render a change in the planetary vorticity since f is a constant. In this case there is no induced relative vorticity anomaly so there is no coherent movement of the perturbation. We thus see the central role of $\beta \neq 0$ for Rossby waves. We also see that the sign of the Coriolis parameter is not relevant; it is only $\beta > 0$ that determines the motion.

for fluid parcels, in the presence of $\beta \neq 0$, gives rise to the westward phase speed for the Rossby wave.

The argument offered in Figure 34.6 does not consider any forces. Rather, we make use of the constraint imposed by material conservation of absolute vorticity and infer the motion of fluid parcels by noting how the relative vorticity anomaly induces a flow that moves surrounding parcels.⁶ These vorticity-based arguments offer a straightforward and insightful means to capture the fundamentals of Rossby waves. An alternative argument, based on examining the pressure force, requires solving the Poisson equation for pressure from Section 34.5. Contrary to the simple examples offered in Section 34.5, the Rossby wave source for the pressure equation is nontrivial, even when linearizing the source by assuming small amplitude waves. We thus find the vorticity arguments to be much more transparent and streamlined.



34.8 Exercises

EXERCISE 34.1: INTEGRAL PROPERTIES OF THE INVISCID 2D NON-DIVERGENT FLOW

In this exercise, we establish some global conservation properties for inviscid two-dimensional non-divergent flow on a β -plane. Assume the geometry is a flat plane defined over a static region \mathcal{S} . The region can either be infinite, in which case all fields decay to zero at infinity, or a finite domain surrounded by static material boundaries. There are no boundary fluxes of matter, so that the fluid domain is material. Many of the properties derived here are discussed in Section 3.1 of [McWilliams \(2006\)](#).

⁶ Analogous arguments hold for a stratified fluid, with material conservation of potential vorticity providing the fundamental constraint. See Chapter 6 of [Vallis \(2017\)](#) for details.

- (a) Show that the domain integrated kinetic energy per mass remains constant in time

$$\mathcal{K} = \frac{1}{2} \int_{\mathcal{S}} \mathbf{u} \cdot \mathbf{u} \, d\mathcal{S}, \quad (34.89)$$

where the horizontal integral extends over the full fluid domain \mathcal{S} .

- (b) Why is the mechanical energy budget only associated with kinetic energy? What about the gravitational potential energy?
(c) Show that the domain integrated relative vorticity (equal also to the relative circulation) is constant in time

$$\mathcal{C} = \int_{\mathcal{S}} \zeta \, d\mathcal{S}. \quad (34.90)$$

- (d) Show that the domain integrated enstrophy is constant in time for f -plane motion ($\beta = 0$)

$$Z^{(\zeta)} = \int_{\mathcal{S}} \zeta^2 \, d\mathcal{S}. \quad (34.91)$$

- (e) Show that the domain integrated potential enstrophy is constant in time even with $\beta \neq 0$

$$Z^{(q)} = \int_{\mathcal{S}} q^2 \, d\mathcal{S}. \quad (34.92)$$

EXERCISE 34.2: CIRCULATION IN A 2D BAROTROPIC FLOW

Consider a non-divergent barotropic flow on a β -plane in the presence of a biharmonic friction operator, where the governing vorticity equation is

$$\frac{\partial \zeta}{\partial t} + J(\psi, \zeta + \beta y) = -\nu \nabla^4 \zeta, \quad (34.93)$$

with $\nu > 0$ a constant biharmonic viscosity with dimensions of $L^4 T^{-1}$. Show that the circulation around a fixed material area, \mathcal{S} , in the fluid evolves according to

$$\frac{d\mathcal{C}}{dt} = - \oint_{\partial\mathcal{S}} \left[\psi \frac{\partial q}{\partial s} + \nu \frac{\partial (\nabla^2 \zeta)}{\partial n} \right] ds, \quad (34.94)$$

where s is the arc-length along the boundary of the region and n is a coordinate normal to the boundary.

EXERCISE 34.3: DYNAMICS OF VORTICITY GRADIENTS

For many purposes it is of interest to develop equations describing the evolution of scalar gradients. We developed a general expression in Exercise 13.2. Here, we develop a similar equation for the gradient of relative vorticity in a non-divergent barotropic flow. For this purpose, consider the inviscid barotropic vorticity equation on an f -plane

$$\frac{\partial \zeta}{\partial t} + J(\psi, \zeta) = 0. \quad (34.95)$$

- (a) Show that the material evolution of the vorticity gradient is given by

$$\frac{D(\nabla \zeta)}{Dt} = -J(\nabla \psi, \zeta). \quad (34.96)$$

- (b) Show that the material evolution of the squared vorticity gradient is given by

$$\frac{D|\nabla\zeta|^2}{Dt} = 2 J(\zeta, \nabla\psi) \cdot \nabla\zeta. \quad (34.97)$$

EXERCISE 34.4: ANGULAR MOMENTUM FOR BAROTROPIC FLOW IN A BASIN

The exercise derives some equations presented in [Holloway and Rhines \(1991\)](#), who offer a specialized example of the shallow water angular momentum discussed in Section 32.7.

As in Section 32.7.1, the relative angular momentum for a region of fluid is given by

$$\mathbf{L} = \int d\mathcal{S} \int (\mathbf{x} \wedge \mathbf{v}) \rho dz, \quad (34.98)$$

where \mathbf{x} is the position vector and the relative angular momentum is that due to the motion of the fluid with respect to the solid body. For a barotropic fluid of constant density and constant thickness, and correspondingly a zero vertical velocity, the relative angular momentum reduces to

$$\mathbf{L} = \rho H \int_{\mathcal{S}} (\mathbf{x} \wedge \mathbf{u}) d\mathcal{S}, \quad (34.99)$$

with \mathbf{u} the horizontal velocity and \mathcal{S} the horizontal region. For barotropic motion on a tangent plane we are interested in the vertical component of the relative angular momentum

$$L^z = \rho H \int_{\mathcal{S}} \hat{\mathbf{z}} \cdot (\mathbf{x} \wedge \mathbf{u}) d\mathcal{S}. \quad (34.100)$$

Show for a simply connected and bounded region, L^z can be written

$$L^z = 2 \rho H \int_{\mathcal{S}} (\psi_b - \psi) d\mathcal{S} \quad (34.101)$$

where ψ is the streamfunction satisfying $\mathbf{u} = \hat{\mathbf{z}} \wedge \nabla\psi$, and ψ_b is the value of the streamfunction evaluated on the region boundary. Hint: note that $\nabla \cdot \mathbf{x} = 2$ for a horizontal position vector. Also recall from Section 17.4.2 that the streamfunction equals to a spatial constant when evaluated along the domain boundary.

EXERCISE 34.5: STEADY AXIALLY SYMMETRIC FLOW

Consider a two-dimensional non-divergent velocity

$$\mathbf{v} = \hat{\mathbf{z}} \wedge \nabla\psi. \quad (34.102)$$

Assume the streamfunction is static and depends only on the radial distance from an arbitrary origin,

$$\psi = \psi(r), \quad (34.103)$$

where $r = \sqrt{x^2 + y^2}$, and assume the velocity is a solution to the steady inviscid non-divergent barotropic dynamics on an f -plane.

- (a) Show that the velocity only has an angular component

$$\mathbf{v} = v^\varphi \hat{\boldsymbol{\varphi}}, \quad (34.104)$$

where $\hat{\boldsymbol{\varphi}}$ is the angular unit vector oriented counter-clockwise from the $\hat{\mathbf{x}}$ axis. Express v^φ in terms of the streamfunction ψ . Hint: see Figure 8.2 and Section 8.3 for a reminder of polar coordinates.

(b) Write the relative vorticity in terms of the streamfunction using polar coordinates.

(c) Consider the circulation

$$\mathcal{C} = \oint_{\partial S} \mathbf{v} \cdot d\mathbf{r}, \quad (34.105)$$

where S is a circular region in the $x - y$ plane centered at $r = 0$. Express the circulation in terms of v^φ and the radius of the circle.

(d) Write the pressure gradient acceleration in terms of v^φ , f , and r . Hint: remember that $\hat{\varphi}$ is a function of the polar angle φ .

(e) Interpret the steady balance of accelerations in terms of the balanced dynamics in Chapter 27.

(f) Why is this axial symmetric solution only valid for an f -plane? Hint: show that if $\beta \neq 0$ that there is an inconsistency in the velocity equation.

EXERCISE 34.6: GALILEAN TRANSFORMATION OF PV ADVECTION AND THE APV METHOD

In Section 13.6 we established the invariance of the material time derivative operator under a Galilean transformation

$$\bar{\mathbf{x}} = \mathbf{x} + \mathbf{U} t \quad \text{and} \quad \bar{\mathbf{u}} = \mathbf{u} + \mathbf{U}, \quad (34.106)$$

where \mathbf{U} is a constant. Here we study the Galilean transformation properties of the non-divergent barotropic model on a β -plane.

(a) Determine the Galilean transformation properties of the potential vorticity equation (34.28)

$$\frac{\partial q}{\partial t} + \mathbf{u} \cdot \nabla q = \frac{\partial q}{\partial t} + J(\psi, q) = 0, \quad (34.107)$$

where $q = \zeta + f$, $\hat{\mathbf{z}} \wedge \psi = \mathbf{u}$, and J is the Jacobian operator.

(b) Determine the Galilean transformation properties of the relative vorticity equation (34.25)

$$\frac{\partial \zeta}{\partial t} + \mathbf{u} \cdot \nabla \cdot \zeta = -\beta v. \quad (34.108)$$

Discuss why there is Galilean invariance *only* for zonal Galilean boosts, $\mathbf{U} = \hat{\mathbf{x}} U$.

(c) An Euler forward time stepping scheme for the PV equation leads to

$$q^{n+1} = q^n - \Delta t \mathbf{u}^n \cdot \nabla q^n, \quad (34.109)$$

where Δt is the discrete time step and the integer n represents the discrete time. Inspired by this time discrete expression, [Sadourny and Basdevant \(1985\)](#) proposed the *anticipated potential vorticity* (APV) method for parameterizing subgrid scale processes. The simplest form of APV is given by

$$\frac{\partial q}{\partial t} = -\mathbf{u} \cdot \nabla [q - \tau \mathbf{u} \cdot \nabla q] = -J[\psi, q - \tau J(\psi, q)], \quad (34.110)$$

with τ a constant time scale. From the time discrete expression (34.109), we see that the APV method makes use of an estimate for the future value of PV in computing the advection operator, thus motivating the term “anticipated” in the method’s name.

Show that $\tau \neq 0$ breaks Galilean invariance for the equation (34.110). Provide a discussion of why invariance is broken. Hint: [Vallis and Hua \(1988\)](#) offer a technical reason for why Galilean invariance is broken, making use of the streamfunction and Jacobian form. You do not necessarily need to follow their approach. Rather, it is sufficient to merely note how velocity appears in the APV operator.

EXERCISE 34.7: ELEMENTS OF THE FOFONOFF GYRE

A Fofonoff gyre is an unforced inviscid solution in a flat bottom bounded domain with a rigid lid. For a single layer of homogeneous fluid with constant thickness, the absolute vorticity is materially invariant, $D(\zeta + f)/Dt = 0$. An explicit solution is derived in Section 19.5.3 of [Vallis \(2017\)](#) for quasi-geostrophic flow using the method of matched asymptotics. We depict elements of a double Fofonoff gyre in Figure 34.7. Provide a narrative for this flow based on material conservation of absolute vorticity. In particular, discuss how the flow enters and leaves the side boundaries and conversely how it leaves and enters the interior region. We are not concerned with how this flow is established. Instead, assume the flow exists and discuss how its existence is consistent with $D(\zeta + f)/Dt = 0$. Hint: recall our discussion of Figure 34.3.

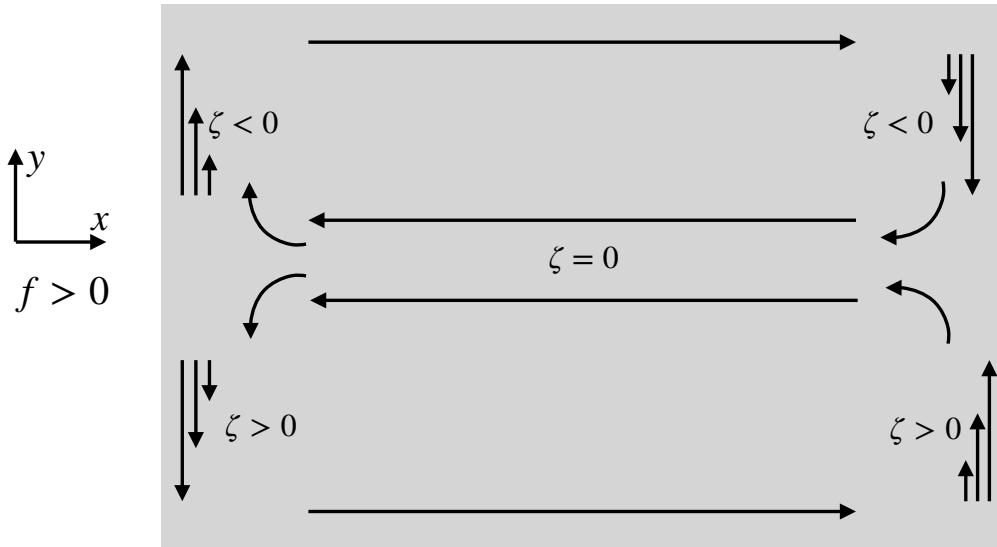


FIGURE 34.7: A Fofonoff gyre is an unforced inviscid flow in a bounded domain where $D(\zeta + f)/Dt = 0$. We here depict elements of this double-gyre flow in the northern hemisphere as part of Exercise 34.7.



Shallow water vorticity and potential vorticity

In this chapter we study vorticity and potential vorticity within the shallow water system. We start by deriving the evolution equation for vorticity by taking the curl of the momentum equation. Combining vorticity evolution with mass continuity then renders the evolution equation for potential vorticity. Potential vorticity is a material invariant for inviscid shallow water motion, thus providing a mechanical constraint on the fluid flow. After developing the basic concepts and equations, we consider a variety of flow regimes and case studies, mostly with an ocean focus, where the study of vorticity, potential vorticity, and circulation enhances our understanding of geophysical fluid mechanics.

READER'S GUIDE FOR THIS CHAPTER

The shallow water fluid offers a fruitful conceptual model to introduce the dynamics of vorticity and potential vorticity while requiring a relatively modest level of mathematical details. Even so, we require vector calculus identities for Cartesian coordinates as detailed in Chapter 2. We also require an understanding of shallow water mechanics from Chapters 31 and 32, as well as the vorticity kinematics introduced in Chapter 33. The concepts and methods developed in this chapter are fundamental to the remaining chapters in this part of the book.

35.1	Vector invariant velocity equation	788
35.1.1	Basic manipulations	788
35.1.2	Dynamical pressure and the Magnus acceleration	789
35.2	Shallow water vorticity equation	789
35.2.1	Vorticity equation for a single layer	790
35.2.2	Vorticity equation for N -layers	790
35.2.3	Vorticity flux divergence and curl of nonlinear advection	791
35.3	Potential vorticity for a rotating cylinder	792
35.3.1	Mass conservation	792
35.3.2	Angular momentum conservation	792
35.3.3	Potential vorticity conservation	793
35.3.4	Connecting angular momentum and vorticity	794
35.3.5	Comments and further study	794
35.4	Shallow water potential vorticity	794
35.4.1	Mass conservation plus the vorticity equation	794
35.4.2	Motivating the name	796
35.4.3	Mass conservation + Kelvin's circulation theorem	796
35.4.4	A fluid column with constant $f \neq 0$	798
35.4.5	Material conservation of an arbitrary function of PV	798

35.4.6	<i>N</i> -layer potential vorticity	798
35.4.7	Further study	799
35.5	Potential vorticity with non-conservative processes	799
35.5.1	Material time evolution of PV	800
35.5.2	The potential vorticity flux	800
35.6	Example implications of PV material invariance	801
35.6.1	Topographic beta effect	801
35.6.2	Planetary geostrophic PV	802
35.6.3	Spin up of converging flow	803
35.6.4	Further study	803
35.7	Geostrophic adjustment	804
35.7.1	Potential vorticity inversion	804
35.7.2	Posing the initial value problem	805
35.7.3	Adjustment in the absence of rotation	806
35.7.4	Adjustment with rotation	807
35.7.5	Comments	808
35.8	Circulation with non-conservative processes	809
35.8.1	Circulation around a closed streamline in steady flow	810
35.8.2	Circulation from wind stress and Rayleigh drag	810
35.9	Western intensification of ocean gyres	811
35.9.1	Steady and large-scale vorticity balance	811
35.9.2	Planetary geostrophic flow and the Sverdrup balance	812
35.9.3	Sverdrup flow in a closed domain with anti-cyclonic wind stress	812
35.9.4	Western intensification and the role of beta	813
35.9.5	A role for topographic form stress	814
35.9.6	Comments and further study	816
35.10	Column vorticity	817
35.10.1	Formulating the column vorticity equation	817
35.10.2	Summary of the column vorticity equation	818
35.10.3	Steady linear column vorticity balance and the island rule	819
35.10.4	Further study	820
35.11	Exercises	821

35.1 Vector invariant velocity equation

Prior to deriving the vorticity equation, we massage the momentum equation into its vector invariant form. Doing so exposes the dynamic pressure effects from kinetic energy per mass as well as the Magnus acceleration, both of which were encountered for the continuously stratified fluid in Section 18.2.3.

35.1.1 Basic manipulations

To start, introduce vorticity for the full velocity field as well as that just for the horizontal flow

$$\boldsymbol{\omega} = \nabla \wedge \mathbf{v} \tag{35.1a}$$

$$\boldsymbol{\omega}^* = \nabla \wedge \mathbf{u} \equiv \zeta \hat{\mathbf{z}}, \tag{35.1b}$$

where

$$\zeta = \hat{\mathbf{z}} \cdot \boldsymbol{\omega} = \hat{\mathbf{z}} \cdot \boldsymbol{\omega}^* = \partial_x v - \partial_y u \tag{35.2}$$

is the vertical component to the relative vorticity. The vector identity (see Section 2.3.4)

$$(\mathbf{u} \cdot \nabla) \mathbf{u} = (1/2) \nabla (\mathbf{u} \cdot \mathbf{u}) - \mathbf{u} \wedge (\nabla \wedge \mathbf{u}) = (1/2) \nabla (\mathbf{u} \cdot \mathbf{u}) + \boldsymbol{\omega}^* \wedge \mathbf{u} \tag{35.3}$$

brings the inviscid shallow water momentum equation (31.6)

$$\frac{\partial \mathbf{u}}{\partial t} + (\mathbf{u} \cdot \nabla) \mathbf{u} + f \hat{\mathbf{z}} \wedge \mathbf{u} = -g \nabla \eta \quad (35.4)$$

into its *vector invariant* form

$$\frac{\partial \mathbf{u}}{\partial t} + (f + \zeta) \hat{\mathbf{z}} \wedge \mathbf{u} = -\nabla (g \eta + \mathbf{u} \cdot \mathbf{u}/2). \quad (35.5)$$

In the following we sometimes make use of the vertical component to the absolute vorticity

$$\boldsymbol{\omega}_a^* = (\zeta + f) \hat{\mathbf{z}} = \zeta_a \hat{\mathbf{z}} \quad (35.6)$$

which is the sum of the relative vorticity of the horizontal flow, $\boldsymbol{\omega}^* = \zeta \hat{\mathbf{z}}$, plus the solid-body vorticity, $f \hat{\mathbf{z}}$, due to motion of the rotating reference frame (recall Section 33.5.1).

35.1.2 Dynamical pressure and the Magnus acceleration

As for the discussion of vector invariant velocity equation in Section 18.2.2, the velocity equation (35.5) exposes two physical features that lend insight into the motion of a shallow water fluid column.

Dynamical pressure from kinetic energy per mass

The kinetic energy per mass, $\mathbf{u} \cdot \mathbf{u}/2$, adds a *dynamical pressure* to the mechanical pressure associated with the free surface undulations. Gradients in the kinetic energy thus drive accelerations towards regions of smaller kinetic energy; i.e., down the kinetic energy gradient.

Magnus acceleration

As discussed in Sections 18.2.2 and 27.2.6, the Magnus acceleration is a body acceleration defined by the nonlinear term

$$\mathbf{A}_{\text{magnus}} = -\boldsymbol{\omega}^* \wedge \mathbf{u} = \zeta (\mathbf{u} \wedge \hat{\mathbf{z}}), \quad (35.7)$$

appearing in the vector-invariant velocity equation (35.5). Hence, there is a non-zero Magnus acceleration when a shallow water fluid column spins while it moves, with this acceleration acting to deflect the spinning column perpendicular to its trajectory. As for the example in Figure 18.1, consider a shallow water fluid column moving zonally, in which case the Magnus acceleration is

$$\mathbf{A}_{\text{magnus}} = u \zeta (\hat{\mathbf{x}} \wedge \hat{\mathbf{z}}) = -\hat{\mathbf{y}} u \zeta. \quad (35.8)$$

With a positive relative vorticity, $\zeta > 0$, the Magnus acceleration is directed to the right of the motion, which is in the same direction as the Coriolis acceleration in the northern hemisphere. For large-scale geophysical flows, the Magnus acceleration is sub-dominant to the Coriolis acceleration. However, the Magnus acceleration is a crucial facet of large (order unity or larger) Rossby number motions in which relative vorticity is sizable.

35.2 Shallow water vorticity equation

In this section we formulate the vorticity equation for the shallow water fluid, starting with a single layer and then extending to multiple layers.

35.2.1 Vorticity equation for a single layer

We make use of the vector identity from Section 2.3.4 to express the curl of the Magnus acceleration plus Coriolis acceleration in the form

$$\nabla \wedge (\boldsymbol{\omega}_a^* \wedge \mathbf{u}) = \boldsymbol{\omega}_a^* (\nabla \cdot \mathbf{u}) + (\mathbf{u} \cdot \nabla) \boldsymbol{\omega}_a^* - \mathbf{u} (\nabla \cdot \boldsymbol{\omega}_a^*) - (\boldsymbol{\omega}_a^* \cdot \nabla) \mathbf{u} \quad (35.9a)$$

$$= \boldsymbol{\omega}_a^* (\nabla \cdot \mathbf{u}) + (\mathbf{u} \cdot \nabla) \boldsymbol{\omega}_a^*, \quad (35.9b)$$

so that

$$\hat{\mathbf{z}} \cdot [\nabla \wedge (\boldsymbol{\omega}_a^* \wedge \mathbf{u})] = \nabla \cdot (\mathbf{u} \zeta_a) \quad (35.10)$$

Equation (35.9b) required setting

$$\nabla \cdot \boldsymbol{\omega}_a^* = \nabla \cdot \boldsymbol{\omega}^* + \nabla \cdot (f \hat{\mathbf{z}}) = 0, \quad (35.11)$$

which follows since this expression involves the divergence of a curl (first right hand side term) and since f has no z dependence. We furthermore set

$$(\boldsymbol{\omega}_a^* \cdot \nabla) \mathbf{u} = |\boldsymbol{\omega}_a^*| \partial_z \mathbf{u} = 0, \quad (35.12)$$

which follows since the horizontal velocity in a shallow water fluid is depth independent within a layer (see Section 31.2). Applying the operator $\hat{\mathbf{z}} \cdot (\nabla \wedge)$ onto the vector-invariant momentum equation (35.5) annihilates the gradient of pressure and kinetic energy, with the identity (35.10) leading to the Eulerian flux-form evolution equation for absolute vorticity

$$\frac{\partial \zeta_a}{\partial t} + \nabla \cdot (\mathbf{u} \zeta_a) = 0. \quad (35.13)$$

This equation says that the vertical component to the absolute vorticity, ζ_a , at a point in the inviscid shallow water fluid changes according to the horizontal convergence of vorticity advected to that point

$$\frac{\partial \zeta_a}{\partial t} = -\nabla \cdot (\mathbf{u} \zeta_a). \quad (35.14)$$

We can write the vorticity equation (35.13) in the material form

$$\frac{D \zeta_a}{Dt} = -\zeta_a \nabla \cdot \mathbf{u}, \quad (35.15)$$

where the material time derivative for the shallow water fluid is determined just by the horizontal flow

$$\frac{D}{Dt} = \frac{\partial}{\partial t} + \mathbf{u} \cdot \nabla = \frac{\partial}{\partial t} + u \partial_x + v \partial_y. \quad (35.16)$$

The material evolution equation (35.15) means that the absolute vorticity of a shallow water fluid particle, moving with the horizontal flow, changes according to the horizontal convergence of the fluid flow as multiplied by the absolute vorticity.

35.2.2 Vorticity equation for N -layers

The previous results for a single layer are readily extended to N -layers. For that purpose recall the velocity for layer- k evolves according to equation (31.77)

$$[\partial_t + (\mathbf{u}_k \cdot \nabla)] \mathbf{u}_k + f \hat{\mathbf{z}} \wedge \mathbf{u}_k = -\nabla p_k / \rho_1, \quad (35.17)$$

where there is no implied summation on the layer index k , and where layer pressure, p_k , is given by equation (31.73). Use of the vector identity (35.3) brings the layer velocity equation (35.17) to the vector invariant form

$$\partial_t \mathbf{u}_k + (f + \zeta_k) \hat{\mathbf{z}} \wedge \mathbf{u}_k = -\nabla(p_k/\rho_1 + \mathbf{u}_k \cdot \mathbf{u}_k/2), \quad (35.18)$$

where $\zeta_k = \hat{\mathbf{z}} \cdot (\nabla \wedge \mathbf{u}_k)$ is the vertical component to the layer- k relative vorticity. Taking the curl and making use of the mathematical identities used for single layer in Section 35.2.1 renders the vorticity equation for layer- k

$$\frac{\partial \zeta_{ak}}{\partial t} + \nabla \cdot (\mathbf{u}_k \zeta_{ak}) = 0 \iff \frac{D_k \zeta_{ak}}{Dt} = -\zeta_{ak} \nabla \cdot \mathbf{u}_k \quad (35.19)$$

where

$$\zeta_{ak} = f + \zeta_k \quad (35.20)$$

is the vertical component to the absolute vorticity of layer- k . Hence, the vorticity equation for an arbitrary layer in a stacked shallow water model is isomorphic to that for a single shallow water layer.

35.2.3 Vorticity flux divergence and curl of nonlinear advection

We revisit the manipulations from Section 35.2.1 to explicitly identify a connection between the nonlinear terms in the vorticity equation. Start by writing the velocity equation in the advective form and the vector invariant form

$$\frac{\partial \mathbf{u}}{\partial t} + (\mathbf{u} \cdot \nabla) \mathbf{u} + f \hat{\mathbf{z}} \wedge \mathbf{u} = -g \nabla \eta \quad (35.21a)$$

$$\frac{\partial \mathbf{u}}{\partial t} + (f + \zeta) \hat{\mathbf{z}} \wedge \mathbf{u} = -\nabla(g \eta + \mathbf{u} \cdot \mathbf{u}/2). \quad (35.21b)$$

Taking their curl yields two expressions of the vorticity equation

$$\frac{\partial \boldsymbol{\omega}^*}{\partial t} + \nabla \wedge [(\mathbf{u} \cdot \nabla) \mathbf{u}] + \nabla \wedge [f \hat{\mathbf{z}} \wedge \mathbf{u}] = 0 \quad (35.22a)$$

$$\frac{\partial \boldsymbol{\omega}^*}{\partial t} + \nabla \wedge [(f + \zeta) \hat{\mathbf{z}} \wedge \mathbf{u}] = 0, \quad (35.22b)$$

whose equality leads to

$$\nabla \wedge [(\mathbf{u} \cdot \nabla) \mathbf{u} - \zeta \hat{\mathbf{z}} \wedge \mathbf{u}] = 0. \quad (35.23)$$

Making use of the identity

$$\hat{\mathbf{z}} \cdot [\nabla \wedge (\zeta \hat{\mathbf{z}} \wedge \mathbf{u})] = \nabla \cdot (\mathbf{u} \zeta) \quad (35.24)$$

renders the relation

$$\hat{\mathbf{z}} \cdot \nabla \wedge [(\mathbf{u} \cdot \nabla) \mathbf{u}] = \nabla \cdot (\mathbf{u} \zeta). \quad (35.25)$$

We thus see that the divergence of the advective vorticity flux (right hand side) equals to the curl of the nonlinear advection (left hand side). This identity holds for each layer in an N -layer shallow water model.

35.3 Potential vorticity for a rotating cylinder

To introduce potential vorticity, consider a fluid cylinder of constant mass M , constant density ρ , variable radius R , and variable height h , and assume the cylinder exhibits solid-body rotation about its central axis. This analysis offers a useful conceptual picture for a rotating column of a shallow water layer, in which time derivatives in the following are interpreted as material derivatives. It also lends intuition for the motion of spinning fluids constrained by mass and angular momentum conservation.

35.3.1 Mass conservation

With a constant density, mass conservation for the cylinder means that its volume is fixed. Hence, mass conservation constrains the relative changes to the radius and height of the cylinder. A materially constant cylinder mass

$$M = \pi R^2 h \rho \quad (35.26)$$

implies

$$\frac{2}{R} \frac{DR}{Dt} = -\frac{1}{h} \frac{Dh}{Dt}. \quad (35.27)$$

That is, mass conservation means that as twice the relative radius increases the relative height decreases. That is, if the cylinder is squashed (h decreases) it thickens (R increases) and conversely the cylinder thins (R decreases) as it extends (h increases).

35.3.2 Angular momentum conservation

A second constraint arises from angular momentum conservation. Since the cylinder is rotating as a solid-body, angular momentum is straightforward to compute. For simplicity, choose the center of mass coordinate axes through the center of the cylinder, with the z -axis along the central line of the cylinder and with $z = 0$ at the cylinder mid-point. The angular rotation vector is thus given by

$$\boldsymbol{\Omega} = \Omega \hat{z}. \quad (35.28)$$

With this axis orientation, the solid-body rotation occurs about the center of mass so that the angular momentum of the center of mass vanishes. The moment of inertia tensor for a cylinder with this axis orientation is given by (e.g., [Marion and Thornton, 1988](#))

$$I_{mn} = \delta_{mn} \frac{MR^2}{2}. \quad (35.29)$$

The moment of inertia is a measure of the rotational inertia of the cylinder, and is seen to be directly related to the cylinder mass (assumed fixed here) and radius (can change). Note that the moment of inertia is not a function of the cylinder height. The angular momentum for the cylinder is thus given by

$$\mathbf{L} = \frac{MR^2}{2} \Omega \hat{z}. \quad (35.30)$$

The familiar *ice skater* example occurs when the cylinder radius changes and thus changes the moment of inertia (e.g., the ice skater's arms are brought in toward the central axis of the body or out away from the body). Maintaining constant angular momentum and constant mass means that

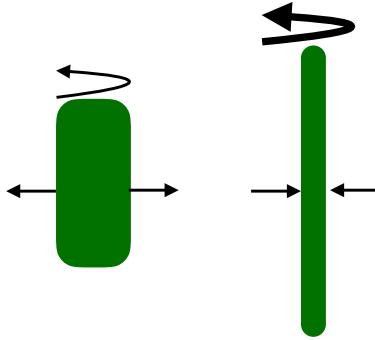


FIGURE 35.1: Illustrating the conservation of angular momentum for a spinning cylinder of constant mass undergoing solid-body rotation around its central axis. The moment of inertia (relative to the central axis) for the left cylinder is larger since more of its mass is distributed away from the central axis than in the right configuration. If the initial spin for the left cylinder is counterclockwise, then for the two configurations to have identical angular momentum means that the right cylinder spins more rapidly since its moment of inertia is smaller. This example exemplifies the familiar ice skater experience, whereby the skater's spin rate increases when bringing arms (mass) inward towards the central axis of the body (depicted by the inward arrows on the right panel), and the skater slows when extending the arms outward (depicted by the outward arrows on the left panel).

the angular velocity, Ω , increases in magnitude (rotates faster) when the cylinder radius decreases, and vice versa. Explicitly for the cylinder we have $d\mathbf{L}/dt = 0$ and $dM/dt = 0$ thus rendering

$$\frac{2}{R} \frac{DR}{Dt} = -\frac{1}{\Omega} \frac{D\Omega}{Dt}. \quad (35.31)$$

This tradeoff between spin rate and radius holds for a fluid and we encounter it again in Figure 35.6 for a layer of shallow water fluid. Namely, reducing the moment of inertia for a constant mass body by bringing its mass distribution towards the central axis (converging mass) leads, through angular momentum conservation, to an increase in rotation speed. The opposite occurs when mass diverges from a region, thus reducing the rotation speed.

Although the angular momentum constraint means that the spin rate changes when changing the moment of inertia, it does not impose a preferred direction. For example, no matter what direction a skater is rotating, decreasing the moment of inertia increases the spin rate in that particular direction. Yet when placing the spinning column on a rotating planet, the planetary rotation breaks the symmetry and thus prescribes the direction for the spin changes. The reason is that planetary rotation contributes to the spin of the column, even if the column has no spin relative to the rotating planetary reference frame. We encounter this additional part of the story in Section 35.4.

35.3.3 Potential vorticity conservation

Combining angular momentum conservation (35.31) with mass conservation (35.27) leads to the conservation law

$$\frac{D}{Dt} \frac{\Omega}{h} = 0. \quad (35.32)$$

Equation (35.32) is a statement of potential vorticity conservation for the material fluid column, with potential vorticity given by

$$Q \equiv \frac{\Omega}{h}. \quad (35.33)$$

For example, if the column thickens then the rotational velocity increases to maintain $Q = \Omega/h$ constant. Equivalently, if the column cross-sectional area decreases, the column thickness increases according to volume conservation, which in turn results in an increase in the spin according to angular momentum conservation.

35.3.4 Connecting angular momentum and vorticity

When allowing the fluid to exhibit motion that is more general than a solid-body, then the angular rotation rate appearing in the potential vorticity (35.33) is generalized to the absolute vorticity, and we will find this generalization in Section 35.4. Furthermore, as shown in Section 33.5, the vorticity for solid body motion equals to twice the rotation rate, 2Ω . Hence, the numerator for the potential vorticity of the solid-body rotating cylinder equals to one-half the vorticity. In Chapter 37, we connect angular momentum and vorticity (and strain) for arbitrary fluid motion.

35.3.5 Comments and further study

The discussion in this section is motivated by Section 2.4 of [Salmon \(1998\)](#). The solid-body rotating cylinder succinctly identifies the two mechanical properties contributing to the potential vorticity conservation law (35.32): a kinematic property (mass conservation) and a dynamic property (angular momentum conservation). For the solid-body rotating cylinder, the implications of PV conservation are well gleaned from the separate mass and angular momentum conservation principles. Hence, PV conservation lends little novel insight for the cylinder. However, PV conservation is of fundamental use for studies of rotating and stratified fluids.

We also note that an important element missing from this discussion is the *beta effect*, which accounts for the changes in planetary vorticity when moving on a rotating spherical planet. We encounter this effect in the following sections.

35.4 Shallow water potential vorticity

We now consider the potential vorticity for a single layer of shallow water fluid. The result is directly analogous to that derived for the rotating cylinder in Section 35.3. However, the derivation here makes use of fluid mechanical equations rather than those from rigid body mechanics. We present two derivations: one based on manipulations of the mass and momentum equations, and one based on the small aspect ratio limit of Kelvin's circulation theorem, with Kelvin's Theorem more thoroughly studied in Chapter 36. Figure 35.2 summarizes key elements leading to PV conservation for a shallow water fluid layer. Namely, as shown in this section, shallow water PV conservation arises from combining the kinematic constraint of mass conservation (material conservation of $h A$) with either the vorticity equation or Kelvin's circulation theorem for a small aspect ratio fluid.

35.4.1 Mass conservation plus the vorticity equation

To derive the potential vorticity equation, we here make use of the vorticity equation (35.15) and combine it with mass conservation.

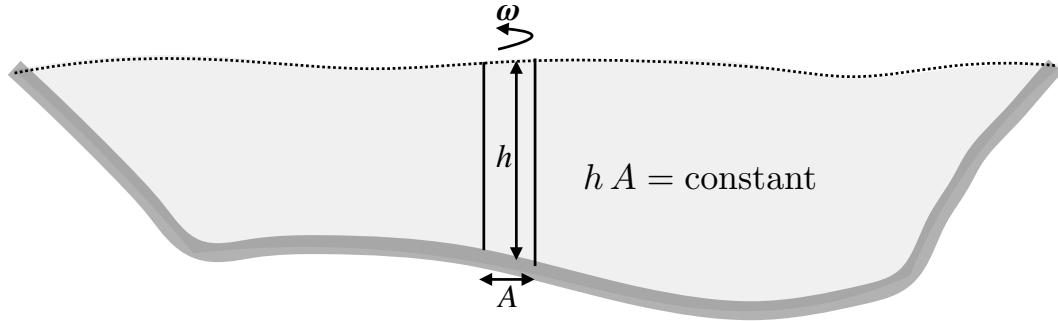


FIGURE 35.2: Illustrating the conservation of PV for a layer of shallow water fluid column. PV conservation results from merging mass conservation (material conservation of the column volume, $h A$), to either the vorticity equation or Kelvin's circulation theorem for a small aspect ratio fluid (material conservation of ζA).

Shallow water vorticity and vortex stretching

Mass conservation in the form of the material thickness equation (31.17) leads to the following expression for the divergence of the horizontal velocity

$$\nabla \cdot \mathbf{u} = -\frac{1}{h} \frac{Dh}{Dt}. \quad (35.34)$$

Making use of this result in the vorticity equation (35.15) allows us to eliminate the horizontal convergence

$$\frac{D\zeta_a}{Dt} = -\zeta_a \nabla \cdot \mathbf{u} = \frac{\zeta_a}{h} \frac{Dh}{Dt}. \quad (35.35)$$

This equation says that material changes in shallow water absolute vorticity arise only from material changes in the layer thickness; i.e., absolute vorticity increases in magnitude if the column stretches and decreases if the column compresses. We see in Section 36.6.3 that vorticity in more general fluids is affected by vortex stretching, as in the shallow water fluid, as well as vortex tilting and torques from baroclinicity. In contrast, equation (35.35) says that the material evolution of absolute vorticity for a shallow water fluid is affected only through vortex stretching. This behavior is a result of the depth independence of the horizontal velocity within a shallow water layer and the associated vertical columnar motion of fluid within the layer.

Material invariance of PV

Equation (35.35) can be written as an expression of the material invariance of the shallow water potential vorticity

$$\frac{DQ}{Dt} = 0, \quad (35.36)$$

where

$$Q = \frac{\zeta_a}{h} = \frac{\zeta + f}{h} \quad (35.37)$$

is the shallow water potential vorticity. As defined, shallow water potential vorticity is the ratio of absolute vorticity to the thickness of the fluid layer. The material conservation law (35.36) says that this ratio remains constant for the shallow water layer in the absence of irreversible processes. Much of the remainder of this chapter explores the many implications of potential vorticity conservation.

35.4.2 Motivating the name

Material invariance of the shallow water potential vorticity in equation (35.37) is most practically a statement about how the relative vorticity, ζ , changes when changing column thickness or latitude. That is, by maintaining Q fixed, ζ must change when either the column thickness, h , changes or when the column moves meridionally to thus alter the Coriolis parameter, f . By focusing on relative vorticity we are offered insights into how the fluid motion is constrained and thus a means to predict changes in that motion. In turn, these changes in ζ motivate the name “potential vorticity” as we now see.

Potential vorticity measures the ability for a shallow water fluid column to either spin up or spin down (change its relative vorticity) relative a standard configuration. For example, let the standard configuration be defined by an arbitrary standard thickness, h_s , at the equator (where $f = 0$). Now move an off-equatorial shallow water fluid column with zero relative vorticity to the equator and stretch/compress the column to the standard thickness. Material invariance of the column’s potential vorticity allows us to deduce the column’s relative vorticity at the equator, given information about the initial column thickness and initial Coriolis parameter (see Figure 35.3). Hence, potential vorticity, as an invariant material property, provides the “potential” for a fluid column to manifest a particular value of the relative vorticity when moved and stretched into a standard configuration. In this manner, the use of “potential” in “potential vorticity” is directly analogous to the use of “potential” in “potential temperature” as described in Section 21.8.

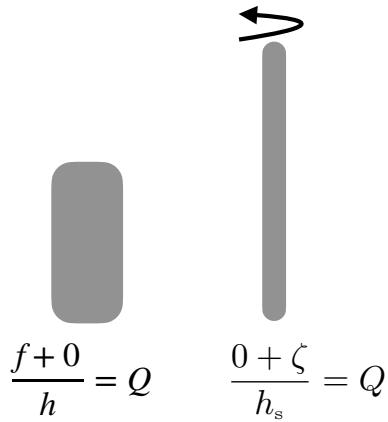


FIGURE 35.3: Left panel: an arbitrary shallow water column with zero relative vorticity and potential vorticity $Q = f/h$, with $f > 0$ assumed for this figure (northern hemisphere). Right panel: the same fluid column moved to the equator (where $f = 0$) and stretched to have the standard thickness, $h_s > h$. The relative vorticity of the column at the equator is given by $\zeta = f(h_s/h)$, where f is the Coriolis parameter at the original latitude where $f > 0$. Potential vorticity thus provides a means to deduce the relative vorticity that can be realized by moving any particular configuration to a standard location and with a standard thickness. This property motivates the “potential” used in the name.

35.4.3 Mass conservation + Kelvin’s circulation theorem

Although we have yet to discuss Kelvin’s Theorem (Section 36.3), we here invoke it to illustrate another way to derive shallow water PV conservation. As we will see, this derivation lends itself to direct analog of the rotating cylinder discussed in Section 35.3.

When applied to an infinitesimal circuit in an inviscid and constant density fluid, Kelvin’s

theorem says that

$$\frac{D(\boldsymbol{\omega}_a \cdot \hat{\mathbf{n}} \delta\mathcal{S})}{Dt} = 0, \quad (35.38)$$

where $\boldsymbol{\omega}_a$ is the absolute vorticity

$$\boldsymbol{\omega}_a = \boldsymbol{\omega} + f \hat{\mathbf{z}}, \quad (35.39)$$

$\hat{\mathbf{n}} \delta\mathcal{S}$ is the infinitesimal surface area enclosed by the closed circuit, with $\hat{\mathbf{n}}$ the outward normal to the surface. Hence, equation (35.38) says that the projection of the absolute vorticity onto the local normal of an area element, multiplied by that area element, remains materially constant. This identity offers a very strong constraint on the flow.

To make use of equation (35.38) for the shallow water layer, decompose absolute vorticity into

$$\boldsymbol{\omega}_a = \hat{\mathbf{z}} (\zeta + f) + \boldsymbol{\omega}_h, \quad (35.40)$$

where ζ is the vertical component to the relative vorticity (equation (35.2)), and

$$\boldsymbol{\omega}_h = \hat{\mathbf{x}} \frac{\partial w}{\partial y} - \hat{\mathbf{y}} \frac{\partial w}{\partial x} \quad (35.41)$$

is the horizontal component to the relative vorticity, making use of the depth independence of the horizontal velocity components for the shallow water fluid ($\partial u / \partial z = \partial v / \partial z = 0$). Inserting the absolute vorticity (35.40) into Kelvin's theorem (35.40) leads to

$$\frac{D}{Dt} [(\zeta + f) \delta A + \boldsymbol{\omega}_h \cdot \hat{\mathbf{n}} \delta\mathcal{S}] = 0, \quad (35.42)$$

where the horizontal area element, δA , is the projection of the surface area element onto the vertical direction

$$\delta A = \hat{\mathbf{z}} \cdot \hat{\mathbf{n}} d\mathcal{S}. \quad (35.43)$$

Shallow water fluid mechanics arises from considering a constant density fluid layer under the small aspect ratio limit ($H/L \ll 1$, with H the vertical length scale and L the horizontal length scale). Under this limit, the second term in equation (35.42) is much smaller than the first. Ignoring this small term then leads to

$$\frac{D}{Dt} \left[\left(\frac{\zeta + f}{h} \right) h \delta A \right] = 0, \quad (35.44)$$

where h is the layer thickness and $h \delta A$ is the volume of an infinitesimal fluid column. Given the incompressible nature of the shallow water layer, the column volume is materially constant

$$\frac{D(h \delta A)}{Dt} = 0, \quad (35.45)$$

so that equation (35.44) yields the material invariance of shallow water PV

$$\frac{D}{Dt} \left[\frac{\zeta + f}{h} \right] = \frac{DQ}{Dt} = 0, \quad (35.46)$$

where $Q = (\zeta + f)/h$ is the same shallow water potential vorticity derived above in Section 35.4.1.

35.4.4 A fluid column with constant $f \neq 0$

Some of the essential features of shallow water potential vorticity conservation are depicted in Figure 35.4 for the case of a fluid column with constant $f \neq 0$. In the left panel, the column thickness increases (column is stretched). Volume conservation means that the column radius decreases. As the material in the column moves radially inward toward the center, it experiences a Coriolis deflection to the right in the northern hemisphere and to the left in the southern. Both of these deflections renders a cyclonic tendency to the relative vorticity, creating a positive relative vorticity tendency in the north and negative relative vorticity tendency in the south.

An equivalent way to understand the cyclonic tendency is to consider the angular momentum of the fluid column. As the radius of the column decreases so too does its moment of inertia. Angular momentum conservation means that the column picks up a tendency that causes its spin to increase. The direction of this spin increase accords with the background f of the environment. Returning to the skater analog in Figure 35.1, we consider f to be the initial spin of the skater so that when the moment of inertia decreases the column picks up a spin in the same direction as f ; i.e., a cyclonic tendency.¹

In both hemispheres, the Coriolis deflection, or equivalently angular momentum conservation, creates a cyclonic relative vorticity tendency when the column stretches, thus maintaining $Q = (f + \zeta)/h$ fixed. Again, the cyclonic relative vorticity adds to the magnitude of the planetary vorticity to ensure that the absolute vorticity magnitude increases in accord with the column thickness increase, thus keeping $Q = \zeta_a/h$ constant. The converse holds when the column is flattened/squashed, whereby the relative vorticity picks up an anti-cyclonic tendency (negative relative vorticity tendency in the northern hemisphere and positive relative vorticity tendency in the southern). Doing so reduces the magnitude of the absolute vorticity in accord with the reduced column thickness.

35.4.5 Material conservation of an arbitrary function of PV

The material conservation law for PV, equation (35.36), means that any function, $F(Q)$ is also materially constant. We see this property through the chain rule

$$\frac{DF}{Dt} = \frac{dF}{dQ} \frac{DQ}{Dt} = 0. \quad (35.47)$$

Since F is arbitrary, there are an infinite number of material invariants corresponding to distinct functions F . This result holds for all materially invariant scalar properties of the fluid.

35.4.6 N -layer potential vorticity

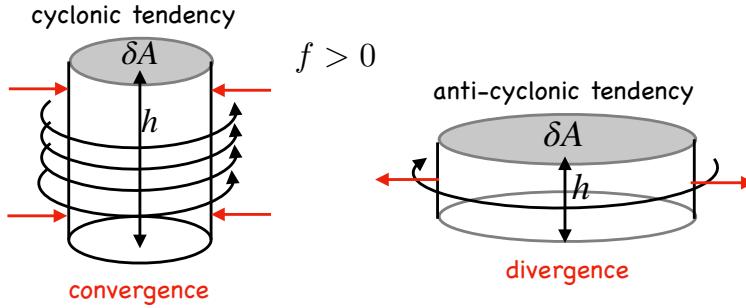
The thickness equation (31.70a) and the vorticity equation (35.19) for an N -layer shallow water model are given by

$$\frac{D_k h_k}{Dt} = -h_k \nabla \cdot \mathbf{u}_k \quad \text{and} \quad \frac{D_k \zeta_{ak}}{Dt} = -\zeta_{ak} \nabla \cdot \mathbf{u}_k, \quad (35.48)$$

where there is no implied summation over the layer index k . These forms are isomorphic to the single layer equations so that the potential vorticity of layer- k is given by

$$Q_k = \frac{f + \zeta_k}{h_k}, \quad (35.49)$$

¹The connection between the Coriolis acceleration and angular momentum conservation was discussed for particle mechanics in Section 11.8.



$$\frac{D(h \delta A)}{Dt} = 0 \quad \frac{D[\delta A (f + \zeta)]}{Dt} = 0 \implies \frac{D[(f + \zeta)/h]}{Dt} = 0$$

FIGURE 35.4: Material invariance for PV in a shallow water layer results from combining material invariance of the volume of the fluid column with the material invariance of the area weighted absolute vorticity. As the cross-sectional area of the column decreases, as in a converging flow, the thickness of the fluid must increase in order to maintain constant volume. Furthermore, PV invariance can be maintained by changing the fluid spin, as measured by the relative vorticity, or by changing the latitude and thus changing its planetary vorticity (the β -effect discussed in Section 36.7). We here depict the case with $f > 0$ constant. Flow converging toward the center of the column picks up a Coriolis acceleration that creates a cyclonic tendency in the relative vorticity, just like the spinning cylinder in Figure 35.1. Equivalently, as the radius of the column decreases so too does its moment of inertia so that the column must pick up a cyclonic tendency to conserve angular momentum. Conversely, as the cross-sectional area increases, the diverging flow creates an anti-cyclonic tendency in the relative vorticity.

and for a perfect fluid this layer potential vorticity is materially constant

$$\frac{D_k Q_k}{Dt} = \frac{\partial Q_k}{\partial t} + \mathbf{u}_k \cdot \nabla Q_k = 0, \quad (35.50)$$

where, again, there is no implied summation over k .

35.4.7 Further study

The shallow water potential vorticity (35.37) was introduced by [Rossby \(1940\)](#) and as such it is sometimes referred to as the *Rossby potential vorticity*. Non-rotating shallow water potential vorticity, ζ/h , is illustrated in this [video from Prof. Shapiro at around the 11 minute mark](#). Note that he does not use the term “potential vorticity”, instead invoking mass conservation and angular momentum conservation to describe the motion.

35.5 Potential vorticity with non-conservative processes

In this section we consider the role of a non-conservative force per mass, \mathbf{F} , with this term arising from friction and boundary stresses. Additionally, we allow for the presence of a boundary volume source, $w^{(\dot{\eta})}$ (as in precipitation minus evaporation), thus changing the volume in the layer. We introduced such processes in Section 31.5, in which case the shallow water equations take on the form

$$\frac{D\mathbf{u}}{Dt} + f \hat{z} \wedge \mathbf{u} = -g \nabla \eta + \mathbf{F} \quad \text{and} \quad \frac{Dh}{Dt} + h \nabla \cdot \mathbf{u} = w^{(\dot{\eta})}. \quad (35.51)$$

35.5.1 Material time evolution of PV

In the presence of friction, the absolute vorticity equation (35.13) takes on the form

$$\frac{\partial \zeta_a}{\partial t} + \nabla \cdot (\mathbf{u} \zeta_a) = \hat{\mathbf{z}} \cdot (\nabla \wedge \mathbf{F}) \implies \frac{D\zeta_a}{Dt} + \zeta_a \nabla \cdot \mathbf{u} = \hat{\mathbf{z}} \cdot (\nabla \wedge \mathbf{F}), \quad (35.52)$$

so that vorticity is now affected by the curl of friction. As before, we make use of the thickness equation to replace the horizontal divergence according to

$$\nabla \cdot \mathbf{u} = \frac{1}{h} \left[-\frac{Dh}{Dt} + w^{(\dot{\eta})} \right]. \quad (35.53)$$

The presence of $w^{(\dot{\eta})}$ modifies the divergence of the horizontal velocity beyond that for a conservative fluid. We are thus led to the PV equation

$$h \frac{DQ}{Dt} = -Q w^{(\dot{\eta})} + \hat{\mathbf{z}} \cdot (\nabla \wedge \mathbf{F}). \quad (35.54)$$

Hence, with $w^{(\dot{\eta})} \neq 0$ and/or $\hat{\mathbf{z}} \cdot (\nabla \wedge \mathbf{F}) \neq 0$, potential vorticity is no longer materially invariant.

35.5.2 The potential vorticity flux

Deriving the Eulerian flux-form equation

We can convert the potential vorticity equation (35.54) into an Eulerian flux-form conservation equation by making use of the thickness equation

$$h \frac{DQ}{Dt} = h \left[\frac{\partial Q}{\partial t} + \mathbf{u} \cdot \nabla Q \right] + Q \left[\frac{\partial h}{\partial t} + \nabla \cdot (h \mathbf{u}) - w^{(\dot{\eta})} \right] = \frac{\partial(h Q)}{\partial t} + \nabla \cdot (\mathbf{u} h Q) - Q w^{(\dot{\eta})}, \quad (35.55)$$

to render

$$\frac{\partial(h Q)}{\partial t} + \nabla \cdot (h \mathbf{u} Q) = \hat{\mathbf{z}} \cdot (\nabla \wedge \mathbf{F}). \quad (35.56)$$

As a final step, make use of the identity

$$\hat{\mathbf{z}} \cdot (\nabla \wedge \mathbf{F}) = -\nabla \cdot (\hat{\mathbf{z}} \wedge \mathbf{F}), \quad (35.57)$$

so that the thickness weighted PV equation (35.56) can be written

$$\frac{\partial(h Q)}{\partial t} = -\nabla \cdot (h Q \mathbf{u} + \hat{\mathbf{z}} \wedge \mathbf{F}). \quad (35.58)$$

Note how the volume source term, $w^{(\dot{\eta})}$, does not explicitly appear in the flux-form equation (35.58) since the effects from $w^{(\dot{\eta})}$ are captured by the divergence, $\nabla \cdot \mathbf{u}$, as per equation (35.53).

For the shallow water fluid, the thickness weighted potential vorticity equals to the absolute vorticity

$$h Q = \zeta_a. \quad (35.59)$$

Consequently, the flux-form conservation form of the PV equation (35.58) is identical to equation (35.52) for the absolute vorticity

$$\frac{\partial(h Q)}{\partial t} = -\nabla \cdot (h Q \mathbf{u} + \hat{\mathbf{z}} \wedge \mathbf{F}) \iff \frac{\partial \zeta_a}{\partial t} = -\nabla \cdot (\zeta_a \mathbf{u} + \hat{\mathbf{z}} \wedge \mathbf{F}). \quad (35.60)$$

Potential vorticity flux vector

It is remarkable that even with non-conservative forcing, the thickness weighted potential vorticity (equivalently, the absolute vorticity) has its Eulerian evolution determined by the convergence of a flux,

$$\mathbf{J}_Q = h Q \mathbf{u} + \hat{\mathbf{z}} \wedge \mathbf{F}. \quad (35.61)$$

The general form for the evolution of “PV-stuff” is not just specific to the shallow water fluid. Rather, it holds for any fluid and it leads to some rather special properties for potential vorticity budgets, with a detailed exploration given in Chapters 39 and 46.

The flux convergence evolution for the potential vorticity equation (35.60) is actually a kinematic result of the definition of potential vorticity. Namely,

$$h Q - f = \zeta = \nabla \cdot (\mathbf{u} \wedge \hat{\mathbf{z}}) \quad (35.62)$$

so that

$$\partial_t(h Q) = \nabla \cdot (\partial_t \mathbf{u} \wedge \hat{\mathbf{z}}) \equiv -\nabla \cdot \mathbf{J}^{\text{kin}}, \quad (35.63)$$

where we defined the kinematic potential vorticity flux vector

$$\mathbf{J}^{\text{kin}} \equiv -\partial_t \mathbf{u} \wedge \hat{\mathbf{z}} = \mathbf{J} - \nabla \wedge (g \eta + \mathbf{u} \cdot \mathbf{u}/2). \quad (35.64)$$

Since the two fluxes differ by a curl; i.e., a rotational term, their divergences are identical

$$\nabla \cdot \mathbf{J}^{\text{kin}} = \nabla \cdot \mathbf{J}, \quad (35.65)$$

so that their convergence leads to the same evolution of $h Q$. Stated more formally, \mathbf{J}^{kin} and \mathbf{J} differ by a gauge, with the gauge function here given by $g \eta + \mathbf{u} \cdot \mathbf{u}/2$.

35.6 Example implications of PV material invariance

The material conservation properties of shallow water PV provide nontrivial constraints on the fluid motion by stating that f, h, ζ cannot change independently of the other. Rather, the combination $Q = (f + \zeta)/h$ must remain materially invariant (in the absence of friction and mixing). There are a huge variety of situations that induce changes in one or two of the terms, with the third term constrained to ensure Q remains unchanged. We here consider some thought experiments to garner experience with shallow water PV-thinking.

35.6.1 Topographic beta effect

Changes in the topography affect the potential vorticity by changing the thickness of a fluid column via (see Figure 31.1)

$$h = H + \Delta\eta - \Delta\eta_b, \quad (35.66)$$

with $H = \bar{h}$ the area mean layer thickness, $\Delta\eta = \eta - \bar{\eta}$ the deviation of the ocean surface from its area mean, and $\Delta\eta_b = \eta_b - \bar{\eta}_b$ the deviation of the ocean bottom from its area mean. For relative vorticity, we note that spatial changes in the topography act to drive a vertical velocity at the layer bottom as per equation (31.25),

$$w = \frac{D\eta_b}{Dt} \quad \text{at } z = \eta_b, \quad (35.67)$$

which then leads to vortex stretching and hence to a change in relative vorticity. In this subsection we highlight the analog between topographic slopes and planetary beta to thus motivate the term *topographic beta effect*.

To mathematically exhibit the topographic beta effect, consider a fluid column whose vorticity is dominated by planetary vorticity and with bottom topography having a small and linear slope in the meridional direction

$$\Delta\eta_b = \delta y, \quad (35.68)$$

where $|\delta| \ll 1$ is the topographic slope. Assuming the free surface undulations are small relative to the resting layer thickness, $\Delta\eta \ll H$, we can expand the potential vorticity according to

$$Q = \frac{f + \zeta}{h} \quad (35.69a)$$

$$= \frac{f_0 + \beta y + \zeta}{H + \Delta\eta - \delta y} \quad (35.69b)$$

$$\approx \frac{f_0 + \beta y + \zeta}{H} [1 - H^{-1}(\Delta\eta - \delta y)] \quad (35.69c)$$

$$\approx \frac{f_0 + \beta y + \zeta}{H} - \frac{f_0}{H^2} (\Delta\eta - \delta y). \quad (35.69d)$$

Setting $DQ/Dt = 0$ and rearranging leads to the material evolution of relative vorticity

$$\frac{D\zeta}{Dt} = -v(\beta + f_0\delta/H) + \frac{f_0 w(\eta)}{H}. \quad (35.70)$$

The second term on the right hand side is the vortex stretching associated with vertical motion at the top of the layer, where $w(\eta) = D\Delta\eta/Dt = D\eta/Dt$ according to the surface kinematic boundary condition (31.33). The first term on the right hand side is vortex stretching arising from both planetary beta and topographic slopes. It is written in a form revealing the parallels between these two contributions, and it is readily generalized to the following for arbitrary topography

$$\beta_{\text{eff}} = (H - \eta_b) \nabla[f/(H - \eta_b)] \approx \beta \hat{\mathbf{y}} + (f_0/H) \nabla\eta_b. \quad (35.71)$$

One of the more prominent roles for topographic beta is in supporting topographic Rossby waves, which are analogous to the Rossby waves supported by planetary beta (see Section 34.7). We also encounter the topographic beta effect in Section 40.6.1 in our study of the quasi-geostrophic shallow water model.

35.6.2 Planetary geostrophic PV

As introduced in Section 26.5, planetary geostrophy (PG) is used to study the large-scale laminar ocean circulation where relative vorticity is ignored. Furthermore, as shown in Section 40.4, the inviscid and adiabatic PG system materially preserves the PG potential vorticity, $Q = f/h$, so that

$$\frac{D(f/h)}{Dt} = 0. \quad (35.72)$$

Consequently, fluid particles respecting the inviscid PG equations follow contours of constant f/h . These contours are referred to as *geostrophic contours* since the flow is under geostrophic balance.

If we furthermore assume the ocean surface undulations are negligible (a useful assumption for large-scale flow), then shallow water columns follow trajectories of constant f/H , where $z = -H(x, y) = \eta_b(x, y)$ is the vertical position of the bottom topography. Figure 35.5 illustrates sample

f/H contours for a topographic seamount (bump) in the southern hemisphere. Fluid columns are steered towards the equator as they encounter the seamount. Those near to the bump are trapped as they circle around the seamount. There are hence two regimes of contours: those that are trapped by the seamount and cannot escape, and those that are untrapped.

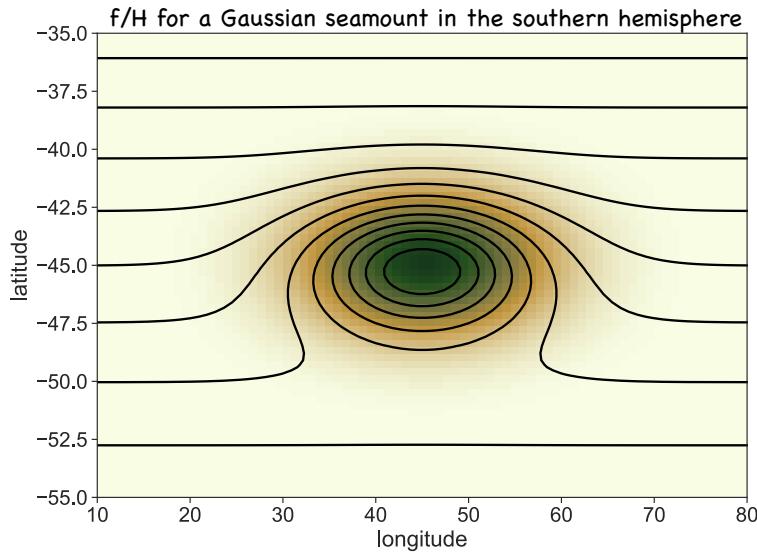


FIGURE 35.5: When a shallow water fluid is governed by the inviscid and adiabatic planetary geostrophic equations, fluid columns maintain fixed planetary geostrophic potential vorticity, $D(f/h)/Dt = 0$. Ignoring free surface undulations relative to changes in the bottom depth means that f/H remains fixed following the inviscid geostrophic flow, where $z = -H(x, y) = \eta_b(x, y)$ is the bottom topography. We illustrate these *geostrophic contours* for a Gaussian seamount (shaded feature) in the southern hemisphere. Contours of constant f/H that are far from the seamount follow lines of constant latitude, on which f is constant. Contours closer to the seamount are diverted according to the requirement of keeping f/H fixed. When H decreases, as per a seamount, then the f/H contours move equatorward. Trajectories near the seamount are trapped as they circle around the bump.

35.6.3 Spin up of converging flow

Consider the flow shown in Figure 35.6, whereby mass in the shallow water layer converges into a region. Just as described in the PV derivation Figure 35.4, increasing the column thickness, without substantially altering the planetary vorticity (e.g., f -plane), requires $\partial\zeta/\partial t > 0$ in order to maintain $Q = (\zeta + f)/h$ materially constant. Following our discussion of the rotating column in Section 35.3, note that convergence of mass reduces the moment of inertia relative to the center of the region. Angular momentum conservation requires the fluid to rotate faster thus picking up a positive relative vorticity. This dynamical process is embedded in the material invariance of PV. Finally, note that the opposite occurs in a region of diverging fluid, whereby the PV invariance implies that the relative vorticity has a negative tendency ($\partial\zeta/\partial t < 0$) (see also Figure 35.4).

35.6.4 Further study

Section 4.3 of Holton (1992) discusses the case of flow over topography where the full shallow water PV is materially invariant, $D(f + \zeta)/Dt = 0$. In that case there is a dramatic difference between easterly and westerly flows. In the northern hemisphere, westerly winds (eastward flow) deflects over the topography and downstream it undulates as topographic leewaves. A rotating tank offers

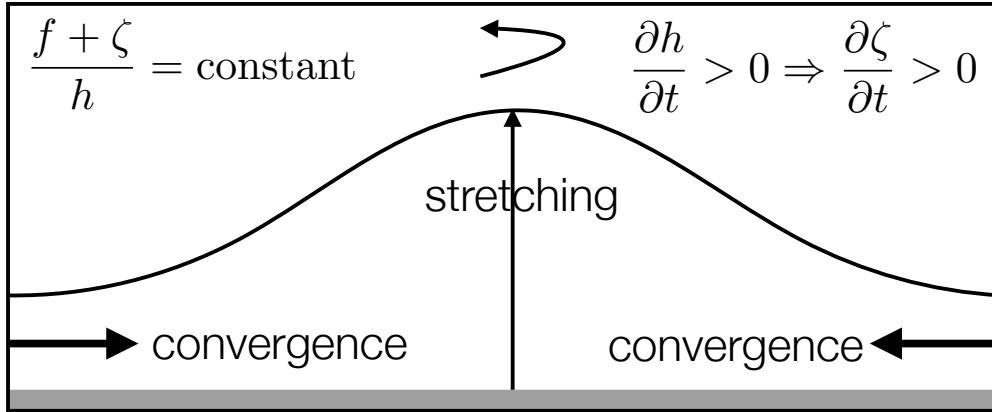


FIGURE 35.6: Illustrating the implications of material PV invariance for a shallow water fluid on an f -plane. If mass converges into a region, thus stretching the fluid column, then PV invariance implies the relative vorticity increases, $\partial\zeta/\partial t > 0$. This result is directly analogous to the rotating cylinder example considered in Figures 35.1 and 35.4. Namely, converging a region of constant mass reduces its moment of inertia so that angular momentum conservation leads to an increase in spin.

a useful controlled setting to observe leewaves, such as shown near the 20 minute mark in [this video from Prof. Fultz of the University of Chicago](#). Easterly winds (westward flow) do not exhibit a wavelike pattern, instead following a trajectory similar to the f/H contours of planetary geostrophic case, though modified by relative vorticity. In general, the study of flow near topography, either in the shallow water or continuously stratified, introduces a wealth of dynamical behaviors where material invariance of potential vorticity provides an important tool to help unravel mechanisms.

35.7 Geostrophic adjustment

The geostrophic balance presented in Sections 26.4 and 32.2 is very well maintained by the observed large-scale atmosphere and ocean. Hence, geostrophy (and the associated thermal wind) is a powerful diagnostic. In this section, we examine how a flow state that is initially not in geostrophic balance evolves towards geostrophy. We thus study the dynamical processes associated with the *geostrophic adjustment* problem.

A single shallow water layer on a flat f -plane is sufficient to introduce the main physical ideas. Furthermore, we focus on linear perturbations so that the governing equations are those derived in Section 32.8 when studying inertia-gravity waves. The adjustment consists of linear inertia-gravity waves that maintain a locally static potential vorticity (Section 32.8.4). For brevity in notation, we here drop all primes on the linear fluctuating terms.

35.7.1 Potential vorticity inversion

Before studying the geostrophic adjustment problem, we offer a few comments about *potential vorticity inversion*, which generally refers to the process of determining the flow field given information about the potential vorticity. In a shallow water layer, the potential vorticity is given by

$$Q = h^{-1} (f + \zeta) = h^{-1} [f + \partial_x v - \partial_y u]. \quad (35.73)$$

If we further assume the flow to be in geostrophic balance (Section 32.2), then

$$Q = \frac{f}{h} + \frac{1}{h} \left[\frac{\partial}{\partial x} \left(\frac{g}{f} \frac{\partial \eta}{\partial x} \right) + \frac{\partial}{\partial y} \left(\frac{g}{f} \frac{\partial \eta}{\partial y} \right) \right]. \quad (35.74)$$

Assuming we know Q throughout the domain; assuming f and Q are uniformly of the same sign within the domain; and assuming we know boundary conditions for η , then equation (35.74) is a nonlinear elliptic boundary value problem (Section 3.5) for η . Nonlinearities come from the $h^{-1} = (\eta - \eta_b)^{-1}$ pre-factor, as well as the boundary conditions discussed below. Linearizing by setting $h^{-1} \approx H^{-1}$ and simplifying the boundary conditions (see below) allows equation (35.74) to be solved for η . This solution process is referred to as *inverting* the elliptic operator, so that this particular inversion process is referred to as *potential vorticity inversion*.

General boundary conditions for η can be rather complex to handle mathematically. Namely, in a domain with a sloping bottom, such as in Figure 31.1, the free surface vanishes along the domain boundaries since the layer thickness vanishes there. Furthermore, the horizontal position of the domain boundary is a function of time since the layer moves up and down the sloping bottom. Vanishing layers and the associated moving boundaries are intrinsically nonlinear; i.e., there is no way to linearize this process. Instead, to facilitate the use of linear physics requires us to assume the layer thickness remains nonzero throughout the domain. Furthermore, we assume the layer thickness deviates only a small amount from the layer averaged thickness: $\eta/H \approx 1$. These assumptions are made in the following discussion of geostrophic adjustment.

35.7.2 Posing the initial value problem

We solve for the $t > 0$ evolution of surface height and velocity by making use of the linearized equations from Section 32.8

$$\frac{\partial \mathbf{u}}{\partial t} + f \hat{\mathbf{z}} \wedge \mathbf{u} = -g \nabla \eta \quad (35.75)$$

$$\frac{\partial \eta}{\partial t} + H \nabla \cdot \mathbf{u} = 0 \quad (35.76)$$

$$\zeta - \frac{f\eta}{H} = q(x, y) \quad (35.77)$$

where $q(x, y)$ is the static potential vorticity determined by the initial conditions (Section 32.8.4). To illustrate the geostrophic adjustment in an analytically tractable manner, consider the following step initial conditions for the surface height

$$\eta(x, t=0) = \begin{cases} +\eta_0 & x < 0 \\ -\eta_0 & x > 0, \end{cases} \quad (35.78)$$

which can be written

$$\eta(x, t=0) = \eta_0 [1 - 2\mathcal{H}(x)] = -\eta_0 \operatorname{sgn}(x), \quad (35.79)$$

where the sign-function is given by

$$\operatorname{sgn}(x) = \begin{cases} -1 & \text{if } x < 0 \\ 0 & \text{if } x = 0 \\ 1 & \text{if } x > 0, \end{cases} \quad (35.80)$$

which can also be written in terms of the Heaviside step function

$$\operatorname{sgn}(x) = 2\mathcal{H}(x) - 1, \quad (35.81)$$

where

$$\mathcal{H}(x) = \begin{cases} 0 & \text{if } x < 0 \\ 1/2 & \text{if } x = 0 \\ 1 & \text{if } x > 0. \end{cases} \quad (35.82)$$

The velocity is assumed to be zero initially

$$\mathbf{u}(x, y, t = 0) = 0. \quad (35.83)$$

Correspondingly, the initial relative vorticity vanishes so that the linearized potential vorticity (equation (32.176)) is

$$q(x, y) = \frac{f\eta_0}{H} \operatorname{sgn}(x). \quad (35.84)$$

Since $\partial q / \partial t = 0$ (equation (32.175)), this value of the potential vorticity is maintained at each point throughout the adjustment process. The velocity and surface height adjustment is thus constrained to keep potential vorticity static. This rather basic point is key to determining evolution of the velocity and surface height, and thus in determining the final (equilibrium) state for these fields.

35.7.3 Adjustment in the absence of rotation

In the absence of rotation ($f = 0$), relative vorticity is constant at each grid point. With a zero initial velocity, relative vorticity remains zero throughout the adjustment. The adjustment is thus quite simple. Namely, it consists of linear gravity waves (which have zero relative vorticity; see Section 32.8.2) propagating away from the initial step, converting the potential energy of the step into kinetic energy of waves that propagate to infinity. As the linear gravity waves are non-dispersive, they carry the initial pulse out to infinity without distortion in the form²

$$\eta(x, t) = -\frac{\eta_0}{2} [\operatorname{sgn}(x + ct) + \operatorname{sgn}(x - ct)], \quad (35.85)$$

where $c = \sqrt{gH}$ is the speed for non-rotating gravity waves (Section 32.8.3). The meridional velocity remains zero, whereas the zonal velocity

$$\frac{\partial u}{\partial t} = -g \frac{\partial \eta}{\partial x} \quad (35.86)$$

is given by

$$u(x, t) = \frac{g\eta_0}{2c} [\operatorname{sgn}(x + ct) - \operatorname{sgn}(x - ct)]. \quad (35.87)$$

After the transient waves have passed, the steady solution is a flat surface height with zero velocity. This steady solution is familiar from the case of a rock dropped into a still pond. After dropping the rock into the pond, the surface gravity waves radiate outward from the rock and are eventually damped upon reaching the shore. In equilibrium, the pond returns to a state of rest with a flat surface height.³

²In Section 3.7 we studied the wave solutions that take on the general form of equation (35.85).

³For some purposes, a pond is better studied using deep water equations rather than shallow water equations; see Section 30.1. Even so, the key physical points in this example are maintained in either case.

35.7.4 Adjustment with rotation

With rotation, the transient solution consists of inertia-gravity waves (Section 32.8) that transmit information about the initial surface height perturbation out to infinity. After the transient waves have passed, the steady solution is either the trivial solution with flat surface height (as for the non-rotating case), or a nontrivial solution that is in geostrophic balance

$$f\hat{z} \wedge \mathbf{u} = -g \nabla \eta \quad \text{and} \quad \nabla \cdot \mathbf{u} = 0 \quad \text{and} \quad q = H^{-1} f \eta_0 \operatorname{sgn}(x). \quad (35.88)$$

Conservation of potential vorticity constrains the solution so that the steady state surface height is sloped according to a geostrophically balanced state. That is, an equilibrium state of no-motion is not allowed by potential vorticity conservation. This is a profound distinction from the non-rotating case.

Computing the equilibrium state

As the flow is geostrophic on an f -plane, we make use of the geostrophic streamfunction

$$\psi = g \eta / f. \quad (35.89)$$

The equilibrium state is written in terms of the streamfunction according to

$$u = -\partial_y \psi \quad \text{and} \quad v = \partial_x \psi \quad \text{and} \quad \zeta = \nabla^2 \psi. \quad (35.90)$$

Making use of these expressions in the linearized potential vorticity expression (35.84) leads to the elliptic partial differential equation for the streamfunction

$$[\nabla^2 - L_d^{-2}] \psi = q(x, y), \quad (35.91)$$

where we introduced the shallow water *deformation radius*

$$L_d = f^{-1} \sqrt{g H}. \quad (35.92)$$

We motivate the name “deformation radius” in the following.

The initial condition (35.79) has no y -dependence. Furthermore, there is nothing in the adjustment process that will break this meridional symmetry. Hence, the equilibrium state is a function only of x , in which case the streamfunction satisfies the ordinary differential equation

$$\frac{d^2 \psi}{dx^2} - L_d^{-2} \psi = \frac{f \eta_0}{H} \operatorname{sgn}(x). \quad (35.93)$$

We solve this equation separately for $x > 0$ and $x < 0$ and then match the function and its first derivative at $x = 0$. Furthermore, we constrain the streamfunction to vanish at $\pm\infty$. The $x > 0$ streamfunction satisfies

$$\frac{d^2 \psi}{dx^2} - L_d^{-2} \psi = \frac{f \eta_0}{H}. \quad (35.94)$$

The particular solution is

$$\psi_p = -L_d^2 q = -L_d^2 \frac{f \eta_0}{H} = -\frac{g \eta_0}{f} \quad (35.95)$$

and the homogeneous solution is

$$\psi_h = \frac{g \eta_0}{f} e^{-x/L_d} \quad (35.96)$$

so that

$$\psi = -\frac{g \eta_0}{f} \left[1 - e^{-x/L_d} \right]. \quad (35.97)$$

The $x < 0$ solution is found similarly, so that

$$\psi = \frac{g \eta_0}{f} \begin{cases} -\left(1 - e^{-x/L_d}\right) & x > 0 \\ \left(1 - e^{x/L_d}\right) & x < 0, \end{cases} \quad (35.98)$$

which means that the equilibrium surface height is

$$\eta = \eta_0 \begin{cases} -\left(1 - e^{-x/L_d}\right) & x > 0 \\ \left(1 - e^{x/L_d}\right) & x < 0. \end{cases} \quad (35.99)$$

Note that the streamfunction vanishes at $x = 0$ and has a first derivative of $-\eta_0 \sqrt{gH}/H$. Since the streamfunction only has a zonal dependence, the equilibrium velocity is purely meridional

$$u = 0 \quad \text{and} \quad v = -\frac{g \eta_0}{f L_d} e^{-|x|/L_d}. \quad (35.100)$$

The equilibrium velocity thus consists of a jet that is perpendicular to the surface height front.

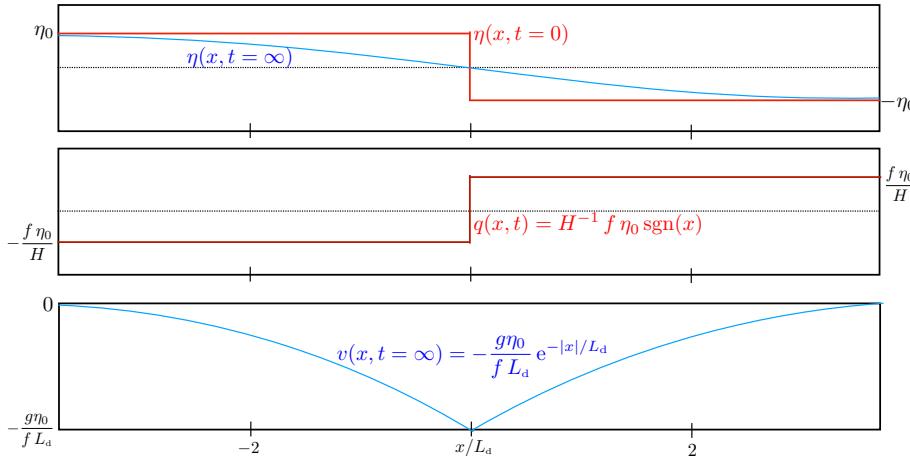


FIGURE 35.7: Depicting solutions to the linear geostrophic adjustment of a rotating shallow water layer on an f -plane. The top panel shows the initial (step-function) surface height (35.79) and the equilibrium (exponential) surface height (35.99). The second panel shows the static (step-function) potential vorticity (35.88). The third panel shows the equilibrium (exponential) meridional velocity (35.100) comprised of a jet centered at $x = 0$. The horizontal axis is scaled according to the shallow water deformation radius, $L_d = f^{-1} \sqrt{gH}$. This figure is adapted from figure 3.10 of [Vallis \(2017\)](#).

35.7.5 Comments

As illustrated in Figure 35.7, the equilibrium profiles for the surface height and velocity both have an exponential decay, with decay length scale given by the deformation radius. It is this length scale over which the solution is affected or “deformed” by rotation, thus motivating the name “deformation radius”.

A key feature of the rotating case is that some of the potential energy contained within the initial perturbed surface height remains in the equilibrium geostrophic flow. The conservation of potential vorticity constrains the flow so that all of the initial potential energy cannot be converted

to kinetic energy. Rather, the adjustment occurs only within a deformation radius distance from the initial perturbation.

Chapter 3 of [Pratt and Whitehead \(2008\)](#) provide a thorough discussion of the geostrophic adjustment problem, both for the linear case and fully nonlinear case.

35.8 Circulation with non-conservative processes

We follow the discussion in Section 35.5 to study the evolution of circulation in the presence of non-conservative processes such as dia-surface transport and boundary stresses (Section 31.5). For this purpose, consider the velocity circulation around a closed horizontal area, \mathcal{S} (see Figure 35.8)

$$\mathcal{C} = \int_{\mathcal{S}} \zeta \, d\mathcal{S} = \oint_{\partial\mathcal{S}} \mathbf{u} \cdot \hat{\mathbf{t}} \, d\ell. \quad (35.101)$$

In this equation, $\hat{\mathbf{t}} \, d\ell$ is the vector line increment around the contour, and $\hat{\mathbf{t}}$ is the tangent vector orienting the contour integral in a counterclockwise direction. We assume the circulation contour extends vertically through the layer so that the outward normal, $\hat{\mathbf{n}}$, to the contour is strictly horizontal, as is the tangent vector, $\hat{\mathbf{t}}$. We now seek an evolution equation for this circulation by making use of the vorticity equation (35.60).

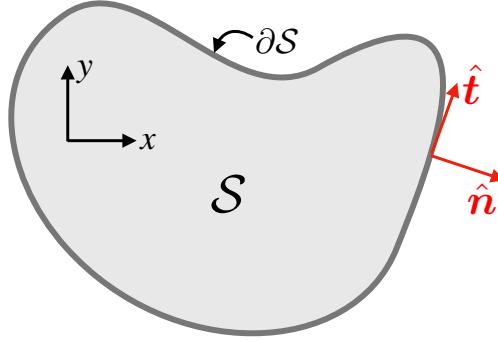


FIGURE 35.8: Illustrating the calculation of circulation around the contour, $\partial\mathcal{S}$, of a static horizontal area, \mathcal{S} , within a layer of shallow water fluid. The circulation theorem (35.105) provides the means to compute the time changes in circulation as a function of the advection of absolute vorticity crossing the contour plus the circulation of friction. Note that the contour has a vertical extent throughout the shallow water layer, so that the unit outward normal, $\hat{\mathbf{n}}$, and the unit tangent direction, $\hat{\mathbf{t}}$, are both horizontal vectors.

To develop an evolution equation for \mathcal{C} , integrate the vorticity equation (35.60) over the area \mathcal{S} to yield

$$\int_{\mathcal{S}} \frac{\partial \zeta}{\partial t} \, d\mathcal{S} = \int_{\mathcal{S}} [-\nabla \cdot (\zeta_a \mathbf{u} + \hat{\mathbf{z}} \wedge \mathbf{F})] \, d\mathcal{S} = - \oint_{\partial\mathcal{S}} [\zeta_a \mathbf{u} + \hat{\mathbf{z}} \wedge \mathbf{F}] \cdot \hat{\mathbf{n}} \, d\ell, \quad (35.102)$$

where \oint symbolizes an integral around the contour. To reach this equation we used the divergence theorem on the right hand side, with $\hat{\mathbf{n}}$ the *horizontal* outward normal vector along the contour, $\partial\mathcal{S}$, and $d\ell$ is the line increment along the contour. We also set $\partial f / \partial t = 0$ as part of the time derivative of the absolute vorticity. The non-conservative forcing term can be written

$$(\hat{\mathbf{z}} \wedge \mathbf{F}) \cdot \hat{\mathbf{n}} = (\hat{\mathbf{n}} \wedge \hat{\mathbf{z}}) \cdot \mathbf{F} = -\hat{\mathbf{t}} \cdot \mathbf{F}, \quad (35.103)$$

thus leading to

$$\int_{\mathcal{S}} \frac{\partial \zeta}{\partial t} d\mathcal{S} = - \oint_{\partial \mathcal{S}} \zeta_a \mathbf{u} \cdot \hat{\mathbf{n}} d\ell + \oint_{\partial \mathcal{S}} \mathbf{F} \cdot \hat{\mathbf{t}} d\ell, \quad (35.104)$$

where \oint is the counter-clockwise oriented closed contour integral. We next assume the area \mathcal{S} is constant in time, so that the Eulerian time derivative can be pulled across the integral to render

$$\frac{dC}{dt} = - \oint_{\partial \mathcal{S}} \zeta_a \mathbf{u} \cdot \hat{\mathbf{n}} d\ell + \oint_{\partial \mathcal{S}} \mathbf{F} \cdot \hat{\mathbf{t}} d\ell. \quad (35.105)$$

The first term on the right hand side arises from the horizontal advection of absolute vorticity across the contour. This term is not oriented and so the integral sign has no arrow. The second term arises from the counter-clockwise oriented circulation of any non-conservative accelerations. We refer to equation (35.105) as a *circulation theorem*. It has many uses under specific cases, some of which are described in the remainder of this chapter.

35.8.1 Circulation around a closed streamline in steady flow

As a particular example of the circulation theorem (35.105), consider a steady flow in the absence of boundary volume sources ($w^{(\dot{\eta})} = 0$). In this case, the thickness equation (31.16) reduces to $\nabla \cdot (h \mathbf{u}) = 0$ so that we can introduce a streamfunction, Ψ (with dimensions of $L^3 T^{-1}$)

$$h \mathbf{u} = \hat{\mathbf{z}} \wedge \nabla \Psi. \quad (35.106)$$

In a bounded domain, the contours of constant Ψ , the streamlines, are closed. Furthermore, the outward normal to a closed streamline is perpendicular to the velocity, $\mathbf{u} \cdot \hat{\mathbf{n}} = 0$. We thus see that the steady state form of the circulation theorem (35.105), computed around a closed streamline, leads to the following constraint on the non-conservative forces

$$\oint_{\text{streamline}} \mathbf{F} \cdot \hat{\mathbf{t}} d\ell = \int_{\text{streamline area}} (\nabla \wedge \mathbf{F}) \cdot \hat{\mathbf{z}} d\mathcal{S} = 0, \quad (35.107)$$

where the second equality follows from Stokes' theorem applied over the area bounded by the streamline. Equation (35.107) provides a constraint on the non-conservative forcing that must be satisfied to enable a steady flow. If this constraint is not satisfied, then the flow cannot reach a steady state. Although we may not know explicit details of the streamlines, we can still make use of this constraint if we assume the flow is steady. The analysis in Section 35.9 offers an example application of these ideas for studies of circulation in steady ocean gyres.

35.8.2 Circulation from wind stress and Rayleigh drag

A particularly simple form for the non-conservative acceleration is given by

$$\mathbf{F} = -\gamma \mathbf{u} + \boldsymbol{\tau}^{\text{wind}} / (h \rho) \equiv -\gamma \mathbf{u} + \mathbf{F}^{\text{wind}}. \quad (35.108)$$

The first term is referred to as *Rayleigh drag* with $\gamma > 0$ a constant with dimensions of inverse time.⁴ Rayleigh drag damps all flow to rest with γ^{-1} the e-folding time for the damping. The second term in equation (35.108) is the acceleration on the layer from wind stress, $\mathbf{F}^{\text{wind}} = \boldsymbol{\tau}^{\text{wind}} / (h \rho)$, with this form following from our discussion in Section 31.5.2. Namely, the homogeneity of the shallow water layer renders the contact stress from winds into a body stress applied throughout the layer.

⁴See Section 28.2.4 for more discussion of Rayleigh drag in the context of Ekman mechanics.

Plugging the acceleration (35.108) into the time dependent circulation theorem (35.105) leads to

$$\frac{dC}{dt} + \gamma C = - \int_{\partial S} \zeta_a \mathbf{u} \cdot \hat{\mathbf{n}} d\ell + \oint_{\partial S} \mathbf{F}^{\text{wind}} \cdot \hat{\mathbf{t}} d\ell. \quad (35.109)$$

Specializing to a steady state and choosing the contour as a closed streamline (along which $\mathbf{u} \cdot \hat{\mathbf{n}} = 0$), renders

$$C = \gamma^{-1} \oint_{\text{streamline}} \mathbf{F}^{\text{wind}} \cdot \hat{\mathbf{t}} d\ell = (\gamma \rho)^{-1} \oint_{\text{streamline}} (\boldsymbol{\tau}^{\text{wind}}/h) \cdot \hat{\mathbf{t}} d\ell. \quad (35.110)$$

This equation says that velocity circulation around a closed streamline is determined by wind stress circulation around that streamline plus knowledge of the Rayleigh drag damping time scale γ^{-1} . This result supports our intuition that the steady circulation around a closed contour is oriented according to the applied wind stress.

35.9 Western intensification of ocean gyres

Ocean gyres are a prominent feature of ocean circulation, with the North Atlantic and North Pacific middle-latitude gyres two canonical examples. A characteristic feature of such gyres is that they are not symmetric in the east-west direction, with a prominent western side where flow is stronger than the more sluggish interior flow. As shown in this section, such asymmetry is not a response to the wind forcing, with the asymmetry found even without any zonal variations in the winds. Instead, it is a manifestation of the beta effect present for flow on a spherical planet (or idealized as the β -plane from Section 23.2). We discuss the beta effect more formally in Section 36.7, but for now it is sufficient to know that it arises from the meridional gradient of the planetary vorticity: $\beta = \partial_y f$.

We have the basic tools in hand to understand the physical balances leading to western intensification in steady ocean gyres. For that purpose, we make use of vorticity balances as they offer a more direct path towards western intensification than the linear momentum or axial angular momentum balances used to explore channel flow in Sections 20.4 and 32.6.

35.9.1 Steady and large-scale vorticity balance

The steady circulation theorem (35.110) holds regardless the bottom topography or surface height undulations. Again, it says that circulation around a closed streamline is in the same sense as the wind circulation. However, we need more information to see how western intensification emerges as a property of ocean gyres. For that purpose, consider the steady absolute vorticity equation (35.60), again in the presence of wind forcing and Rayleigh drag

$$\nabla \cdot (\mathbf{u} \zeta_a) = -\gamma \zeta + \hat{\mathbf{z}} \cdot (\nabla \wedge \mathbf{F}^{\text{wind}}). \quad (35.111)$$

Introducing the shallow water potential vorticity, $Q = \zeta_a/h$, allows us to write

$$\nabla \cdot (\mathbf{u} \zeta_a) = \nabla \cdot (h \mathbf{u} Q). \quad (35.112)$$

The steady state thickness equation (31.16) means that $\nabla \cdot (h \mathbf{u}) = 0$, so that the vorticity equation (35.111) takes the form

$$h \mathbf{u} \cdot \nabla Q = -\gamma \zeta + \hat{\mathbf{z}} \cdot (\nabla \wedge \mathbf{F}^{\text{wind}}). \quad (35.113)$$

This equation says that in the absence of the irreversible terms on the right hand side, the steady horizontal flow follows contours of constant potential vorticity. This result follows directly from the material invariance of PV in the absence of non-conservative processes. However, in the presence of Rayleigh drag and/or wind stress curl, PV is modified when following the flow so that $\mathbf{u} \cdot \nabla Q \neq 0$.

35.9.2 Planetary geostrophic flow and the Sverdrup balance

For large-scale flow away from lateral boundaries, the flow has an absolute vorticity that is dominated by planetary vorticity so that

$$Q \approx Q_{\text{PG}} = f/h, \quad (35.114)$$

which is the PV for planetary geostrophic flow (Section 26.5 and 40.4). Correspondingly, we assume the Rayleigh drag term is negligible in this region since the relative vorticity is small, so that the PV equation (35.113) takes the form

$$h \mathbf{u} \cdot \nabla Q_{\text{PG}} = \hat{\mathbf{z}} \cdot (\nabla \wedge \mathbf{F}^{\text{wind}}). \quad (35.115)$$

Expanding the left hand side and introducing the planetary vorticity gradient renders the *shallow water Sverdrup balance*

$$\beta v = Q_{\text{PG}} \mathbf{u} \cdot \nabla h + \hat{\mathbf{z}} \cdot (\nabla \wedge \mathbf{F}^{\text{wind}}) \quad \text{shallow water Sverdrup balance.} \quad (35.116)$$

This balance states how horizontal advection of layer thickness (first right hand side term) plus the wind stress curl (second term) lead to meridional motion for flow on a rotating sphere (beta effect on left hand side). Gradients in the layer thickness arise from free surface undulations as well as gradients in the bottom topography (see Figure 31.1). The traditional *Sverdrup balance* arises when we assume the flow takes place over a flat bottom and the free surface undulations are negligible, in which case $\nabla h \approx 0$ so that

$$\beta v = \hat{\mathbf{z}} \cdot (\nabla \wedge \mathbf{F}^{\text{wind}}) \quad \text{traditional Sverdrup balance.} \quad (35.117)$$

35.9.3 Sverdrup flow in a closed domain with anti-cyclonic wind stress

As a particular example, consider a closed northern hemisphere middle latitude β -plane domain driven by an anti-cyclonic wind stress

$$\hat{\mathbf{z}} \cdot (\nabla \wedge \mathbf{F}^{\text{wind}}) < 0 \quad \text{northern hemisphere anti-cyclonic wind stress.} \quad (35.118)$$

This situation is depicted in Figure 35.9, where we also illustrate a commonly used wind stress profile that is purely zonal and has a co-sinusoidal meridional structure that is symmetric about the central latitude of the domain

$$\mathbf{F}^{\text{wind}} = -\hat{\mathbf{x}} A \cos[\pi(y - y_0 + L/2)/L] \quad (35.119a)$$

$$\hat{\mathbf{z}} \cdot (\nabla \wedge \mathbf{F}^{\text{wind}}) = -(\pi A/L) \sin[\pi(y - y_0 + L/2)/L]. \quad (35.119b)$$

In these equations, $A > 0$ is the magnitude of the wind stress acceleration applied to the layer, and the domain extends meridionally from $y_0 - L/2 \leq y \leq y_0 + L/2$ with $y = y_0$ the central latitude. This wind stress has westerlies on the poleward side of the domain and easterlies (trade winds) on the equatorward side so that $\hat{\mathbf{z}} \cdot (\nabla \wedge \boldsymbol{\tau}) < 0$ throughout the domain.

The Sverdrup balance (35.117) indicates that an anti-cyclonic wind stress curl drives an equatorward Sverdrup flow. We emphasize that this flow is *not* the result of meridional winds pushing the fluid to the south. Instead, it arises in response to the constraints of vorticity balance with an anti-cyclonic wind stress curl in the presence of the beta effect. Indeed, for the idealized wind stress (35.119a) there is no meridional wind component. Although water satisfying Sverdrup balance flows south, all the water in the domain cannot be moving to the south. Rather, volume conservation requires a poleward return flow somewhere outside the region of Sverdrup balance.

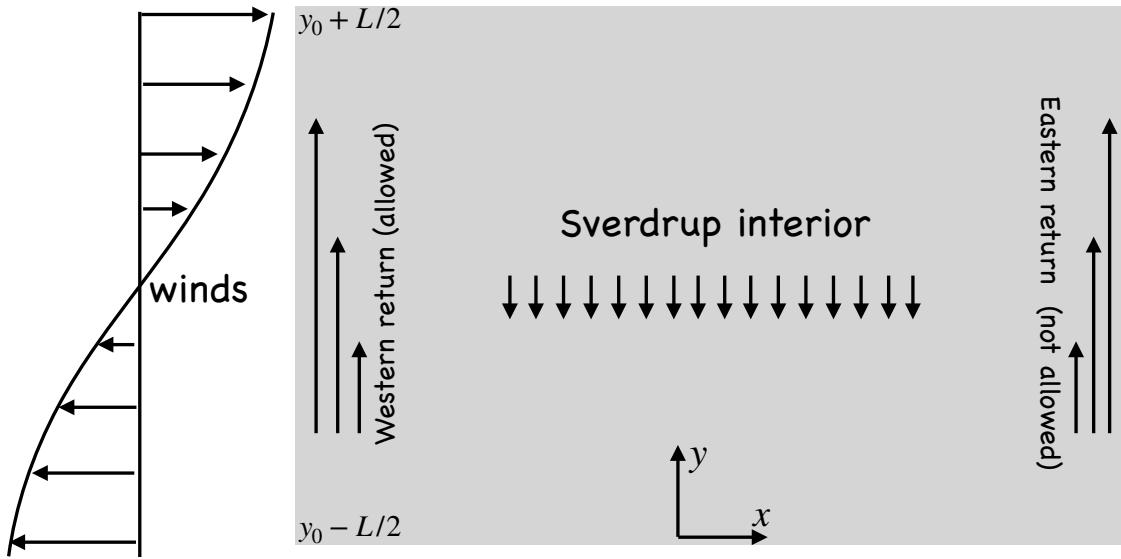


FIGURE 35.9: Illustrating the southward Sverdrup flow in response to an anti-cyclonic wind stress forcing in a bounded northern hemisphere domain. A northward return flow is required to satisfy volume conservation. As seen in Section 35.9.4, a linear vorticity balance between beta, winds, and friction lead to a western boundary return flow and corresponding western intensification.

35.9.4 Western intensification and the role of beta

Volume conservation is a kinematic constraint that requires a return flow on either the eastern or western side of the domain, outside the region of Sverdrup balance. But what side? We offer the following arguments for the western side.

Inertial entry into a boundary layer region

Recall the discussion of Figure 34.4 where we considered how inviscid flow in a homogeneous fluid over a flat bottom region materially preserves absolute vorticity in the presence of a meridional barrier. To materially preserve absolute vorticity (and ignoring free surface undulations relative to the depth of the fluid), the flow can deviate meridionally, either northward or southward, when encountering a western wall. In contrast, such meridional deviation is prohibited for absolute vorticity preserving flow that encounters an eastern wall. So in referring to Figure 35.9, southward flow can make a turn westward towards the western boundary, enter the boundary layer, and move northward within the boundary layer. It cannot do so for the eastern side.

Steady velocity circulation follows wind circulation

The steady circulation theorem (35.110) means that the flow circulation is in the same sense as the wind circulation. With an anti-cyclonic wind forcing, the circulation also must be anti-cyclonic. This result again leads us to conclude that the western boundary is the region for the return flow, in which case the full gyre circulation will be anti-cyclonic.

Steady vorticity balance and the need for dissipation

The two previous arguments point to the western side as the return flow region. But we have yet to provide an argument for western intensification. For that purpose, recall that the return flow region

is not in traditional Sverdrup balance. To see what terms can break that balance, consider again the steady vorticity balance (35.113). Continuing to assume a flat bottom and rigid lid surface leads to

$$\beta v = -\gamma \zeta + \hat{z} \cdot (\nabla \wedge \mathbf{F}^{\text{wind}}), \quad (35.120)$$

where we dropped the nonlinear advection term $\mathbf{u} \cdot \nabla \zeta$ since we wish to determine whether a linear balance can give rise to western intensification. The Rayleigh drag on the right hand side breaks the Sverdrup balance in regions where the relative vorticity is nontrivial. Since we know there must be a return flow somewhere in the domain, we know there must be a region where dissipation is sufficiently strong to break Sverdrup balance.

What is required for the steady and linear dissipative vorticity balance (35.120) to be maintained in the northward return flow region? To answer this question, expose the signs on the terms in equation (35.120)

$$\underbrace{\beta v}_{\text{positive}} + \underbrace{\gamma \zeta}_{\text{unspecified}} = \underbrace{\hat{z} \cdot (\nabla \wedge \mathbf{F}^{\text{wind}})}_{\text{negative}}. \quad (35.121)$$

We have $\beta v > 0$ since we are concerned with the region of northward return flow, and $\hat{z} \cdot (\nabla \wedge \mathbf{F}^{\text{wind}}) < 0$ by assumption of anti-cyclonic wind stress over the full domain. Hence, for the balance (35.121) to be realized requires $\zeta < 0$, with the value large enough to balance both the winds and the meridional advection of planetary vorticity

$$\underbrace{\gamma \zeta}_{\text{negative}} = \underbrace{-\beta v}_{\text{negative}} + \underbrace{\hat{z} \cdot (\nabla \wedge \mathbf{F}^{\text{wind}})}_{\text{negative}}. \quad (35.122)$$

For anti-cyclonic gyre flow, as required by the circulation condition (35.110), the only way to realize $\zeta < 0$ of sufficient magnitude is to have an intensified flow along the western side of the gyre. In this region, $\zeta \approx \partial v / \partial x < 0$ can become sufficiently large in magnitude. Furthermore, since the wind stress is applied throughout the domain, all streamlines feel the winds and must pass through the western boundary region where vorticity is enhanced and Rayleigh drag is able to balance the winds and planetary advection.

The importance of beta

In the absence of beta, there would be no interior region in Sverdrup balance driving southward flow. The flow would thus only be subject to the circulation condition (35.110) whereby linear flow can symmetrically dissipate the wind stress. Hence, the beta effect is the fundamental element that causes poleward flow to intensify along the western side of the gyre in response to the equatorward interior flow.

35.9.5 A role for topographic form stress

Recall the shallow water Sverdrup balance (35.116), here with the addition of Rayleigh drag

$$\beta v = Q_{\text{PG}} \mathbf{u} \cdot \nabla h - \gamma \zeta + \hat{z} \cdot (\nabla \wedge \mathbf{F}^{\text{wind}}). \quad (35.123)$$

For the right hand side, we have thus far considered a flat bottom and ignored free surface undulations, in which case $\nabla h = 0$. We here sketch the situation with nonzero free surface and topographic variations, thus illustrating how boundary pressure gradients can render flow even in the absence of bottom drag.

Curl of the form stresses

We are just concerned with that portion of $\mathbf{u} \cdot \nabla h$ due to pressure gradients, in which case we insert the geostrophic flow

$$\mathbf{u} \cdot \nabla h = (g/f) \nabla h \cdot (\hat{\mathbf{z}} \wedge \nabla \eta) = (g/f) \hat{\mathbf{z}} \cdot (\nabla \eta \wedge \nabla h). \quad (35.124)$$

Making use of the hydrostatic relation $p_b = p_a + \rho g h$, and the expression for the layer thickness, $h = \eta - \eta_b$, leads to

$$\mathbf{u} \cdot \nabla h = (g/f) \hat{\mathbf{z}} \cdot (\nabla \eta \wedge \nabla h) \quad (35.125a)$$

$$= 1/(\rho f) \hat{\mathbf{z}} \cdot [\nabla \wedge (p_a \nabla \eta) + \nabla \eta \wedge \nabla p_b] \quad (35.125b)$$

$$= 1/(\rho f) \hat{\mathbf{z}} \cdot [\nabla \wedge (p_a \nabla \eta) - \nabla \wedge (p_b \nabla \eta_b) + \nabla h \wedge \nabla p_b] \quad (35.125c)$$

$$= 1/(\rho f) \hat{\mathbf{z}} \cdot [\nabla \wedge (p_a \nabla \eta) - \nabla \wedge (p_b \nabla \eta_b) + \nabla p_b \wedge \nabla p_a / (\rho g)]. \quad (35.125d)$$

The first and second terms are the curls of the pressure form stresses applied to the surface (atmospheric) and bottom boundaries (Section 20.2), respectively, which we refer to as the atmospheric and bottom pressure torques. The third term is a torque arising from the misalignment of the applied pressure and bottom pressure. Making use of this result in the vorticity balance (35.123) leads to

$$\beta \rho h v = \hat{\mathbf{z}} \cdot [\nabla \wedge (p_a \nabla \eta) - \nabla \wedge (p_b \nabla \eta_b) + \nabla p_b \wedge \nabla p_a / (\rho g)] + \rho h [-\gamma \zeta + \hat{\mathbf{z}} \cdot (\nabla \wedge \mathbf{F}^{\text{wind}})]. \quad (35.126)$$

We consider a few special cases to see how the pressure torques affect the meridional flow in a gyre circulation, with a focus on bottom pressure torque since this term is generally far larger than those involving the atmospheric pressure.

Vanishing topographic form stress curl for geostrophic f -plane motion

In the absence of friction, wind forcing, and with a uniform atmospheric pressure, β times the depth integrated meridional mass transport in equation (35.126) is balanced by the curl of the topographic form stress

$$\beta \rho h v = -\hat{\mathbf{z}} \cdot (\nabla \wedge p_b \nabla \eta_b) \quad \text{linear inviscid and unforced.} \quad (35.127)$$

As a corollary of this balance, we see that a linear inviscid and unforced geostrophic flow on an f -plane satisfies $\hat{\mathbf{z}} \cdot (\nabla \wedge p_b \nabla \eta_b) = \hat{\mathbf{z}} \cdot (\nabla p_b \wedge \nabla \eta_b) = 0$, so that the isolines of surface height align with isolines of bottom topography since (with $\nabla p_a = 0$) $p_b = g \rho h = g \rho (\eta - \eta_b)$.

Vanishing integrated topographic form stress curl on closed isobaths

Consider a closed domain comprised of topography with isobaths closed within the domain. Use of Stokes' theorem reveals that an integral of $\hat{\mathbf{z}} \cdot (\nabla \wedge p_b \nabla \eta_b)$ over the area defined by an isobath vanishes

$$\int_S \hat{\mathbf{z}} \cdot (\nabla \wedge p_b \nabla \eta_b) dS = \oint_{\partial S} p_b \nabla \eta_b \cdot \hat{\mathbf{t}} d\ell = 0, \quad (35.128)$$

where ∂S is determined by an isobath so that $\nabla \eta_b \cdot \hat{\mathbf{t}} = 0$. We can employ the same argument for an area bounded by isobars of bottom pressure. Hence, if there is initially no velocity circulation around an isobar or isobath, then topographic form stress will not spontaneously generate a circulation. This result is sensible because topographic form stress is *not* an external force akin to wind stress. Whereas wind stress curl is not constrained to vanish when integrated over a closed domain, topographic form stress, as it arises only in the presence of flow, is highly constrained.

Local generation of bottom pressure torque

The beta effect provides an inviscid means to create misalignment between the surface height and the bottom topography, with misalignment required to generate a nonzero topographic form stress curl, which we refer to as *bottom pressure torque*. We discussed an analogous misalignment in Section 32.6 when studying the force balances in a steady zonally re-entrant channel with bottom topography. Wind stress forcing and dissipation offer another means for (η, η_b) -misalignment, as seen merely by rewriting the vorticity balance (35.126) as an expression for the bottom pressure torque

$$\hat{z} \cdot (\nabla \wedge p_b \nabla \eta_b) = -\beta \rho h v + \hat{z} \cdot [\nabla \wedge (p_a \nabla \eta) + \nabla p_b \wedge \nabla p_a / (\rho g)] + \rho h [-\gamma \zeta + \hat{z} \cdot (\nabla \wedge \mathbf{F}^{\text{wind}})]. \quad (35.129)$$

In our discussion of western intensification in Section 35.9.4, we ignored the role of bottom pressure torque. However, as seen by this balance, bottom pressure torque plays a role when topography and surface height are misaligned, with that role in some locations more important than friction (see [Hughes and de Cueves \(2001\)](#), [Jackson et al. \(2006\)](#) and [Patmore et al. \(2019\)](#) for examples). Even so, as shown above, the integral of the topographic form stress vanishes over a closed domain bounded by isobaths or isobars, so that friction is ultimately required to balance the input of vorticity from the winds. We contrast this situation to the zonally re-entrant channel from Section 32.6, whereby topographic form stress, of sufficient strength, can balance the momentum input by the winds thus making friction less important for the channel flow.

35.9.6 Comments and further study

The study of ocean gyres originates from the pioneering work of Henry Stommel in the 1950s. This video [from SciencePrimer](#) provides a concise summary of the dynamics of ocean gyres and western boundary intensification due to the beta effect. Chapter 19 of [Vallis \(2017\)](#) provides a lucid treatment of ocean gyre dynamics by working through the key features of Stommel's model as well as variants such as that from Walter Munk who considered a viscous closure rather than Rayleigh drag. The Stommel model and its variants are themselves very idealized renditions of the ocean gyres occurring in Nature. Notable further factors become important in studying Nature's gyres, such as topography (briefly discussed in Section 35.9.5), nonlinearities and instabilities, turbulent boundary layers, and coupled air-sea processes. Each of these processes render the study of western boundary currents one of the most complex and timeless areas of physical oceanography.

For the studies of Stommel and Munk, friction plays a key role in the western intensification of wind driven ocean gyres. As described in Chapter 19 of [Vallis \(2017\)](#), there have been attempts to produce an inviscid (and unforced) gyre solution, with the study from Fofonoff of note. We consider elements of the Fofonoff gyre in Exercise 34.7. Additionally, we raised the importance of sloping sides in Section 35.9.5 and in the earlier discussion of axial angular momentum in Section 20.4. Sloping sides enable bottom topographic form stress and bottom pressure torques to dominate over bottom turbulent stresses and turbulent torques. In so doing, sloping sides allow for an inviscid closure of gyres much like that for zonally reentrant channels. The role of sloping sides for gyre circulations was emphasized by [Hughes and de Cueves \(2001\)](#). As Nature eschews vertical sidewalls, the balances of [Hughes and de Cueves \(2001\)](#) are more geophysically compelling than the frictional closures required for vertical sidewalls.

35.10 Column vorticity

Throughout this chapter we have focused on vorticity as defined by the curl of the horizontal velocity, $\zeta = \hat{z} \cdot (\nabla \wedge \mathbf{u})$. Since flow within a shallow water layer moves in coherent and extensible vertical columns (Section 31.2), the shallow water vorticity measures the spin of a shallow water column. However, we can also measure the column spin by considering the curl of the thickness weighted velocity

$$\Sigma = \hat{z} \cdot \nabla \wedge (h \mathbf{u}) = h \zeta + \hat{z} \cdot \nabla h \wedge \mathbf{u}. \quad (35.130)$$

We refer to Σ as the *column vorticity*. In addition to the thickness weighted relative vorticity, Σ measures the misalignment between the layer thickness gradient and the velocity, with $\hat{z} \cdot \nabla h \wedge \mathbf{u}$ generally nonzero especially for geostrophic flows. Hence, the second term offers an extra measure of the spin for the fluid column beyond the relative vorticity measure.⁵ As we will see, the layer thickness factor ensures that Σ directly probes stresses acting at the layer interfaces, as well as stresses within the fluid layer. Besides its intrinsic interest, this analysis is motivated by studies of the depth integrated flow commonly pursued in oceanography, with particular examples provided in Chapter 41.

35.10.1 Formulating the column vorticity equation

To formulate the dynamical equation for Σ , we start from the momentum equation (32.30)

$$\frac{\partial(h \mathbf{u})}{\partial t} + \nabla \cdot (h \mathbf{u} \otimes \mathbf{u}) + f \hat{z} \wedge (h \mathbf{u}) = -g h \nabla \eta^{\text{eff}} + \mathbf{f}^{\text{nc}} \quad (35.131)$$

where \mathbf{f}^{nc} is the thickness weighted non-conservative accelerations, such as arise from horizontal strains in the presence of viscosity, and the boundary transfer of turbulent momentum such as through winds and bottom drag (see Section 31.5). Taking the vertically projected curl of equation (35.131) leads to

$$\partial_t \Sigma + \hat{z} \cdot \nabla \wedge [\nabla \cdot (h \mathbf{u} \otimes \mathbf{u})] + \nabla \cdot (f h \mathbf{u}) = -g \hat{z} \cdot (\nabla h \wedge \nabla \eta^{\text{eff}}) + \hat{z} \cdot (\nabla \wedge \mathbf{f}^{\text{nc}}). \quad (35.132)$$

Let us now examine each of these terms and offer physical interpretations.

Boundary pressure torques and turbulent torques

The term $\hat{z} \cdot (\nabla \wedge \mathbf{f}^{\text{nc}})$ in equation (35.132) provides a torque from boundary stresses and interior viscous stresses. The term $g \nabla h \wedge \nabla \eta^{\text{eff}}$ provides a torque whenever the thickness gradients are not aligned with the gradients in the effective surface height. We can write this term in the equivalent form based on the following

$$\rho g \nabla h \wedge \nabla \eta^{\text{eff}} = \rho g \nabla h \wedge \nabla \eta + \nabla h \wedge \nabla p_a \quad (35.133a)$$

$$= -\nabla p_a \wedge \nabla \eta + \nabla p_b \wedge \nabla \eta + \nabla(p_b - p_a) \wedge \nabla p_a / (\rho g) \quad (35.133b)$$

$$= -\nabla p_a \wedge \nabla \eta + \nabla p_b \wedge \nabla[\eta + p_a / (\rho g)] \quad (35.133c)$$

$$= -\nabla p_a \wedge \nabla \eta + \nabla p_b \wedge \nabla[\eta_b + h + p_a / (\rho g)] \quad (35.133d)$$

$$= -\nabla p_a \wedge \nabla \eta + \nabla p_b \wedge \nabla \eta_b \quad (35.133e)$$

$$= \nabla \wedge (-p_a \nabla \eta + p_b \nabla \eta_b), \quad (35.133f)$$

⁵Exercise 35.7 offers a particular example of a geostrophic flow that is aligned perpendicular to ∇h , as found in a geostrophic shallow water front.

where we made use of the hydrostatic relation $p_b = p_a + \rho g h$. We thus see that $-g \nabla h \wedge \nabla \eta^{\text{eff}} = \rho^{-1} \nabla \wedge (p_a \nabla \eta - p_b \nabla \eta_b)$ arises from pressure torques acting on the surface and bottom of the shallow water column. These torques spin the column if there is a misalignment between the boundary pressure gradients and the boundary surface slopes.

Nonlinear transport

The nonlinear term $\hat{\mathbf{z}} \cdot \nabla \wedge [\nabla \cdot (h \mathbf{u} \otimes \mathbf{u})]$ can be written

$$\hat{\mathbf{z}} \cdot \nabla \wedge [\nabla \cdot (h \mathbf{u} \otimes \mathbf{u})] = \hat{z}_m \epsilon_{mst} \partial_s [\partial_n (h u_n u_t)] \quad (35.134a)$$

$$= \partial_n \partial_s [(h u_n) \epsilon_{mst} \hat{z}_m u_t] \quad (35.134b)$$

$$= \partial_n [\partial_s (h u_n) \epsilon_{mst} \hat{z}_m u_t + h u_n \zeta] \quad (35.134c)$$

$$= -\partial_n [\epsilon_{smt} \hat{z}_m u_t \partial_s (h u_n)] + \nabla \cdot (h \zeta \mathbf{u}), \quad (35.134d)$$

which exposes the divergence of the thickness weighted advective flux of relative vorticity. Further manipulations lead to

$$-\partial_n [\epsilon_{smt} \hat{z}_m u_t \partial_s (h u_n)] = -\partial_n [\epsilon_{smt} \hat{z}_m u_t u_n \partial_s h + \epsilon_{smt} \hat{z}_m u_t h \partial_s u_n] \quad (35.135a)$$

$$= \partial_n [\hat{z}_m (\epsilon_{mst} \partial_s h u_t) u_n - h \epsilon_{smt} \hat{z}_m u_t \partial_s u_n] \quad (35.135b)$$

$$= \nabla \cdot [\hat{\mathbf{z}} \cdot (\nabla h \wedge \mathbf{u}) \mathbf{u}] - \partial_n [h (\hat{\mathbf{z}} \wedge \mathbf{u}) \cdot \nabla u_n] \quad (35.135c)$$

$$= \nabla \cdot [\hat{\mathbf{z}} \cdot (\nabla h \wedge \mathbf{u}) \mathbf{u}] + \nabla \cdot [h (v \partial_x - u \partial_y) \mathbf{u}], \quad (35.135d)$$

so that the nonlinear term takes the form

$$\hat{\mathbf{z}} \cdot \nabla \wedge [\nabla \cdot (h \mathbf{u} \otimes \mathbf{u})] = \nabla \cdot [\Sigma \mathbf{u} + h (v \partial_x - u \partial_y) \mathbf{u}]. \quad (35.136)$$

The first term is the advective flux of Σ , and the second term is the thickness weighted transport of \mathbf{u} by the counter-clockwise rotated velocity.

35.10.2 Summary of the column vorticity equation

The above manipulations bring the Σ equation (35.132) into the Eulerian flux-form

$$\partial_t \Sigma = -\nabla \cdot \mathbf{J}^\Sigma + \rho^{-1} \hat{\mathbf{z}} \cdot \nabla \wedge [-p_a \nabla \eta + p_b \nabla \eta_b + \rho \mathbf{f}^{\text{nc}}], \quad (35.137)$$

where

$$\mathbf{J}^\Sigma \equiv [\Sigma + h f + h (v \partial_x - u \partial_y)] \mathbf{u} = h \zeta_a \mathbf{u} + [v (\partial_x h + h \partial_x) - u (\partial_y h + h \partial_y)] \mathbf{u} \quad (35.138)$$

is the flux of Σ . The corresponding material time derivative form of equation (35.137) is given by

$$\frac{D\Sigma}{Dt} + \Sigma \nabla \cdot \mathbf{u} + \nabla \cdot [h f \mathbf{u} + h (v \partial_x - u \partial_y) \mathbf{u}] = \rho^{-1} \hat{\mathbf{z}} \cdot \nabla \wedge [-p_a \nabla \eta + p_b \nabla \eta_b + \rho \mathbf{f}^{\text{nc}}], \quad (35.139)$$

with the thickness equation (31.81)

$$\frac{Dh}{Dt} = -h \nabla \cdot \mathbf{u} + w^{(\dot{\eta})}, \quad (35.140)$$

yielding the equivalent expression

$$h \frac{D(\Sigma/h)}{Dt} + \nabla \cdot [h f \mathbf{u} + h (v \partial_x - u \partial_y) \mathbf{u}] = -\frac{\Sigma w^{(\dot{\eta})}}{h} + \rho^{-1} \hat{\mathbf{z}} \cdot \nabla \wedge [-p_a \nabla \eta + p_b \nabla \eta_b + \rho \mathbf{f}^{\text{nc}}]. \quad (35.141)$$

35.10.3 Steady linear column vorticity balance and the island rule

To garner experience with the column vorticity equation (35.137), consider the steady state and assume \mathbf{J}^Σ is dominated by the planetary vorticity

$$\mathbf{J}^\Sigma \approx h f \mathbf{u}, \quad (35.142)$$

in which case equation (35.137) reduces to the balance

$$\nabla \cdot (h f \mathbf{u}) = \rho^{-1} \hat{\mathbf{z}} \cdot \nabla \wedge [-p_a \nabla \eta + p_b \nabla \eta_b + \rho \mathbf{f}^{nc}]. \quad (35.143)$$

Hence, torques due to interface pressures and turbulent stresses are balanced by the divergence of the thickness weighted advective flux of planetary vorticity.

Formulating the integral balance

The local balance (35.143) leads to an integral balance through an area integral over an arbitrary closed domain

$$\oint_{\mathcal{S}} \nabla \cdot (h f \mathbf{u}) d\mathcal{S} = \rho^{-1} \oint_{\mathcal{S}} \hat{\mathbf{z}} \cdot \nabla \wedge [-p_a \nabla \eta + p_b \nabla \eta_b + \rho \mathbf{f}^{nc}] d\mathcal{S}. \quad (35.144)$$

Making use of Gauss' divergence theorem on the left hand side and Stokes' theorem on the right hand side renders

$$\oint_{\partial\mathcal{S}} h f \mathbf{u} \cdot \hat{\mathbf{n}} d\ell = \rho^{-1} \oint_{\partial\mathcal{S}} [-p_a \nabla \eta + p_b \nabla \eta_b + \rho \mathbf{f}^{nc}] \cdot \hat{\mathbf{t}} d\ell. \quad (35.145)$$

The left hand side is a contour integral with $\hat{\mathbf{n}}$ the outward normal along the contour. The right hand side is also a contour integral, yet with $\hat{\mathbf{t}}$ the tangent along the contour oriented in the counter-clockwise sense. This integral balance says that the advective transport of planetary vorticity leaving the closed region (left hand side) is balanced by the oriented contour integral of the pressure and turbulent stresses.

The integral balance (35.145) is a rather remarkable statement that equates the transport leaving a region to pressure and turbulent stresses integrated along the region boundary. [Godfrey \(1989\)](#) chose a particularly clever contour, such as that depicted in Figure 35.10, and made some assumptions about the stresses that then allowed for an estimate of the transport. We here present the arguments.

Assumptions about the stresses

A portion of the red contour in Figure 35.10 traverses the eastern boundary on both Island A and Island B. As shown in Section 35.9, friction in the eastern boundary region of an ocean gyre is much less than the western boundary, motivating us to ignore friction (either viscous or bottom drag) along the eastern boundary portions of the contour. Likewise, friction is generally small for the open ocean portion along the two latitudinal lines. Hence, wind stress is the only non-conservative process that affects the right hand side of equation (35.145).

[Godfrey \(1989\)](#) furthermore ignored pressure form stresses and their associated torques. This assumption is reasonable for the atmospheric form stress, $-p_a \nabla \eta$, which is generally quite small (it is zero for a rigid lid approximation). However, the bottom form stress, $p_b \nabla \eta_b$, and the associated bottom pressure torque, can be larger than the wind stress and wind stress curl, especially near the continental margins and with strong currents. Hence, ignoring bottom pressure torque is perhaps the least satisfying assumption.

The island rule

The islands are material surfaces so that $\mathbf{u} \cdot \hat{\mathbf{n}} = 0$ along those portions of the contour that are adjacent to the coasts. Integrating along the northern latitude, with $\hat{\mathbf{n}} = \hat{\mathbf{y}}$, yields

$$f_{\text{north}} \int_{y_{\text{north}}} h v dx \equiv f_{\text{north}} \mathcal{T}, \quad (35.146)$$

where \mathcal{T} is the meridional transport. Ignoring any volume transport through the layer surface (e.g., $w^{(\eta)} = 0$), the steady transport crossing the northern boundary equals to that crossing the southern boundary, so that

$$f_{\text{south}} \int_{y_{\text{south}}} h v dx \equiv f_{\text{south}} \mathcal{T}. \quad (35.147)$$

Making use of these results in the integral balance (35.145), and following the above assumptions about the boundary stresses, leads to *Godfrey's island rule*

$$\rho \mathcal{T} = \frac{1}{f_{\text{north}} - f_{\text{south}}} \oint_{\partial S} \boldsymbol{\tau}^{\text{wind}} \cdot \hat{\mathbf{t}} d\ell, \quad (35.148)$$

where $\boldsymbol{\tau}^{\text{wind}}$ is the wind stress, and where the minus sign on f_{south} arises since $\hat{\mathbf{n}} = -\hat{\mathbf{y}}$ along the southern latitude contour. This expression provides an approximation to the flow around Island A, noting that the only nonzero flow normal to the contour is through the two latitudinal segments.

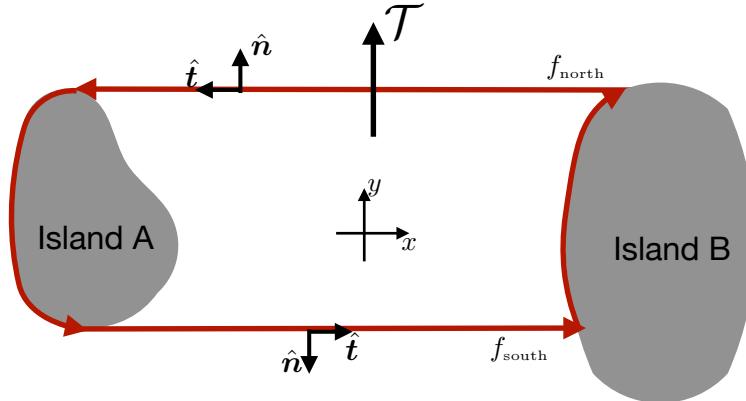
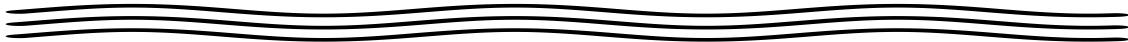


FIGURE 35.10: Illustrating the island rule with two land masses (“islands”). The red contour surrounds all of Island A and traverses along the eastern side of both Island A and Island B. The northern contour extends along a constant latitude line with a corresponding Coriolis parameter f_{north} , whereas the southern contour has Coriolis parameter f_{south} . Ignoring any precipitation or evaporation crossing the surface, steady state volume conservation means that the same meridional transport, \mathcal{T} , crosses both the southern and northern contours. Godfrey's island rule (35.148) provides an estimate for this transport when given the wind stress along the contour.

35.10.4 Further study

[Godfrey \(1989\)](#) applied the island rule (35.148) to estimate transport around Austral-Asia, New Zealand, and Malagasy. Further discussion of Godfrey's island rule can be found in [Tomczak and Godfrey \(1994\)](#), [Pedlosky et al. \(1997\)](#), and [Klinger and Haine \(2019\)](#).



35.11 Exercises

EXERCISE 35.1: FLOW NEAR A TOPOGRAPHIC BUMP IN A REDUCED GRAVITY MODEL

Consider a reduced gravity model with a dynamic lower layer and stagnant upper layer. Place a topographic bump (e.g., seamount or mountain) fully within the lower layer as shown in Figure 35.11.

- What is the potential vorticity for the lower layer? Hint: recall the equations for a reduced gravity model with a stagnant upper layer were derived in Exercise 31.4.
- Now assume an f -plane. If the lower layer potential vorticity is (somehow) horizontally homogenized (e.g., some form of mixing produces a horizontally constant PV), draw the resulting layer interface η_1 . Assume the relative vorticity is negligible compared to the planetary vorticity so that the flow satisfies the planetary geostrophic scaling introduced in Section 35.6.2 and further pursued in Section 40.4. Also, ignore any non-steady processes; we are only interested here in the steady flow.
- For the case of horizontally homogenized potential vorticity from the previous part, what is the direction for a geostrophically balanced flow: cyclonic or anti-cyclonic?

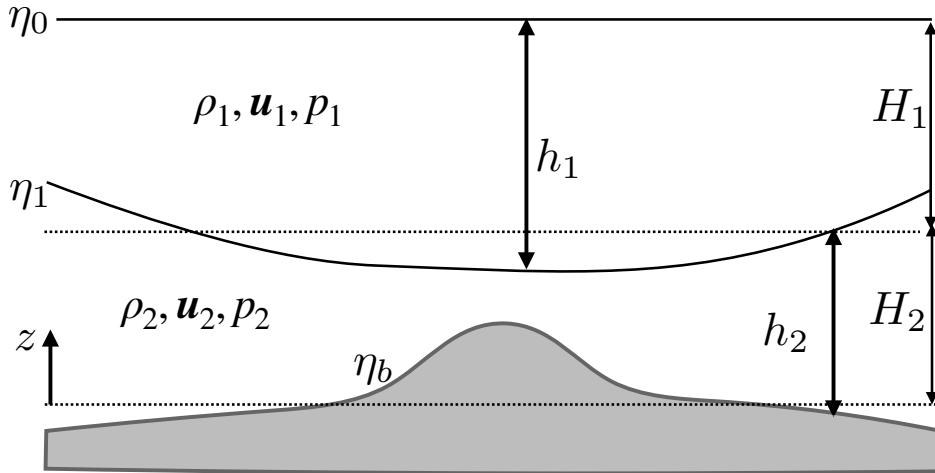


FIGURE 35.11: A reduced gravity model (see Section 31.3) with a stagnant upper layer and a dynamic lower layer as in Exercise 31.4. A seamount sits fully within the lower layer.

EXERCISE 35.2: PV FOR TWO SHALLOW WATER LAYERS

Consider the inviscid Boussinesq two-layer shallow water model as discussed in Section 31.4. Derive the potential vorticity equation for each layer, showing the mathematical steps used in the derivation. Hint: the answer is given in Section 35.4.6.

EXERCISE 35.3: AVERAGE VORTICITY IN A SHALLOW WATER LAYER

Consider a single layer of shallow water fluid on a rotating plane with rotation rate $\Omega = \hat{z}\Omega$. Assume the fluid is contained in an arbitrary horizontal region and that it has a constant total volume given by

$$\mathcal{V} = \int \left[\int dz \right] d\mathcal{S} = \int h d\mathcal{S} = \int (H + \Delta\eta - \eta_b) d\mathcal{S} = H \mathcal{S}, \quad (35.149)$$

where \mathcal{S} is the horizontal area of the domain, $h(x, y, t) = H + \Delta\eta(x, y, t) - \eta_b(x, y)$ is the layer thickness, H is the resting depth relative to $z = 0$, $\Delta\eta$ is the sea level deviation from resting, and η_b is the undulation of the bottom topography (see Figure 31.1). Additionally, recall that $z = 0$ is set according to

$$\int \eta_b \, d\mathcal{S} = 0, \quad (35.150)$$

and volume conservation ensures that

$$\int \Delta\eta \, d\mathcal{S} = 0. \quad (35.151)$$

- (a) Determine the volume average of the vorticity $\hat{\mathbf{z}} \cdot \boldsymbol{\omega}_{\text{solid}}$ arising from the solid-body rotation

$$\langle \hat{\mathbf{z}} \cdot \boldsymbol{\omega}_{\text{solid}} \rangle = \mathcal{V}^{-1} \int \hat{\mathbf{z}} \cdot \boldsymbol{\omega}_{\text{solid}} \, dV. \quad (35.152)$$

- (b) Determine the area average of the relative vorticity,

$$\bar{\zeta} = \mathcal{S}^{-1} \int \hat{\mathbf{z}} \cdot \boldsymbol{\omega} \, d\mathcal{S}, \quad (35.153)$$

in terms of the circulation around the boundary of the domain.

- (c) Determine the volume average of the relative vorticity

$$\langle \zeta \rangle = \mathcal{V}^{-1} \int \hat{\mathbf{z}} \cdot \boldsymbol{\omega} \, dV. \quad (35.154)$$

Write the expression in terms of the area average vorticity, $\bar{\zeta}$, and the correlation, $\overline{\zeta' h'}$, where primes are deviations from the area mean.

EXERCISE 35.4: APPLICATIONS OF SHALLOW WATER PV CONSERVATION

In an adiabatic shallow water fluid in a rotating reference frame, show that the potential vorticity conservation law is

$$\frac{D}{Dt} \left[\frac{\zeta + f}{\eta - \eta_b} \right] = 0, \quad (35.155)$$

where η is the height of the free surface and η_b is the height of the bottom topography (see Figure 31.1). For both of the following questions, assume constant volume for the fluid column. Also, assume the column rotates about its axis as a solid-body.

- (a) A cylindrical column of air at 30° latitude with radius 100 km expands horizontally to twice its original radius. If the air is initially at rest, what is the mean tangential velocity at the perimeter after the expansion?
- (b) An air column at 60°N with zero relative vorticity ($\zeta = 0$) stretches from the surface to the tropopause, which we assume is a rigid lid at 10 km. The air column moves zonally onto a plateau 2.5 km high. What is its relative vorticity? Suppose it then moves southward along the plateau to 30°N , starting from the relative vorticity it obtained from the plateau. What is its new relative vorticity?

EXERCISE 35.5: APPLICATION OF SHALLOW WATER PV CONSERVATION

An air column at 60°N with $\zeta = 0$ initially reaches from the surface to a fixed tropopause at 10 km height. If the air column moves across a mountain 2.5 km high at 45°N , what is its absolute vorticity and relative vorticity as it passes the mountaintop? Hint: Use PV conservation for a shallow-water fluid, and assume the top of the column remains at 10 km.

EXERCISE 35.6: SOME PROPERTIES OF THE STEADY STATE SHALLOW WATER FLUID

Consider a single layer of shallow water fluid in steady state (i.e., all Eulerian time derivatives vanish).

- (a) Show that there exists a streamfunction for the steady state thickness weighted horizontal flow

$$h \mathbf{u} = \hat{z} \wedge \nabla \Psi. \quad (35.156)$$

- (b) What are the physical dimensions of Ψ ?

- (c) Show that the shallow water potential vorticity is a constant along the steady state streamlines of the thickness weighted flow

$$Q = Q(\Psi). \quad (35.157)$$

- (d) Show that the Bernoulli function,

$$B = g \eta + \mathbf{u} \cdot \mathbf{u}/2 \quad (35.158)$$

is also a constant along the same streamlines; i.e.,

$$B = B(\Psi). \quad (35.159)$$

- (e) What is the functional relation between the Bernoulli function and the potential vorticity?

EXERCISE 35.7: ZONALLY SYMMETRIC SHALLOW WATER FRONT

Consider a single layer of shallow water fluid on a β -plane with a flat bottom. Assume all fields possess zonal symmetry as in the zonal front shown in Figure 35.12.

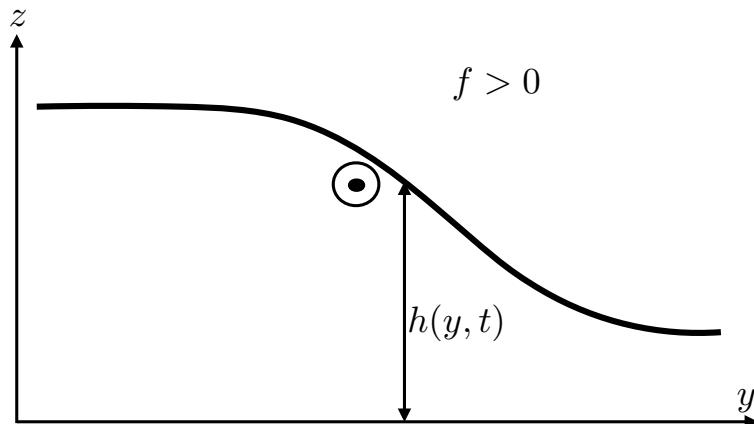


FIGURE 35.12: Schematic of a zonally symmetric front in a shallow water layer in the northern hemisphere ($f > 0$). The thickness decreases to the north. If the flow is in geostrophic balance, then the northward pressure gradient is in geostrophic balance with a southward Coriolis acceleration arising from an eastward (out of the page) geostrophic current (see also Figure 32.1).

- (a) Write the potential vorticity, Q , assuming the fluid is in geostrophic balance. Write in terms of meridional derivatives of the layer thickness.
- (b) From the shallow water equations, explicitly show that the PV is materially constant (i.e., it is a Lagrangian invariant). To do so, work through the usual shallow water PV conservation derivation yet make use of the zonally symmetric equations of motion. Show all relevant steps.
- (c) Show that the potential vorticity can be written as $Q = -(\partial_y M)/h$, where h is the layer thickness. What is the expression for M ?
- (d) Potential vorticity is not the only material constant for this system. Due to the zonal symmetry, Noether's Theorem indicates there is another. Show that M is materially constant.

EXERCISE 35.8: RAYLEIGH DRAG AND GALILEAN INVARIANCE

Recall the discussion of Galilean invariance from Section 13.6. Is the Rayleigh drag used in equation (35.108) Galilean invariant? Why? If not, then should that be of concern for its use in studying flow in a closed and bounded domain?

EXERCISE 35.9: GYRES IN THE PRESENCE OF CYCLONIC WINDS

Consider the ocean gyre discussion in Section 35.9. Rather than anti-cyclonic winds, now apply a cyclonic wind stress to the domain. As per the circulation theorem (35.110), the gyre flow will have a cyclonic sense. Will the resulting gyre exhibit eastern intensification or western? Appeal to whatever arguments you wish.

EXERCISE 35.10: GEOSTROPHIC ADJUSTMENT (BASED ON EXERCISE 4.6 OF *Vallis (2019)*)

Consider the linear geostrophic adjustment problem on an f -plane with a single layer of shallow water fluid over a flat bottom. Rather than assume an initial free surface profile, as we did in Section 35.7, here we assume an initial meridional velocity profile given by

$$v(x, t=0) = v_0 \operatorname{sgn}(x) = v_0 (2 \mathcal{H}(x) - 1), \quad (35.160)$$

where v_0 is a constant, sgn is the sign-function (equation (35.80)), and \mathcal{H} is the Heaviside step function (equation (35.82)). The free surface is assumed to be initially flat.

- (a) Show that the linearized potential vorticity (equation (32.176)) is given by

$$q(x) = 2 v_0 \delta(x), \quad (35.161)$$

where $\delta(x)$ is the Dirac delta distribution whose dimensions are inverse length.

- (b) As we did in Section 35.7, solve for the geostrophic streamfunction $\psi = g \eta/f$.
- (c) Discuss this solution and draw a sketch of ψ and v .

Here are some hints.

- We discuss properties of the Dirac distribution in Section 4.2. However, to answer the first part of this question it is sufficient to know that the derivative of the Heaviside step function (a dimensionless step function) is the Dirac delta distribution

$$\delta(x) = \frac{d\mathcal{H}(x)}{dx}, \quad (35.162)$$

so that the Dirac delta distribution has dimensions of inverse length.

- For the second part, note that the streamfunction is exponentially decaying on either side of the $x = 0$ according to $\psi = \psi_0 e^{-|x|/L_d}$, which then leads to a jump in the derivative approaching the origin from each side. Carefully use equations (35.162) and (35.91) to determine ψ_0 .

EXERCISE 35.11: SHALLOW WATER EQUATIONS WITH DIVERGENCE-DAMPING

When breaking the continuous symmetry of the equations of motion, a discretized numerical simulation admits unphysical flow features sometimes referred to as *computational modes*. Some computational modes can evolve in time with energy accumulating at high wave numbers, in which case the numerical simulation produces unphysical grid noise and becomes of little physical use. To suppress grid noise, numerical models commonly introduce numerical dissipation, even if the continuous equations have zero dissipation. The formulation of numerical dissipation is largely an art guided by the dual needs of suppressing grid noise without otherwise damaging physical properties of the simulated flow. We here consider physical properties of a specific form of numerical dissipation known as *divergence-damping*. We work within the framework of the continuous equations so to develop generic physical properties of the divergence-damping operator. No knowledge of numerical methods is required to solve this problem.

Divergence-damping is motivated by the desire to leave the vorticity equation untouched while damping divergent motion that can arise in numerical simulations. This motivation is based on noting that much of the large-scale circulation in a rotating fluid has a nontrivial absolute vorticity yet a relatively small horizontal divergence. For example, geostrophic flow on an f -plane has vorticity dominated by planetary vorticity f , while it has zero horizontal divergence (see Section 26.4 or the 2d barotropic equation in Section 34.2). The divergence-damping operator is thus designed to reduce the magnitude of the horizontal divergence while leaving the vorticity untouched.

We here examine the impacts of divergence-damping on mechanical energy and angular momentum. For this purpose, consider a single layer of shallow water fluid with divergence-damping. This system is described by the momentum and thickness equations

$$\frac{D\mathbf{u}}{Dt} + f \hat{\mathbf{z}} \wedge \mathbf{u} = -\nabla(g\eta + \alpha\Gamma) \quad (35.163a)$$

$$\frac{Dh}{Dt} = -h\nabla \cdot \mathbf{u}. \quad (35.163b)$$

The parameter $\alpha > 0$ is a constant and the field Γ is given by the Laplacian of the horizontal flow divergence

$$\Gamma = \nabla^2 \mathcal{D}, \quad (35.164)$$

where

$$\mathcal{D} = \nabla \cdot \mathbf{u}. \quad (35.165)$$

The divergence has physical dimensions of inverse time (T^{-1}), so that its Laplacian, Γ , has dimensions of $L^{-2} T^{-1}$, and the coefficient α has dimensions $L^4 T^{-1}$.

Divergence damping leads to a modification to the horizontal pressure gradient. We may think of this modification as arising from the horizontal gradient of a modified free surface height

$$\tilde{\eta} = \eta + \frac{\alpha\Gamma}{g}. \quad (35.166)$$

Notably, mass conservation remains the same since the thickness equation is unchanged. Hence, momentum evolution is modified by changing the pressure gradient, yet the thickness equation remains the same.

- (a) Show that the vorticity equation (35.15) remains unchanged in the presence of divergence-damping.
- (b) Show that the potential vorticity equation (35.36) remains unchanged in the presence of divergence-damping.
- (c) Show that the horizontal divergence evolves according to

$$\frac{\partial \mathcal{D}}{\partial t} = \left[\frac{\partial \mathcal{D}}{\partial t} \right]_{\alpha=0} - \alpha \nabla^2 \Gamma. \quad (35.167)$$

- (d) Show that the evolution of gravitational potential energy per horizontal area

$$\mathcal{P} = g \rho \int_{\eta_b}^{\eta} z \, dz \quad (35.168)$$

remains unchanged from that determined in Section 32.5.1.

- (e) Show that the kinetic energy per horizontal area evolves according to

$$\frac{\partial \mathcal{K}}{\partial t} + \nabla \cdot (\mathbf{u} \mathcal{K}) = -h \rho g \nabla \tilde{\eta}, \quad (35.169)$$

where

$$\mathcal{K} = \frac{1}{2} \int_{\eta_b}^{\eta} \rho \mathbf{u}^2 \, dz = \rho h \mathbf{u}^2 / 2, \quad (35.170)$$

is the horizontal kinetic energy per area (Section 32.5.2).

- (f) Determine the evolution equation for global integrated kinetic energy

$$\frac{\partial}{\partial t} \left[\int \mathcal{K} \, dA \right] = \frac{\partial}{\partial t} \left[\int \int_{\eta_b}^{\eta} (\rho \mathbf{u} \cdot \mathbf{u} / 2) \, dz \, dA \right]. \quad (35.171)$$

Hint: drop all lateral boundary terms by assuming either solid lateral walls or periodicity.

- (g) Consider a single shallow water layer in a rotating tank as in Section 32.7. Show that the material evolution of angular momentum relative to the vertical rotational axis is given by

$$\frac{1}{\delta M} \frac{DL^z}{Dt} = -g \frac{\partial \eta}{\partial \phi} + \mathcal{T}. \quad (35.172)$$

What is the mathematical form for \mathcal{T} ? Hint: check your answer with the next part of this exercise.

- (h) Show that the domain integrated angular momentum satisfies the equation

$$\frac{\partial}{\partial t} \int L^z = \alpha \rho \int \Gamma \frac{\partial \eta}{\partial \phi} \, dA. \quad (35.173)$$

where we assume the bottom topography is flat so that $h = \eta$.

- (i) The linearized thickness equation (see Section 32.8) for a flat bottom is given by

$$\frac{\partial \eta}{\partial t} + H \nabla \cdot \mathbf{u} = 0, \quad (35.174)$$

where H is the thickness of the resting fluid layer. Show that the time change for the global integrated angular momentum is given by

$$\frac{\partial}{\partial t} \int L^z = -\frac{\alpha \rho}{H} \int \left[\frac{\partial}{\partial t} \nabla^2 \eta \right] \frac{\partial \eta}{\partial \phi} dA. \quad (35.175)$$



Vorticity and circulation mechanics

This chapter develops the kinematics and dynamics of vorticity and circulation, extending the shallow water discussion of Chapter 35 to three-dimensional continuously stratified flows. To ground the conceptual presentation, we include sample applications of particular importance to geophysical flows.

READER'S GUIDE FOR THIS CHAPTER

This chapter assumes an understanding of vorticity developed in earlier chapters in this part of the book. We also make use of fluid kinematics from Part III and fluid dynamics from Part IV. As for the shallow water vorticity discussed in Chapter 35, we here make use of vector calculus identities for Cartesian coordinates as detailed in Chapter 2. The concepts and methods developed in this chapter are fundamental to the notions of vorticity and circulation, much of which is encountered in the remainder of this part of the book and elsewhere such as the balanced models considered in Part VII.

36.1	Loose threads	830
36.2	Kinematics of vortex lines and vortex tubes	830
36.2.1	Vortex lines and vortex tubes	830
36.2.2	Kinematic properties	831
36.2.3	Helmholtz's theorems	832
36.2.4	Further study	833
36.3	Kelvin's circulation theorem	833
36.3.1	Formulation	834
36.3.2	Barotropic flow	835
36.3.3	Pressure gradient force for barotropic fluids	835
36.4	Vorticity dynamics	836
36.4.1	Vector-invariant velocity equation	836
36.4.2	Basic form of the vorticity equation	836
36.4.3	Massaged form of the vorticity equation	837
36.4.4	Vorticity, angular momentum, and torques	837
36.5	Mechanics of baroclinicity	838
36.5.1	Curl of the pressure gradient body force	838
36.5.2	Kelvin's circulation theorem and contact pressure forces	839
36.5.3	Bottom pressure contributions at the solid-earth boundary	840
36.5.4	Further study	841
36.6	Vorticity filaments and material line elements	841
36.6.1	A role for the rate of strain	842
36.6.2	Frozen-in nature of vorticity	842

36.6.3	Stretching and tilting of vortex tubes	843
36.6.4	Comments	845
36.7	Circulation and vorticity for rotating fluids	846
36.7.1	Material evolution of absolute circulation	847
36.7.2	The beta effect	848
36.7.3	A two-dimensional incompressible fluid	849
36.7.4	Further study	850
36.8	Vorticity in a hydrostatic and Boussinesq fluid	850
36.8.1	Deriving the vorticity equation	851
36.8.2	Boussinesq form of the baroclinicity vector	852
36.8.3	Vertical vorticity equation	853
36.9	Depth integral of the vertical vorticity equation	854
36.9.1	Leibniz rule expressions	854
36.9.2	Bottom boundary contribution	855
36.9.3	Surface boundary contribution	856
36.10	Vorticity for the depth integrated horizontal flow	857
36.10.1	Comparing the two vorticities	857
36.10.2	Evolution of the depth integrated momentum	858
36.10.3	Evolution of vorticity for the depth integrated horizontal flow	860
36.10.4	Boundary pressure torques	861
36.10.5	Steady state vorticity budget	862
36.10.6	Budget in terms with the vector invariant velocity formulation	863
36.10.7	Vorticity of the depth averaged flow	863
36.10.8	Comments and further study	864
36.11	Exercises	864

36.1 Loose threads

- More on the point that velocity in a Boussinesq fluid is a non-local function of the pressure. However, the vorticity in a barotropic incompressible fluid is a local function of vorticity and velocity. So vorticity diffuses and advects in a local manner, whereas velocity is affected non-locally. This point is nicely discussed on p.43 of [Davidson \(2015\)](#).

36.2 Kinematics of vortex lines and vortex tubes

We here develop the basics of vorticity kinematics, with this discussion closely following from the kinematics of material line elements discussed in Section 14.2.3.

36.2.1 Vortex lines and vortex tubes

A *vortex line* is a line drawn through the fluid that is tangent, at each instance in time, to the vorticity at each spatial point. A vortex line is mathematically parameterized just like any other line, whereby we write the spatial coordinates along the line as a function of a suitable parameter φ (e.g., the arc-length)

$$\boldsymbol{x}(\varphi) = x(\varphi) \hat{\boldsymbol{x}} + y(\varphi) \hat{\boldsymbol{y}} + z(\varphi) \hat{\boldsymbol{z}}. \quad (36.1)$$

The three coordinates of the line are constrained so that the line is tangent to vorticity at each point, which means

$$\frac{dx/d\varphi}{\omega_x} = \frac{dy/d\varphi}{\omega_y} = \frac{dz/d\varphi}{\omega_z}. \quad (36.2)$$

These equations are directly analogous to those satisfied by velocity streamlines (Section 13.8.2)

$$\frac{dx/d\zeta}{u} = \frac{dy/d\zeta}{v} = \frac{dz/d\zeta}{w}, \quad (36.3)$$

where ζ is the parameter along the streamline. Notably, the velocity is not constant along a velocity streamline, nor is vorticity constant along a vortex line. In a steady state, streamlines map the trajectory of a fluid particle (see Section 13.8). However, a vortex line does not offer an interpretation in terms of trajectories.

A *vortex tube* is a bundle of vortex lines that pass through a simple closed curve, with Figure 36.1 illustrating a sample tube. By definition, the sides of the vortex tube are parallel to the vorticity field, since the sides are constructed from vortex lines. We defined a similar notion, the streamtube, for a non-divergent velocity field in Figure 13.5.

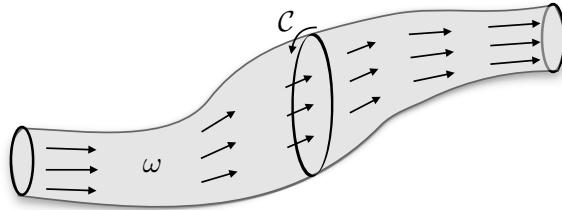


FIGURE 36.1: A vortex line is a line in the fluid that is everywhere tangent to the vorticity vector. A vortex tube is the accumulation of vortex lines passing through a closed loop. A vortex tube is sometimes referred to as a vortex filament. We here depict a vortex tube within the fluid and illustrate the circulation around the tube $\mathcal{C} = \oint_{\partial S} \mathbf{v} \cdot d\mathbf{r} = \int \boldsymbol{\omega} \cdot \hat{\mathbf{n}} dS$. Since vorticity has zero divergence, the circulation is the same for any loop around the vortex tube (Helmholtz's first theorem from Section 36.2.3). A uniform circulation along the tube means that the magnitude of the vorticity is larger in regions where the tube has a small area and conversely where the tube has a large area.

36.2.2 Kinematic properties

Vorticity has zero divergence

$$\nabla \cdot \boldsymbol{\omega} = \nabla \cdot (\nabla \wedge \mathbf{v}) = 0, \quad (36.4)$$

which follows since the divergence of a curl vanishes. Integrating the non-divergence relation over an arbitrary closed volume within the fluid leads to

$$\int_{\mathcal{R}} \nabla \cdot \boldsymbol{\omega} dV = \oint_{\partial \mathcal{R}} \boldsymbol{\omega} \cdot \hat{\mathbf{n}} dS = 0, \quad (36.5)$$

where we made use of Gauss's divergence theorem to reach the surface integral expression, with $\hat{\mathbf{n}} dS$ the oriented area element on the boundary of the volume, $\partial \mathcal{R}$, and $\hat{\mathbf{n}}$ the outward normal on the boundary. This result means there is no net vorticity entering or leaving an arbitrary closed region. That is, there is a vanishing net integrated “flux” of vorticity across the closed region. Consequently, there are no sources or sinks of vorticity within the fluid. In turn, there is no accumulation of vorticity within any arbitrary closed region.

Now specialize the surface integral in equation (36.5) to a volume along a chosen vortex tube such as in Figure 36.1.¹ The two ends of the tube generally have different cross-sectional areas. The integral over the sides of the vortex tube vanishes, since the vorticity is parallel to the tube sides. Hence, the surface integral only picks up contributions from the two ends of the tube

$$\int_A \boldsymbol{\omega} \cdot \hat{\mathbf{n}} d\mathcal{S}_A + \int_B \boldsymbol{\omega} \cdot \hat{\mathbf{n}} d\mathcal{S}_B = 0. \quad (36.6)$$

Note that the outward normals point in the opposite direction, so that we see that the flux of vorticity is independent of position along the tube. Correspondingly, Stoke's theorem transfers the vorticity constraint to a constraint on the circulation around the circumference of the tube, so that

$$\oint_A \mathbf{v} \cdot d\mathbf{r} + \oint_B \mathbf{v} \cdot d\mathbf{r} = 0. \quad (36.7)$$

We thus see that the circulation around the vortex tube is the same no matter where it is computed. The circulation constraints (36.6) and (36.7) are kinematic, holding for any vorticity field. We now consider some consequences of this constraint.

36.2.3 Helmholtz's theorems

There are a few basic properties of vorticity that follow from its vanishing divergence. These properties are known as Helmholtz's theorems.

Helmholtz's first theorem

Since the cross-sectional slices used to derive the circulation constraint (36.7) are arbitrary, the constraint holds throughout the full extent of the vortex tube. Hence, as noted following equation (36.7), the circulation is the same for any position along the vortex tube; i.e., the strength of a vortex tube is constant along its length (see Figure 36.1). This result is known as *Helmholtz's first theorem*.

As a corollary, we refer to the vorticity constraint (36.6) to note that changes in the vortex tube cross-sectional area are compensated by changes in vorticity. For example, let the vortex tube shrink over some region. To maintain constant circulation along the tube, the vorticity magnitude must increase where the area decreases, which in turn means that the velocity circulating around the tube increases in magnitude as the area reduces. Think of a tornado as in Figure 36.2, which is a natural expression of a vortex tube. Near the ground, the cross-sectional area of the tornado is small, with the tangential velocity of a fluid particle within the tube relatively large. Near the tornado top, the cross-sectional area is large so the tangential velocity is relatively small.

Helmholtz's second theorem

The vorticity constraint (36.6) cannot be satisfied by a finite vorticity if the area of a vortex tube vanishes anywhere. Hence, a vortex tube cannot begin or end within the fluid. This result follows from the absence of vortex sources and sinks within the fluid. Hence, a vortex tube can only loop with itself (e.g., a smoke ring as in Figure 36.3), or intersect a boundary (as for a tornado in Figure 36.2, where the ground and clouds form the boundary).

¹In Exercise 17.7, we developed a similar set of results for a streamtube in an incompressible fluid.

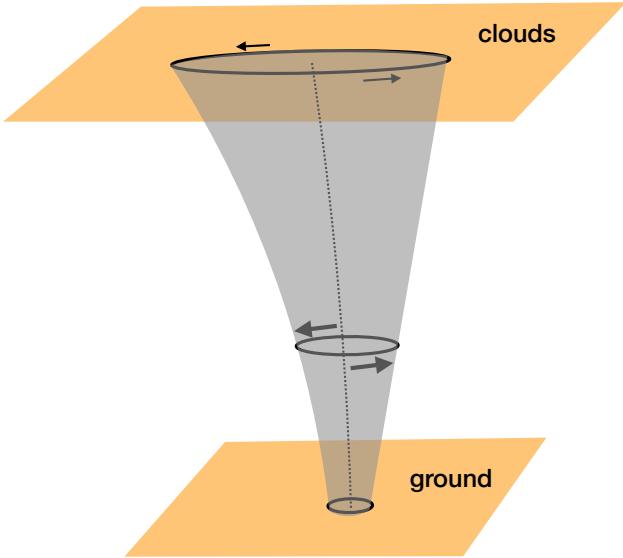


FIGURE 36.2: A vortex tube idealization of a tornado. Since the circulation around the tube is uniform (Helmholtz's first theorem), the tangential velocity of a fluid particle has a larger magnitude in regions where the vortex area is smaller, such as near the ground. As the tornado reaches into the clouds, it generally has a larger cross-sectional area and thus a smaller magnitude for the tangential velocity.

Helmholtz's third theorem

Helmholtz's third theorem states that an unforced inviscid barotropic fluid that has zero vorticity will remain irrotational forever. This theorem is a special case of Kelvin's circulation theorem, which is discussed in Section 36.3.

36.2.4 Further study

A particularly insightful and pedagogical discussion of these ideas can be found in Chapter 5 of [Acheson \(1990\)](#). Additionally, the following videos offer laboratory demonstrations of vorticity in non-rotating and rotating flows.

- Helmholtz's theorems are vividly exhibited by [this video](#) from the *Physics Girl* of flow generated by a paddle in a swimming pool. She also discusses vortex rings in [this video](#).
- Vorticity and Helmholtz's theorems are also described by [this classic video](#) from Prof. Shapiro.
- A rotating tank experiment shown near the 15 minute mark of [this video](#) from Prof. Fultz shows how vorticity is affected by vortex stretching.
- [This video](#) offers a classic tutorial on vorticity in non-rotating fluids from Prof. Shapiro.

36.3 Kelvin's circulation theorem

Kelvin's circulation theorem is concerned with the evolution of circulation around a closed material loop, or equivalently (through Stokes' theorem) with the change in vorticity penetrating the enclosed area

$$\frac{d\mathcal{C}}{dt} = \frac{d}{dt} \oint_{\partial\mathcal{S}(\mathbf{v})} \mathbf{v} \cdot d\mathbf{r} = \frac{d}{dt} \int_{\mathcal{S}(\mathbf{v})} \boldsymbol{\omega} \cdot \hat{\mathbf{n}} d\mathcal{S}, \quad (36.8)$$

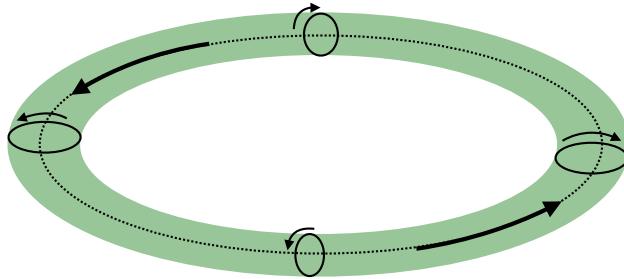


FIGURE 36.3: A vortex ring (torus) is a vortex tube that closes on itself. We here depict a vortex ring with vorticity pointing counter-clockwise around the ring. The tangential velocity is oriented so that the vorticity points according to the right hand rule. That is, orient the fingers on the right hand according to the tangential velocity. The thumb of the right hand then points in the direction of the vorticity vector.

where $\mathcal{S}(\mathbf{v})$ designates a surface whose points all move with the fluid flow. We here consider the case of non-rotating flow. When considering the rotating case in Section 36.7, we will see that circulation in the non-rotating case directly corresponds to the absolute circulation (relative circulation plus circulation due to planetary rotation).

36.3.1 Formulation

We are here concerned with circulation for material loops. Hence, the time derivative acting on the circulation integral in equation (36.8) moves inside the integral as a material time operator

$$\frac{dC}{dt} = \frac{d}{dt} \oint_{\partial\mathcal{S}(\mathbf{v})} \mathbf{v} \cdot d\mathbf{r} = \oint_{\partial\mathcal{S}(\mathbf{v})} \frac{D(\mathbf{v} \cdot d\mathbf{r})}{Dt}. \quad (36.9)$$

The material evolution of \mathbf{v} is determined by Newton's law of motion, which for a non-rotating flow is given by (see Section 22.10)

$$\frac{D\mathbf{v}}{Dt} = -\frac{1}{\rho} \nabla p - \nabla \Phi + \mathbf{F}. \quad (36.10)$$

In this equation, p is the pressure, ρ is the mass density, Φ is the geopotential and/or the potential for any conservative force, and \mathbf{F} is a non-conservative force such as from viscous stresses and/or boundary interactions.

The material time derivative of the differential line element moving around the circuit equals to the differential of the velocity on the circuit

$$\frac{D(d\mathbf{r})}{Dt} = d\mathbf{v}. \quad (36.11)$$

This result follows since all points along the circuit are defined by fluid particles. Consequently,

evolution of circulation following a material loop becomes

$$\frac{dC}{dt} = \frac{d}{dt} \oint_{\partial S(\mathbf{v})} \mathbf{v} \cdot d\mathbf{r} \quad (36.12a)$$

$$= \oint_{\partial S(\mathbf{v})} \left[\left(-\frac{1}{\rho} \nabla p - \nabla \Phi + \mathbf{F} \right) \cdot d\mathbf{r} + \frac{1}{2} d\mathbf{v}^2 \right] \quad (36.12b)$$

$$= \oint_{\partial S(\mathbf{v})} \left[-\frac{1}{\rho} \nabla p + \mathbf{F} \right] \cdot d\mathbf{r} \quad (36.12c)$$

$$= \int_{S(\mathbf{v})} \left[-\nabla \wedge \left(\frac{1}{\rho} \nabla p \right) + \nabla \wedge \mathbf{F} \right] \cdot \hat{\mathbf{n}} dS \quad (36.12d)$$

$$= \int_{S(\mathbf{v})} (\mathbf{B} + \nabla \wedge \mathbf{F}) \cdot \hat{\mathbf{n}} dS. \quad (36.12e)$$

The third equality follows since $\nabla \Phi \cdot d\mathbf{r} = d\Phi$ has zero integral around a closed circuit, as does $d\mathbf{v}^2$. The fourth equality made use of Stokes' theorem. The final equality introduced the solenoidal vector

$$\mathbf{B} = \frac{\nabla \rho \wedge \nabla p}{\rho^2}, \quad (36.13)$$

which we also refer to as the *baroclinicity vector*. The baroclinicity vector has physical dimensions of inverse squared time, T^{-2} . Equation (36.12e) says that the circulation around a material loop is affected by two processes: baroclinicity and the curl of any non-conservative (e.g., friction) forces.

36.3.2 Barotropic flow

The solenoidal/baroclinicity vector vanishes for a constant density fluid, in which $\nabla \rho = 0$ such as for a single layer of shallow water fluid. More generally, the baroclinicity vector vanishes for barotropic flow, in which

$$p = p(\rho) \implies \text{barotropic flow.} \quad (36.14)$$

Kelvin's theorem then follows, which states that for inviscid barotropic flow the circulation around any closed material circuit remains constant

$$\frac{dC}{dt} = \frac{d}{dt} \oint_{\partial S(\mathbf{v})} \mathbf{v} \cdot d\mathbf{r} = \frac{d}{dt} \int_{S(\mathbf{v})} \boldsymbol{\omega} \cdot \hat{\mathbf{n}} dS = 0 \quad \Leftarrow \text{inviscid barotropic flow.} \quad (36.15)$$

That is, the circulation around any vortex tube in a perfect barotropic fluid moves in a manner that keeps the circulation materially constant. This remarkable result greatly constrains the motion of a barotropic perfect fluid.

36.3.3 Pressure gradient force for barotropic fluids

As noted above, a barotropic fluid has pressure that is a function just of density, and conversely the density is a function just of the pressure. Hence, the curl of the pressure gradient force vanishes

$$\nabla \wedge (\rho^{-1} \nabla p) = 0 \quad (36.16)$$

and there is a corresponding scalar potential for the pressure gradient force

$$\nabla \Phi_p = \rho^{-1} \nabla p. \quad (36.17)$$

Integration renders²

$$\Phi_p = \int_{p_0}^p \frac{dp'}{\rho(p')}, \quad (36.18)$$

where p_0 is an arbitrary reference pressure. We make use of the scalar potential, Φ_p , in Exercise 36.8.

36.4 Vorticity dynamics

We now move from the circulation around a macroscopic circuit to the vorticity at a point. In particular, we seek information for how vorticity changes in time. What physical processes lead to these changes? As for Kelvin's theorem, we make use of Newton's law of motion, written here in the form for a rotating fluid (see Section 22.10)

$$\rho \left[\frac{D}{Dt} + 2\boldsymbol{\Omega} \wedge \right] \mathbf{v} = -\nabla p - \rho \nabla \Phi + \rho \mathbf{F}, \quad (36.19)$$

where $\boldsymbol{\Omega}$ is the angular velocity of the rotating reference frame.

36.4.1 Vector-invariant velocity equation

As for the shallow water fluid in Section 35.2, we find it useful to convert the advective-form momentum equation to vector-invariant form. For this purpose, make use of the vector identity (see Section 2.3.4)

$$\boldsymbol{\omega} \wedge \mathbf{v} = -(1/2) \nabla(\mathbf{v} \cdot \mathbf{v}) + (\mathbf{v} \cdot \nabla) \mathbf{v} \quad (36.20)$$

to eliminate velocity self-advection in favor of vorticity

$$\frac{\partial \mathbf{v}}{\partial t} + \boldsymbol{\omega}_a \wedge \mathbf{v} = -\frac{1}{\rho} \nabla p - \nabla \left[\frac{1}{2} \mathbf{v}^2 + \Phi \right] + \mathbf{F}. \quad (36.21)$$

We here introduced the absolute vorticity

$$\boldsymbol{\omega}_a = \boldsymbol{\omega} + 2\boldsymbol{\Omega}, \quad (36.22)$$

which is the sum of the relative vorticity and the planetary vorticity (see Section 33.5.1).

36.4.2 Basic form of the vorticity equation

Taking the curl of the vector-invariant momentum equation (36.21) removes the mechanical energy per mass, $\mathbf{v}^2/2 + \Phi$, thus leaving

$$\frac{\partial \boldsymbol{\omega}}{\partial t} + \nabla \wedge (\boldsymbol{\omega}_a \wedge \mathbf{v}) = \frac{1}{\rho^2} (\nabla \rho \wedge \nabla p) + \nabla \wedge \mathbf{F}. \quad (36.23)$$

For geophysical fluids we generally assume that $\boldsymbol{\Omega}$ has zero time tendency, so that

$$\frac{\partial \boldsymbol{\omega}_a}{\partial t} = \frac{\partial (\boldsymbol{\omega} + 2\boldsymbol{\Omega})}{\partial t} = \frac{\partial \boldsymbol{\omega}}{\partial t}, \quad (36.24)$$

in which case equation (36.23) can be written

$$\frac{\partial \boldsymbol{\omega}_a}{\partial t} + \nabla \wedge (\boldsymbol{\omega}_a \wedge \mathbf{v}) = \mathbf{B} + \nabla \wedge \mathbf{F}, \quad (36.25)$$

where \mathbf{B} is the baroclinicity vector introduced by equation (36.13).

²See page 79 of Müller (2006) for more on equation (36.17).

36.4.3 Massaged form of the vorticity equation

Physical interpretation of the term $\nabla \wedge (\omega_a \wedge \mathbf{v})$ appearing in the prognostic equation (36.25) can be made more transparent by using yet another vector identity

$$\nabla \wedge (\omega_a \wedge \mathbf{v}) = (\mathbf{v} \cdot \nabla) \omega_a - (\omega_a \cdot \nabla) \mathbf{v} + \omega_a \nabla \cdot \mathbf{v} - \mathbf{v} \nabla \cdot \omega_a \quad (36.26a)$$

$$= (\mathbf{v} \cdot \nabla) \omega_a - (\omega_a \cdot \nabla) \mathbf{v} - \frac{\omega_a}{\rho} \frac{D\rho}{Dt}. \quad (36.26b)$$

The second equality required the continuity equation

$$\frac{D\rho}{Dt} = -\rho \nabla \cdot \mathbf{v}, \quad (36.27)$$

and the non-divergent nature of the absolute vorticity

$$\nabla \cdot \omega_a = \nabla \cdot (\nabla \wedge \mathbf{v} + 2\Omega) = 0. \quad (36.28)$$

Equation (36.25) thus takes the form

$$\frac{D\omega_a}{Dt} - \frac{\omega_a}{\rho} \frac{D\rho}{Dt} = (\omega_a \cdot \nabla) \mathbf{v} + \frac{1}{\rho^2} (\nabla \rho \wedge \nabla p) + \nabla \wedge \mathbf{F}, \quad (36.29)$$

which can be written in vector form as

$$\rho \frac{D(\omega_a/\rho)}{Dt} = (\omega_a \cdot \nabla) \mathbf{v} + \mathbf{B} + \nabla \wedge \mathbf{F}. \quad (36.30)$$

Equation (36.30) is the desired form of the vorticity evolution equation. Each term on the right hand side represents a distinct physical process that impacts material evolution of vorticity. The first term, $(\omega_a \cdot \nabla) \mathbf{v}$, will be explored in Section 36.6 in the simplified context of a barotropic fluid. The second term arises from baroclinicity as introduced in equation (36.13) and given a mechanical interpretation in Section 36.5. The third term arises from the curl of the friction vector, which contributes especially in boundary layer regions where friction curls are relatively large.

36.4.4 Vorticity, angular momentum, and torques

Like vorticity, angular momentum provides a measure of the rotational motion of a fluid. However, there are key distinctions. Perhaps the most fundamental distinction is that vorticity measures the rotation or spin without reference to an origin, whereas angular momentum is computed relative to a subjectively chosen origin. Vorticity is thus an intrinsic property of the fluid flow, whereas angular momentum depends on the chosen origin.

Angular momentum of motion relative to an origin changes in the presence of torques computed about the chosen origin, with the torque equal to the cross product of the position vector of a point and the force vector acting at that point. In contrast, vorticity at a point is affected by the curl of the force per mass acting at the point. Furthermore, angular momentum is a property of any mechanical system, including point particles and rigid bodies, whereas vorticity is a property only of a continuous media where we can compute spatial derivatives. In Chapter 37, we further pursue the kinematic relations between angular momentum and vorticity. In that discussion we see that angular momentum and vorticity are fundamentally distinct in a fluid that has straining motion between fluid elements. In contrast, angular momentum and vorticity are directly proportional (through the moment of inertia tensor) when the fluid exhibits solid body motion in which strain vanishes.

When the curl of a force per mass is applied to a fluid and thus changes its vorticity, we commonly use the term “torque” in reference to this force curl. For example, in Section 36.5 we explore baroclinicity, which is the key mechanism for how inviscid torques from pressure modify vorticity. That is, baroclinicity provides a vorticity source when the pressure force acting on a fluid element does not pass through the center of mass of that element. When there is baroclinicity, the pressure force spins the fluid element thus affecting vorticity. Analogous inviscid and viscous force curls act on boundaries, such as when a fluid interacts with the solid earth. It is within this context that we use the term “torque” when referring to a vorticity source. Correspondingly, the torques providing a vorticity source have the dimension of force per mass per length, whereas the torques altering angular momentum have the dimension of force times length.

36.5 Mechanics of baroclinicity

Baroclinicity is present in most geophysical flows, thus affecting the material evolution of circulation and vorticity. Flow with a nonzero baroclinicity vector is generally referred to as *baroclinic flow*, whereas *barotropic* flows have zero baroclinicity. We illustrate the basic distinction between barotropic and baroclinic fluids in Figure 36.4. We observe that a baroclinic fluid is associated with fluid motion whereas a barotropic fluid that has its pressure and density surfaces aligned with geopotentials support no motion. We further develop these points in the following as we explore the mechanics of baroclinicity.

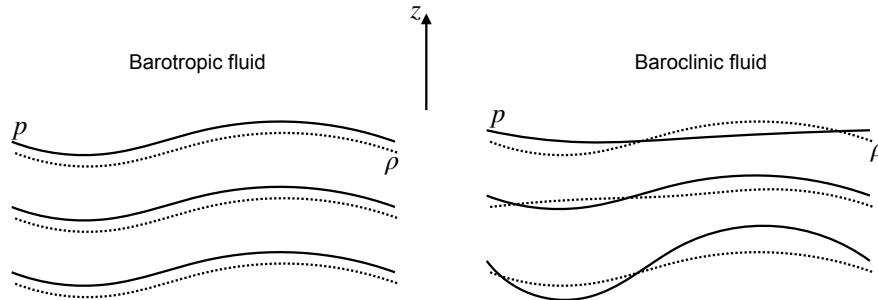


FIGURE 36.4: Left panel: a barotropic fluid, whereby density is a function just of pressure, $\rho = \rho(p)$, so that density surfaces (dashed lines) and pressure surfaces (solid lines) are parallel. Horizontal density and pressure surfaces in a barotropic fluid support no motion. Right panel: a baroclinic fluid, whereby density and pressure surfaces generally differ so that density is a function of more than just the pressure. A baroclinic fluid is associated with fluid motion.

36.5.1 Curl of the pressure gradient body force

Baroclinicity is the curl of the pressure gradient body force

$$\mathbf{B} = \nabla \wedge \mathbf{F}_{\text{press}} = \nabla \wedge (-\nabla p / \rho) = -\nabla \rho^{-1} \wedge \nabla p = \frac{\nabla \rho \wedge \nabla p}{\rho^2}. \quad (36.31)$$

As discussed in Section 36.4.4, the curl of a force provides a torque that spins the fluid, thus rendering a vorticity source. Geometrically, baroclinicity arises when there is nonzero change in pressure along contours of constant density, or conversely changes in density along contours of constant pressure.³ It can be useful to introduce the notion of a *solenoid*, which is a tube region in

³See Exercise 5.1 for a two-dimensional example of this geometry.

the fluid that is perpendicular to both $\nabla\rho$ and ∇p . Solenoids vanish for barotropic flows, whereby $p = p(\rho)$ (see equation (36.14)). For baroclinic flow, solenoids introduce a torque at each point that affects vorticity.

To further understand the mechanical interpretation of solenoids in terms of a torque, consider the cross product

$$\rho \mathbf{B} = \mathbf{F}^{\text{press}} \wedge \nabla\rho = (-\rho^{-1} \nabla p) \wedge \nabla\rho. \quad (36.32)$$

The first term on the right hand side is the pressure gradient force that acts down the pressure gradient. Now consider a tiny fluid element such as shown in Figure 36.5. By construction, the pressure force acts at the geometric center of the element. However, the nonzero density gradient means that the center of mass for the fluid element is not at the geometric center. Since the pressure force does not pass through the center of mass, it imparts a torque to the fluid element. This torque then modifies the vorticity and hence the circulation around the boundary of the element. Only when the pressure force is aligned with the density gradient (barotropic flow), or if the density is spatially uniform (e.g., constant density homogeneous fluid) does the pressure force pass through the center of mass, thus creating no torque and not affecting vorticity.

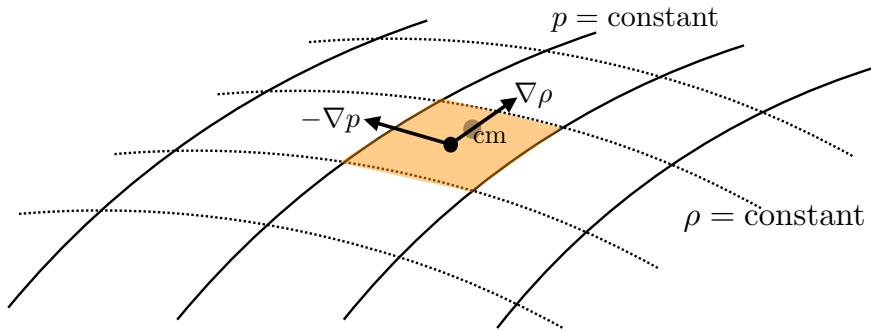


FIGURE 36.5: A mechanical interpretation of the baroclinicity vector. We consider a tiny fluid element bounded by surfaces of constant pressure and density. By construction, the pressure force acts at the geometric center of the element, whereas the center of mass for the element is off-center due to the density gradient across the element. The pressure force thus provides a torque for the fluid element, with the moment-arm for the torque determined by the distance between the geometric center and the center of mass. This torque modifies the vorticity of the fluid element, and in turn modifies the circulation computed around the element's boundary. As depicted here, the baroclinicity vector points into the page (right hand rule for $(-\rho^{-1} \nabla p) \wedge \nabla\rho$), so that this baroclinicity spins-up a clockwise circulation around the element, or equivalently a clockwise vorticity. This figure is based on Figure 14.9 of [Thorne and Blandford \(2017\)](#).

36.5.2 Kelvin's circulation theorem and contact pressure forces

We are afforded another means to understand baroclinicity by returning to the formulation of Kelvin's circulation theorem in Section 36.3.1. Focusing just on the baroclinicity contribution in equation (36.12e) we have

$$\left[\frac{d\mathcal{C}}{dt} \right]_{\text{baroclinicity}} = \oint_{\partial\mathcal{S}(\mathbf{v})} -\rho^{-1} \nabla p \cdot d\mathbf{r} = \oint_{\partial\mathcal{S}(\mathbf{v})} \rho^{-2} (-p \nabla\rho) \cdot d\mathbf{r}, \quad (36.33)$$

which follows since

$$\oint_{\partial\mathcal{S}(\mathbf{v})} \nabla(p/\rho) \cdot d\mathbf{r} = 0. \quad (36.34)$$

The term $-p \nabla \rho$ in equation (36.33) is proportional to the compressive contact pressure force (Section 19.1) acting normal to a constant density surface. Consequently, if the material surface on which we are computing circulation happens to be parallel to a constant density surface, then pressure cannot generate any circulation around that material circuit. The left circuit in Figure 36.6 illustrates this situation. In contrast, in the more general case where a material surface crosses constant density surfaces, pressure modifies circulation around such material circuits (right circuit in Figure 36.6).

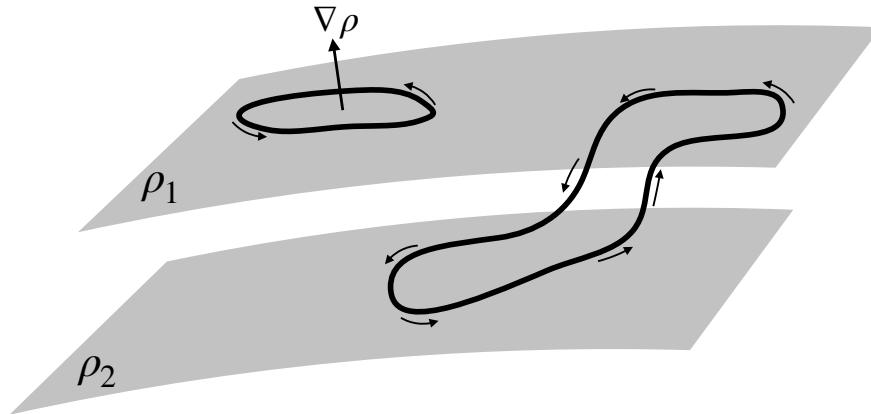


FIGURE 36.6: The material circuit on the left is assumed to be within a constant density surface. In this case, $\nabla \rho \cdot d\mathbf{r} = 0$ so that pressure cannot modify the circulation around this circuit. If material surfaces are not within constant density surface, such as depicted on the right, then pressure is able to modify the circulation computed around any such material circuits.

36.5.3 Bottom pressure contributions at the solid-earth boundary

As an application of the above ideas from Section 36.5.2, consider a fluid region that intersects the solid-earth boundary. The solid-earth boundary is material so that we can apply Kelvin's circulation theorem to a circuit on the boundary. Consider the situation in Figure 36.7, which shows a vertical slice next to a sloping bottom with constant density surface intersecting the bottom. As in our considerations in Section 36.5.2, any material circuit that sits within the bottom boundary crosses density surfaces, so that circulation will be affected by the bottom pressure. Indeed, even if the bottom is flat, so long as density is not constant along the bottom, then a material circuit within the bottom will have its circulation modified by bottom pressure.

To develop the mathematics of the above ideas, write the differential line element within the bottom circuit as

$$d\mathbf{r} = \hat{\mathbf{x}} dx + \hat{\mathbf{y}} dy + \hat{\mathbf{z}} dz = (\hat{\mathbf{x}} + \hat{\mathbf{z}} \partial_x \eta_b) dx + (\hat{\mathbf{y}} + \hat{\mathbf{z}} \partial_y \eta_b) dy. \quad (36.35)$$

To reach this result we set $z = \eta_b(x, y)$ since the circuit is along the bottom boundary, which in turn means that

$$dz = d\eta_b = \nabla_z \eta_b \cdot (\hat{\mathbf{x}} dx + \hat{\mathbf{y}} dy). \quad (36.36)$$

Consequently, the projection of the density gradient onto the circuit is given by

$$\nabla \rho \cdot d\mathbf{r} = (\nabla_z \rho + \partial_z \rho \nabla_z \eta_b) \cdot d\mathbf{x}. \quad (36.37)$$

Making use of this result in Kelvin's circulation theorem and focusing on the pressure contribution, as in equation (36.33), leads to

$$\left[\frac{dC}{dt} \right]_{\text{bottom}} = - \oint_{\partial S_{\text{bottom}}} p \rho^{-2} \nabla \rho \cdot d\mathbf{r} = - \oint_{\partial S_{\text{bottom}}} \frac{p_b}{\rho^2} (\nabla_z \rho + \partial_z \rho \nabla_z \eta_b) \cdot dx. \quad (36.38)$$

There are two contributions to the circulation changes revealed by equation (36.38). The first arises from the sloped density surfaces next to the bottom, and the second arises from the sloped bottom multiplied by the vertical density gradient. These two contributions are weighted by the bottom pressure, p_b , which is normalized by the squared density. Circulation modifications are enhanced by increased horizontal density gradients next to the bottom, as well as increased topographic slopes. For the special case of flat topography and flat density there are no bottom pressure-induced changes to the circulation around a bottom material circuit.

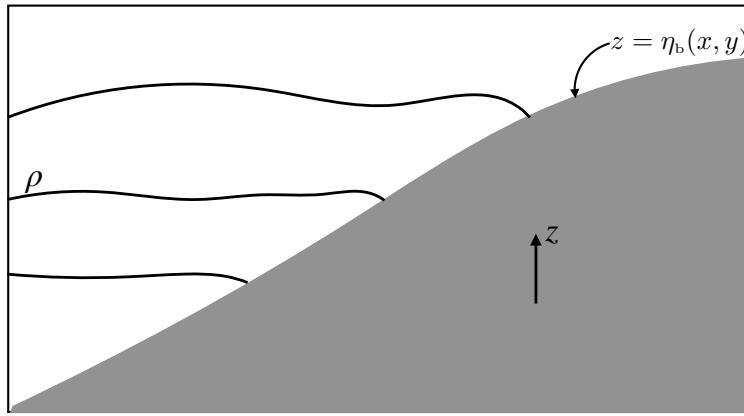


FIGURE 36.7: Constant density surfaces intersecting a sloped solid-earth boundary. Any circuit that sits along the boundary is material since the bottom is material. For circuits that cross density surfaces, the bottom pressure acts to modify circulation computed for this circuit.

36.5.4 Further study

[This video from Prof. Shapiro](#) provides a lucid discussion of baroclinicity and its role in affecting vorticity.

36.6 Vorticity filaments and material line elements

We here study the physics of the source term

$$(\boldsymbol{\omega}_a \cdot \nabla) \mathbf{v} = f \partial_z \mathbf{v} + (\boldsymbol{\omega} \cdot \nabla) \mathbf{v}. \quad (36.39)$$

The contribution from $\hat{z} f \partial_z w$ to the first term is further explored when studying the planetary geostrophic equations in Sections 26.5 and 41.3, given its importance for large-scale meridional motion on a spherical planet. It is the the second term, $(\boldsymbol{\omega} \cdot \nabla) \mathbf{v}$, that is the focus of this section.

36.6.1 A role for the rate of strain

To help unpack the physics of the source, $(\boldsymbol{\omega} \cdot \nabla) \mathbf{v}$, we find it useful to write it in the following form found by exposing Cartesian tensor labels

$$\omega_m \partial_m v_n = (\omega_m/2) [(\partial_m v_n + \partial_n v_m) + (\partial_m v_n - \partial_n v_m)] \quad (36.40a)$$

$$= \omega_m \mathbb{S}_{mn} - \omega_m \mathbb{A}_{mn}, \quad (36.40b)$$

where $\mathbb{S}_{mn} = (1/2)(\partial_m v_n + \partial_n v_m)$ are components to the rate of strain tensor and $\mathbb{A}_{mn} = (1/2)(\partial_n v_m - \partial_m v_n)$ are components to the rotation tensor. These tensors were introduced in Section 14.2 when studying the kinematics of line elements. As shown in that discussion, the rotation tensor is related to the vorticity by the identity (14.44), whose use leads to

$$\omega_m \mathbb{A}_{mn} = -\omega_m \epsilon_{mnp} \omega_p / 2 = 0. \quad (36.41)$$

This identity follows since the outer product $\omega_m \omega_p$ is symmetric on the labels m, p , whereas these labels are anti-symmetric in the permutation symbol, ϵ_{mnp} . We are thus left just with

$$(\boldsymbol{\omega} \cdot \nabla) \mathbf{v} = \boldsymbol{\omega} \cdot \mathbb{S}, \quad (36.42)$$

so that this source term is unaffected by fluid rotation. Hence, regions where the vorticity projects onto the rate of strain tensor see an enhancement of the vorticity magnitude, thus highlighting the fundamental role for fluid strains as a vorticity source.

36.6.2 Frozen-in nature of vorticity

Consider an incompressible, inviscid, non-rotating barotropic fluid, in which case the vorticity equation (36.30) reduces to

$$\frac{D\boldsymbol{\omega}}{Dt} = (\boldsymbol{\omega} \cdot \nabla) \mathbf{v}. \quad (36.43)$$

This evolution equation is kinematically identical to evolution of a material line element detailed in Section 14.2. Nonetheless, we emphasize a key distinction, as exploited in deriving the identity (36.41), whereby the rotation tensor has elements comprised of vorticity so that the vorticity source only involves the rate of strain tensor. In contrast, the source, $(\delta \mathbf{x} \cdot \nabla) \mathbf{v}$, affecting motion of a line element also involves the rotation tensor. Even so, in furthering our understanding of the vorticity source, $(\boldsymbol{\omega} \cdot \nabla) \mathbf{v}$, we find much insights by examining the relation between the kinematics of vorticity and the kinematics of line elements.

Recall from Section 36.2.1 that a vortex line is a line drawn through the fluid that is everywhere parallel to the vorticity. Such a line connects material fluid particles, so that a vortex line constitutes a material line. At time $t = 0$, let the vorticity on an infinitesimal vortex line be related to the initial material line element

$$\delta \mathbf{x}(0) = \Gamma \boldsymbol{\omega}(0), \quad (36.44)$$

where Γ is scalar with dimensions $L T$ that is determined by the initial vorticity and initial line element. Importantly, this relation follows by construction. We are free to draw an infinitesimal vortex line and call it a material line. The key point is that the vorticity equation (36.43) has precisely the same mathematical form as the material line element equation (14.19)

$$\frac{D(\delta \mathbf{x})}{Dt} = (\delta \mathbf{x} \cdot \nabla) \mathbf{v}. \quad (36.45)$$

Consequently, the relation (36.44) holds for all time with Γ a constant. That is, the line element and vorticity evolve according to the same material equation, so the line element and vorticity forever maintain the relation $\delta\mathbf{x} = \Gamma\boldsymbol{\omega}$. Our only qualification to this identification is to, again, note that vorticity is affected only by the rate of strain, so that particles following a vortex line are unaffected by the rotation tensor. Even so, these particles remain attached to the vortex line thus motivating one to say that vorticity is a “frozen-in” property, with an illustration given by Figure 36.8. Although this property holds only for the case of an incompressible, inviscid, barotropic fluid, it offers insight into the more general situation occurring in real fluids.

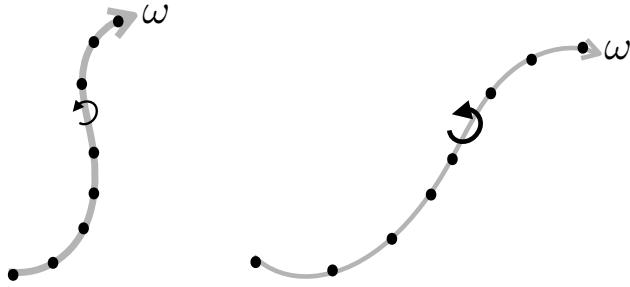


FIGURE 36.8: For an incompressible perfect barotropic fluid, vortex lines (also known as vortex filaments) are also material lines. This property means that for an arbitrary vortex line drawn in the fluid, the fluid particles that are initially on the vortex line will remain on the line as it moves through the fluid. We here show two instances of the same vortex line along with sample test fluid particles. The left configuration stretches into the right configuration, with the vorticity increasing as the vortex line stretches according to the discussion in Section 36.6.3. The material property of vortex lines is known as their *frozen-in nature*. The frozen-in nature of vortex lines strictly holds only for perfect incompressible barotropic fluid, yet it offers useful insight into vorticity dynamics for more general fluids.

36.6.3 Stretching and tilting of vortex tubes

Vorticity responds when vortex lines or tubes are stretched or bent. To help understand the response, consider again the perfect fluid barotropic incompressible vorticity equation (36.43) and focus just the vertical vorticity component

$$\frac{D\omega^z}{Dt} = \omega^x \frac{\partial w}{\partial x} + \omega^y \frac{\partial w}{\partial y} + \omega^z \frac{\partial w}{\partial z}. \quad (36.46)$$

Note that since $\nabla \cdot \boldsymbol{\omega} = 0$, we can write this equation as

$$\frac{D\omega^z}{Dt} = \nabla \cdot (\boldsymbol{\omega} \boldsymbol{\omega}), \quad (36.47)$$

though in the following we find it more useful to focus on the form given by equation (36.46). The following discussion closely emulates that given for a material line element in Section 14.2.6.

Stretching

Consider the vortex tube to be initially aligned with the (vertical) \hat{z} -axis, so that $\omega^x = \omega^y = 0$, in which case there is only a single term impacting vertical vorticity⁴

$$\frac{D\omega^z}{Dt} = \omega^z \frac{\partial w}{\partial z}. \quad (36.48)$$

⁴Be mindful to distinguish the symbols for the vertical component of vorticity, ω^z , and the vertical component of velocity, w .

Since the fluid is incompressible, the volume of an infinitesimal portion of the vortex tube is materially constant

$$\frac{D(\delta V)}{Dt} = 0, \quad (36.49)$$

which means that the vertical extent, δz , and cross-sectional area, δA , are constrained

$$\frac{1}{\delta z} \frac{D(\delta z)}{Dt} + \frac{1}{\delta A} \frac{D(\delta A)}{Dt} = 0. \quad (36.50)$$

As the tube stretches vertically, its horizontal area reduces, and vice versa. Making use of the expression for the evolution of a material line segment (equation (36.45)) allows us to write

$$\frac{1}{\delta z} \frac{D(\delta z)}{Dt} = \frac{\partial w}{\partial z}, \quad (36.51)$$

so that the vorticity equation (36.48) becomes

$$\frac{D\omega^z}{Dt} = \omega^z \frac{\partial w}{\partial z} \quad (36.52a)$$

$$= \omega^z \left[\frac{1}{\delta z} \frac{D(\delta z)}{Dt} \right] \quad (36.52b)$$

$$= -\omega^z \left[\frac{1}{\delta A} \frac{D(\delta A)}{Dt} \right]. \quad (36.52c)$$

Rearrangement leads to

$$\frac{D(\omega^z \delta A)}{Dt} = 0, \quad (36.53)$$

which is an expression of Kelvin's circulation theorem (equation (36.15)) for a cross-section of the vortex tube.

The above manipulations suggest the following interpretation for the *stretching* term, $\omega^z (\partial w / \partial z)$, appearing in the vertical vorticity equation (36.46) and illustrated in Figure 36.9. Namely, as the vortex tube is stretched and its cross-sectional area is compressed, the vorticity magnitude increases so to maintain a constant circulation around the tube, as per Kelvin's theorem (or equivalently as per Helmholtz's first theorem discussed in Section 36.2.3). Stretching a vortex tube increases the magnitude of the vorticity in the direction of the stretching whereas compressing a tube reduces the vorticity magnitude. This result also accords with our understanding of angular momentum conservation as discussed for the rotating cylinder in Section 35.3.2 and depicted by Figure 35.1.

Tilting

Now consider an initially horizontal vortex tube as in the lower left panel of Figure 36.9 so that $\omega^z = 0$. Focus on just one of the two horizontal directions, so that equation (36.46) for the vertical vorticity becomes

$$\frac{D\omega^z}{Dt} = \omega^x \frac{\partial w}{\partial x}. \quad (36.54)$$

If there is no horizontal shear in the vertical velocity ($\partial w / \partial x = 0$), then the vortex tube remains horizontal. However, in the presence of $\partial w / \partial x \neq 0$, the vorticity vector picks up a nonzero vertical projection. To help visualize this process, recall the frozen-in nature of vortex lines, and consider the evolution of an infinitesimal line segment on the vortex tube. With the vortex tube initially

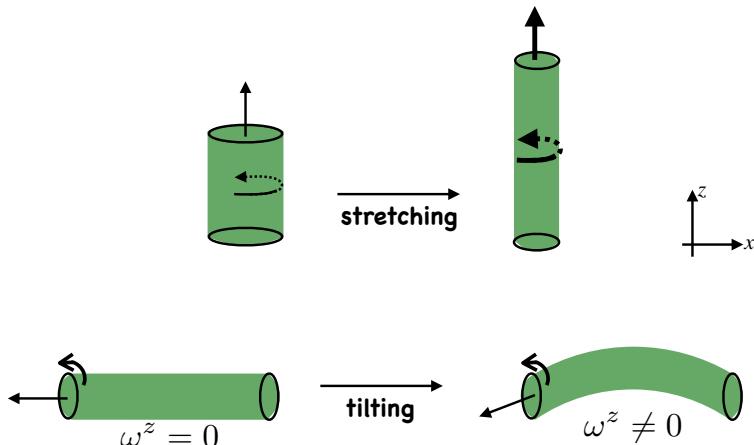


FIGURE 36.9: Illustrating how stretching and tilting of a vortex tube impacts on the vorticity. Top panels: As the cross-sectional area of the vortex tube shrinks, and the vertical extent of the tube stretches, the magnitude of the vorticity along the axis of the tube increases. This result accords with our understanding of angular momentum conservation as discussed for the rotating cylinder in Section 35.3.2 and depicted by Figure 35.1, as well as with Helmholtz's first theorem in Section 36.2.3 and Figure 36.2. Lower panels: The initial vortex tube is assumed to be aligned parallel to the x -axis, so that it has zero projection in the vertical direction. A horizontal shear of the vertical velocity ($\partial w / \partial x \neq 0$) deforms the vortex tube. Upon deforming (or tilting), the tube picks up a nonzero projection in the vertical, which means that it now has a nonzero vertical component to vorticity.

aligned parallel to the x -axis, the evolution of a material line segment (equation (36.45)) is given by

$$\frac{D(\delta\mathbf{x})}{Dt} = \delta\mathbf{x} \frac{\partial\mathbf{v}}{\partial x}. \quad (36.55)$$

The initially horizontal line segment thus picks up a projection in the vertical so long as $\partial w / \partial x \neq 0$. Correspondingly, the vorticity picks up a vertical component. We can think of this process as a tilting or deforming of the initially horizontal vortex tube, with the tilted tube having a nonzero vertical projection.

36.6.4 Comments

As a vortex tube is stretched in the presence of straining motion, it spins faster and its radius reduces. Hence, its mechanical energy moves from larger to smaller spatial scales. This process of downscale energy cascade (i.e., the movement of mechanical energy from large to small scales) is a fundamental property of three dimensional turbulence, and vortex stretching is the dominant mechanism for the cascade. In contrast, two dimensional turbulence, which occurs in horizontal non-divergent flows, does not support vortex stretching and consequently does not support the downscale energy cascade. Instead, two dimensional turbulence supports an inverse cascade whereby there is a net flow of energy to larger scales, with that flow related to the material conservation of vorticity in two dimensional non-divergent flows (see Chapter 34). We have little more to say about the turbulent cascade in this book, with the interested reader encouraged to study the presentations given by [Vallis \(2017\)](#).

36.7 Circulation and vorticity for rotating fluids

The previous sections focused on circulation and vorticity for non-rotating fluids. The discussion here for rotating fluids is a straightforward extension. Even so, the implications are quite profound for the motion of geophysical fluids.

We start by recalling the expression from Section 10.9.1 for the inertial or absolute velocity (i.e., velocity measured in an inertial frame)

$$\mathbf{v}_a = \mathbf{v} + \boldsymbol{\Omega} \wedge \mathbf{x}, \quad (36.56)$$

where \mathbf{v} is the velocity measured in the rotating frame (relative velocity), and \mathbf{x} is the position vector relative to the origin (e.g., center of earth). The absolute circulation around an arbitrary circuit (a circuit that is not necessarily material) is thus given by

$$\mathcal{C}_a = \oint_{\partial S} (\mathbf{v} + \boldsymbol{\Omega} \wedge \mathbf{x}) \cdot d\mathbf{r} = \mathcal{C} + \mathcal{C}_{\text{planet}}, \quad (36.57)$$

where the circulation measured in the rotating frame is

$$\mathcal{C} = \oint_{\partial S} \mathbf{v} \cdot d\mathbf{r} \quad (36.58)$$

and the circulation associated with the rotating planet is

$$\mathcal{C}_{\text{planet}} = \oint_{\partial S} (\boldsymbol{\Omega} \wedge \mathbf{x}) \cdot d\mathbf{r}. \quad (36.59)$$

Again, $d\mathbf{r}$ is the differential line element moving around the circuit as mentioned in Section 33.3.

A fluid particle at rest in the rotating frame will still have a nonzero absolute circulation as given by the planetary circulation. Making use of Stokes' theorem leads to the equivalent forms for the circulations

$$\mathcal{C} = \oint_{\partial S} \mathbf{v} \cdot d\mathbf{r} = \int_S \boldsymbol{\omega} \cdot \hat{\mathbf{n}} dS \quad \text{relative circulation} \quad (36.60a)$$

$$\mathcal{C}_{\text{planet}} = \oint_{\partial S} (\boldsymbol{\Omega} \wedge \mathbf{x}) \cdot d\mathbf{r} = \int_S \boldsymbol{\omega}_{\text{planet}} \cdot \hat{\mathbf{n}} dS \quad \text{planetary circulation} \quad (36.60b)$$

$$\mathcal{C}_a = \oint_{\partial S} \mathbf{v}_a \cdot d\mathbf{r} = \int_S \boldsymbol{\omega}_a \cdot \hat{\mathbf{n}} dS \quad \text{absolute circulation,} \quad (36.60c)$$

where

$$\boldsymbol{\omega} = \nabla \wedge \mathbf{v} \quad \text{relative vorticity} \quad (36.61a)$$

$$\boldsymbol{\omega}_{\text{planet}} = \nabla \wedge (\boldsymbol{\Omega} \wedge \mathbf{x}) = 2\boldsymbol{\Omega} \quad \text{planetary vorticity} \quad (36.61b)$$

$$\boldsymbol{\omega}_a = \boldsymbol{\omega} + \boldsymbol{\omega}_{\text{planet}} \quad \text{absolute vorticity.} \quad (36.61c)$$

Thus far we have merely substituted in the expression (36.56) for the inertial velocity and then decomposed the vorticity and circulation into its relative and planetary components. Next we consider how circulation evolves, in which case we will see how the relative and planetary circulations interact.

36.7.1 Material evolution of absolute circulation

We now consider how the absolute circulation evolves for a material circuit that moves with the fluid

$$\frac{dC_a}{dt} = \frac{d}{dt} \oint_{\partial S(v)} (\mathbf{v} + \boldsymbol{\Omega} \wedge \mathbf{r}) \cdot d\mathbf{r}, \quad (36.62)$$

where we choose to write $\mathbf{r} = \mathbf{x}$ for the position of a fluid particle on the circuit, and $d\mathbf{r}$ is a differential line element around a circuit of material fluid particles. We measure fluid motion in the rotating frame so that the material time derivative contains advection by the velocity \mathbf{v} rather than the absolute velocity \mathbf{v}_a . Following the discussion for non-rotating Kelvin's circulation theorem in Section 36.3 leads to

$$\frac{dC_a}{dt} = \frac{d}{dt} \oint_{\partial S(v)} (\mathbf{v} + \boldsymbol{\Omega} \wedge \mathbf{r}) \cdot d\mathbf{r} \quad (36.63a)$$

$$= \oint_{\partial S(v)} \left[\frac{D\mathbf{v}}{Dt} + \boldsymbol{\Omega} \wedge \frac{D\mathbf{r}}{Dt} \right] \cdot d\mathbf{r} + \oint_{\partial S(v)} (\mathbf{v} + \boldsymbol{\Omega} \wedge \mathbf{r}) \cdot d\mathbf{v} \quad (36.63b)$$

$$= \oint_{\partial S(v)} \left[\frac{D\mathbf{v}}{Dt} + \boldsymbol{\Omega} \wedge \mathbf{v} \right] \cdot d\mathbf{r} + \oint_{\partial S(v)} (\boldsymbol{\Omega} \wedge \mathbf{r}) \cdot d\mathbf{v} \quad (36.63c)$$

$$= \oint_{\partial S(v)} \left[\frac{D\mathbf{v}}{Dt} + 2\boldsymbol{\Omega} \wedge \mathbf{v} \right] \cdot d\mathbf{r}. \quad (36.63d)$$

To reach this result we set

$$\mathbf{v} = \frac{D\mathbf{r}}{Dt}, \quad (36.64)$$

for the velocity of a fluid particle on the circuit. We also used the identity

$$\oint_{\partial S(v)} \mathbf{v} \cdot d\mathbf{v} = \frac{1}{2} \oint_{\partial S(v)} d\mathbf{v}^2 = 0 \quad (36.65)$$

as well as

$$\oint_{\partial S(v)} (\boldsymbol{\Omega} \wedge \mathbf{r}) \cdot d\mathbf{v} = \oint_{\partial S(v)} d[(\boldsymbol{\Omega} \wedge \mathbf{r}) \cdot \mathbf{v}] - \oint_{\partial S(v)} (\boldsymbol{\Omega} \wedge d\mathbf{r}) \cdot \mathbf{v} = \oint_{\partial S(v)} (\boldsymbol{\Omega} \wedge \mathbf{v}) \cdot d\mathbf{r}. \quad (36.66)$$

Now insert the momentum equation (36.19) into equation (36.63d) to yield

$$\frac{dC_a}{dt} = \oint_{\partial S(v)} \left[\frac{D\mathbf{v}}{Dt} + 2\boldsymbol{\Omega} \wedge \mathbf{v} \right] \cdot d\mathbf{r}. \quad (36.67a)$$

$$= \oint_{\partial S(v)} \left[-\frac{1}{\rho} \nabla p - \nabla \Phi + \mathbf{F} \right] \cdot d\mathbf{r}. \quad (36.67b)$$

$$= \oint_{\partial S(v)} \left[-\frac{dp}{\rho} + \mathbf{F} \cdot d\mathbf{r} \right]. \quad (36.67c)$$

Making use of Stokes' theorem leads to the evolution of absolute circulation around a material loop

$$\frac{dC_a}{dt} = \oint_{\partial S(v)} \left[-\frac{dp}{\rho} + \mathbf{F} \cdot d\mathbf{r} \right] = \int_{S(v)} (\mathbf{B} + \nabla \wedge \mathbf{F}) \cdot \hat{\mathbf{n}} d\mathcal{S}, \quad (36.68)$$

where $\mathbf{B} = \rho^{-2} \nabla p \wedge \nabla p$ is the baroclinicity vector from equation (36.31).

The circulation theorem (36.68) is the same as obtained for the non-rotating Kelvin's circulation theorem discussed in Section 36.3 (see equation (36.12e)). This equivalence proves that absolute circulation is an objective (frame invariant) property of the fluid, in which its evolution is unchanged when moving to a non-inertial rotating frame.

36.7.2 The beta effect

As given by equation (36.57), the absolute circulation around an arbitrary circuit equals to the circulation of fluid measured in the rotating frame (relative circulation) plus circulation of the rotating frame itself (planetary circulation)

$$\mathcal{C}_a = \mathcal{C} + \mathcal{C}_{\text{planet}} = \mathcal{C} + 2 \int_{\mathcal{S}} \boldsymbol{\Omega} \cdot \hat{\mathbf{n}} d\mathcal{S} \iff \frac{d\mathcal{C}_a}{dt} = \frac{d\mathcal{C}}{dt} + \frac{d\mathcal{C}_{\text{planet}}}{dt}. \quad (36.69)$$

We can determine the processes that affect the absolute circulation around a material loop by using the circulation equation (36.68)

$$\frac{d\mathcal{C}}{dt} = -\frac{d\mathcal{C}_{\text{planet}}}{dt} + \frac{d\mathcal{C}_a}{dt} \quad (36.70a)$$

$$= -2 \frac{d}{dt} \left[\int_{\mathcal{S}(\mathbf{v})} \boldsymbol{\Omega} \cdot \hat{\mathbf{n}} d\mathcal{S} \right] + \int_{\mathcal{S}(\mathbf{v})} (\mathbf{B} + \nabla \wedge \mathbf{F}) \cdot \hat{\mathbf{n}} d\mathcal{S}. \quad (36.70b)$$

We generally assume that the planetary rotation is a constant in time and points through the north pole of the sphere⁵ $\boldsymbol{\Omega} = \Omega \hat{\mathbf{Z}}$, so that

$$\int_{\mathcal{S}(\mathbf{v})} \boldsymbol{\Omega} \cdot \hat{\mathbf{n}} d\mathcal{S} = \Omega \int_{\mathcal{S}(\mathbf{v})} \hat{\mathbf{Z}} \cdot \hat{\mathbf{n}} d\mathcal{S} = \Omega A_{\perp}. \quad (36.71)$$

The area A_{\perp} is the projection of the spherical area enclosed by the circuit onto the horizontal equatorial plane, with Figure 36.10 illustrating the geometry. This result has profound impact on large scale geophysical fluid motion, whereby relative circulation around a material circuit in the rotating frame changes according to

$$\frac{d\mathcal{C}}{dt} = \underbrace{-2\Omega \frac{dA_{\perp}}{dt}}_{\text{beta effect}} + \underbrace{\int_{\mathcal{S}} (\mathbf{B} + \nabla \wedge \mathbf{F}) \cdot \hat{\mathbf{n}} d\mathcal{S}(\mathbf{v})}_{\text{solenoids plus friction}}. \quad (36.72)$$

Equation (36.72) is sometimes referred to as the *Bjerknes circulation theorem* (see Holton (1992) equation (4.5)). The second term, comprised of baroclinicity and friction, also appears in the non-rotating case. It has already been discussed in Sections 36.3 and 36.5.

The first term in the circulation theorem (36.72) is fundamentally new. It is nonzero in the presence of both rotation and curvature of the sphere. The spherical effect arises from latitudinal movement of a material circuit, with the area, A_{\perp} , changing under such motion. When the circuit moves poleward, the projected area A_{\perp} increases whereas it decreases to zero as it moves equatorward. The material change in A_{\perp} , when multiplied by the planetary vorticity, modifies the relative circulation around the material circuit. We refer to *planetary induction* as the process whereby relative circulation changes due to latitudinal motion of a material circuit on a rotating sphere. Or more commonly, planetary induction is referred to as the *beta effect*, given its connection to the latitudinal gradient of the Coriolis parameter, $\beta = \partial_y f$. Notably, longitudinal motion of the circuit has no impact on A_{\perp} , so that longitudinal motion imparts no planetary induction of relative circulation.

In theories of large-scale laminar planetary flows, the baroclinicity and friction terms are typically sub-dominant. For these flows, the material evolution of relative circulation is dominated

⁵We follow the notational conventions of Figure 8.1 with one exception. Here, the vertical Cartesian direction through the north pole is written $\hat{\mathbf{Z}}$ to avoid confusion with the local vertical direction $\hat{\mathbf{z}}$ determined by the geopotential.

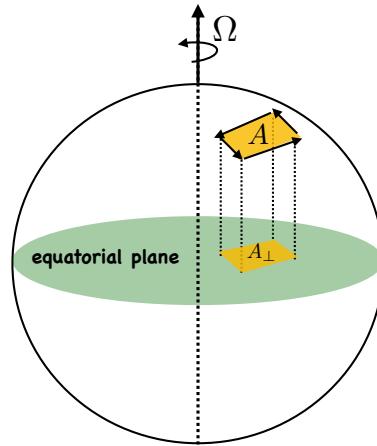


FIGURE 36.10: Geometry of the beta effect. According to the Bjerknes circulation theorem (36.72), the circulation for a material loop on the surface of a rotating sphere is affected by baroclinicity and friction, as for a non-rotating sphere, as well as latitudinal motion of the loop. The latitudinal motion alters the area of the loop as projected onto the equatorial plane, with the projected area increasing as the loop moves poleward. When multiplied by the magnitude of the planetary vorticity, 2Ω , the area contribution is termed *planetary induction* (i.e., relative circulation is induced by latitudinal motion), or more commonly it is called the *beta effect*. The beta effect requires both rotation (2Ω) and curvature of the sphere ($\partial_y f = \beta$); it is therefore absent on the *f*-plane.

by the beta effect. Planetary geostrophic flow is the canonical example of such flow, as studied in Section 26.5 as well as Chapters 40 and 41. In such flows, forces that lead to meridional motion also give rise to changes in the relative circulation. Conversely, forces that change the circulation around a material loop in turn affect meridional motion of the loop.

36.7.3 A two-dimensional incompressible fluid

To garner more insight into the beta effect, consider a perfect incompressible and two-dimensional fluid (zero vertical velocity) on a rotating sphere. In this case there is only a vertical component to vorticity and both baroclinicity and friction have no role in vorticity dynamics. Hence, vorticity is affected only via the beta effect. This *two-dimensional barotropic fluid* system is discussed in more detail in Chapter 34. For now, we use it as a toy example to help expose the essence of the beta effect (see also Section 34.3.2).

In the rotating frame, circulation around an infinitesimal closed material loop is

$$\mathcal{C} = A \zeta, \quad (36.73)$$

where ζ is the relative vorticity and A is the area enclosed by the loop. Because the fluid is incompressible, the loop area A remains constant even as the loop becomes contorted (see Section 17.6). The material evolution of circulation is therefore given by

$$\frac{D\mathcal{C}}{Dt} = A \frac{D\zeta}{Dt} = -2\Omega \frac{DA_{\perp}}{Dt}, \quad (36.74)$$

where the second equality follows from Bjerknes' circulation theorem (36.72). Let the material circuit be at a latitude ϕ so that the projection of the loop area onto the equatorial plane is (see Figure 36.10)

$$A_{\perp} = A \sin \phi. \quad (36.75)$$

Hence, material evolution of the circulation is

$$\frac{DC}{Dt} = A \frac{D\zeta}{Dt} \quad (36.76a)$$

$$= -2\Omega \frac{DA_{\perp}}{Dt} \quad (36.76b)$$

$$= -2A\Omega \frac{D\sin\phi}{Dt} \quad (36.76c)$$

$$= -2A\Omega \cos\phi \frac{D\phi}{Dt} \quad (36.76d)$$

$$= -A \left[\frac{2\Omega \cos\phi}{R} \right] \left[R \frac{D\phi}{Dt} \right] \quad (36.76e)$$

$$= -A\beta v, \quad (36.76f)$$

where we introduced the meridional velocity component

$$v = R \frac{D\phi}{Dt} \quad (36.77)$$

and the meridional derivative of the planetary vorticity

$$\beta = \frac{df}{dy} = \frac{1}{R} \frac{d}{d\phi} (2\Omega \sin\phi) = \frac{2\Omega \cos\phi}{R}. \quad (36.78)$$

The result (36.76f) shows how meridional motion on a rotating sphere induces relative circulation. It furthermore motivates the name *beta effect* for planetary induction of relative vorticity.

36.7.4 Further study

The beta effect and its role in vorticity is nicely summarized in [this video from Science Primer](#) in the context of Rossby waves.

36.8 Vorticity in a hydrostatic and Boussinesq fluid

In this section we develop the vorticity budget for a hydrostatic Boussinesq fluid in the presence of diabatic sources and frictional forcing. This system is of particular importance for models of the ocean circulation with the governing equations given by (see Section 24.1.6)

$$\frac{Du}{Dt} + \mathbf{f} \wedge \mathbf{v} = -\nabla_z \varphi + \mathbf{F} \quad (36.79a)$$

$$\frac{\partial \varphi}{\partial z} = b \quad (36.79b)$$

$$\nabla \cdot \mathbf{v} = 0 \quad (36.79c)$$

$$\frac{Db}{Dt} = \dot{b}, \quad (36.79d)$$

with the non-divergent velocity field written

$$\mathbf{v} = (\mathbf{u}, w) = \mathbf{u} + w \hat{z}. \quad (36.80)$$

The perturbation pressure is given by

$$\rho_0 \varphi = \delta p = p - p_0, \quad (36.81)$$

with the reference pressure, $p_0 = p_0(z)$, in hydrostatic balance with the constant reference density

$$\frac{dp_0}{dz} = -g \rho_0 \quad (36.82)$$

and p the hydrostatic pressure satisfying the hydrostatic balance

$$\frac{\partial p}{\partial z} = -g \rho. \quad (36.83)$$

The globally referenced Archimedean buoyancy is given by

$$b = -g \left[\frac{\rho - \rho_0}{\rho_0} \right], \quad (36.84)$$

with this field introduced in Section 25.4.2. We assume the Coriolis parameter of the form $\mathbf{f} = f \hat{z}$, in which case

$$\mathbf{f} \wedge \mathbf{v} = \mathbf{f} \wedge \mathbf{u}. \quad (36.85)$$

Finally, the horizontal friction acceleration vector is given by

$$\mathbf{F} = (F^x, F^y, 0) \quad (36.86)$$

and the gradient operator is

$$\nabla = \nabla_z + \hat{z} \partial_z. \quad (36.87)$$

36.8.1 Deriving the vorticity equation

To derive the vorticity equation, it is useful to combine the horizontal momentum equation with the hydrostatic balance, in which case

$$\frac{D\mathbf{u}}{Dt} + \mathbf{f} \wedge \mathbf{v} = -\nabla \varphi + b \hat{z} + \mathbf{F}. \quad (36.88)$$

As for the non-hydrostatic case (Section 36.4.1), we rewrite the self-advection operator, $(\mathbf{v} \cdot \nabla) \mathbf{u}$, before taking the curl. In turn, we introduce the hydrostatic relative vorticity given by the curl of the horizontal velocity

$$\boldsymbol{\omega}^{\text{hy}} = \nabla \wedge \mathbf{u} = \hat{z} \wedge \partial_z \mathbf{u} + \hat{z} \zeta = -\hat{x} \partial_z v + \hat{y} \partial_z u + \hat{z} \zeta, \quad (36.89)$$

where $\zeta = \partial_x v - \partial_y u$ is the vertical component to the relative vorticity, and

$$\nabla \cdot \boldsymbol{\omega}^{\text{hy}} = 0. \quad (36.90)$$

It is then straightforward to show that

$$\boldsymbol{\omega}^{\text{hy}} \wedge \mathbf{v} = \hat{x} (w \partial_z u - v \partial_x v + v \partial_y u) + \hat{y} (w \partial_z v - u \partial_y u + u \partial_x v) - \hat{z} \partial_z (u^2 + v^2)/2 \quad (36.91a)$$

$$= w \partial_z \mathbf{u} + \zeta (-v \hat{x} + u \hat{y}) - \hat{z} \partial_z (u^2 + v^2)/2, \quad (36.91b)$$

in which case

$$\nabla(\mathbf{u}^2/2) + \boldsymbol{\omega}^{\text{hy}} \wedge \mathbf{v} = \nabla(u^2 + v^2)/2 - \hat{z} \partial_z (u^2 + v^2)/2 + w \partial_z \mathbf{u} + \zeta (-v \hat{x} + u \hat{y}) \quad (36.92a)$$

$$= (u \partial_x + v \partial_y + w \partial_z) \mathbf{u} \quad (36.92b)$$

$$= (\mathbf{v} \cdot \nabla) \mathbf{u}. \quad (36.92c)$$

The material time derivative of the horizontal velocity can thus be written as

$$\frac{D\mathbf{u}}{Dt} = \frac{\partial \mathbf{u}}{\partial t} + (\mathbf{v} \cdot \nabla) \mathbf{u} = \frac{\partial \mathbf{u}}{\partial t} + \boldsymbol{\omega}^{\text{hy}} \wedge \mathbf{v} + \nabla(\mathbf{u}^2/2), \quad (36.93)$$

which then leads to the vector invariant horizontal velocity equation

$$\frac{\partial \mathbf{u}}{\partial t} + (f \hat{\mathbf{z}} + \boldsymbol{\omega}^{\text{hy}}) \wedge \mathbf{v} = -\nabla(\varphi + \mathbf{u}^2/2) + b \hat{\mathbf{z}} + \mathbf{F}, \quad (36.94)$$

which can be written in the equivalent form

$$\frac{\partial \mathbf{u}}{\partial t} + w \frac{\partial \mathbf{u}}{\partial z} + (f + \zeta) \hat{\mathbf{z}} \wedge \mathbf{u} = -\nabla_z(\varphi + \mathbf{u}^2/2) - (\partial_z \varphi - b) \hat{\mathbf{z}} + \mathbf{F}. \quad (36.95)$$

Now take the curl of the vector invariant velocity equation (36.94), and make use of the identity

$$\nabla \wedge (\boldsymbol{\omega}_a^{\text{hy}} \wedge \mathbf{v}) = (\mathbf{v} \cdot \nabla) \boldsymbol{\omega}_a^{\text{hy}} - (\boldsymbol{\omega}_a^{\text{hy}} \cdot \nabla) \mathbf{v}, \quad (36.96)$$

where we introduced the absolute vorticity for a hydrostatic fluid

$$\boldsymbol{\omega}_a^{\text{hy}} = f \hat{\mathbf{z}} + \boldsymbol{\omega}^{\text{hy}}. \quad (36.97)$$

The result is the vorticity equation

$$\frac{\partial \boldsymbol{\omega}^{\text{hy}}}{\partial t} + (\mathbf{v} \cdot \nabla) \boldsymbol{\omega}_a^{\text{hy}} = (\boldsymbol{\omega}_a^{\text{hy}} \cdot \nabla) \mathbf{v} + \nabla \wedge \hat{\mathbf{z}} b + \nabla \wedge \mathbf{F}. \quad (36.98)$$

Since the Coriolis parameter is time independent, we can add it to the time derivative to yield

$$\frac{D\boldsymbol{\omega}_a^{\text{hy}}}{Dt} = \underbrace{(\boldsymbol{\omega}_a^{\text{hy}} \cdot \nabla) \mathbf{v}}_{\text{stretching} + \text{tilting}} + \underbrace{\nabla \wedge \hat{\mathbf{z}} b}_{\text{baroclinicity}} + \underbrace{\nabla \wedge \mathbf{F}}_{\text{friction curl}}. \quad (36.99)$$

36.8.2 Boussinesq form of the baroclinicity vector

Equation (36.99) is the vorticity equation for a hydrostatic Boussinesq fluid. We can compare this equation to the vorticity equation for a non-hydrostatic and non-Boussinesq fluid (equation (36.30)). Both equations have a vorticity source due to stretching and tilting, and both have a source due to the curl of friction. However, the baroclinicity vector for the Boussinesq fluid is given by

$$\mathbf{B}_{\text{bouss}} = \nabla \wedge \hat{\mathbf{z}} b = \nabla b \wedge \hat{\mathbf{z}}, \quad (36.100)$$

which is simpler than baroclinicity in a compressible fluid

$$\mathbf{B} = (\nabla \rho \wedge \nabla p) / \rho^2 \quad (36.101)$$

(equation (36.31)). We see in Exercise 38.2 that the same baroclinicity vector (36.100) appears for the non-hydrostatic Boussinesq fluid. So in general, we can diagnose the presence of baroclinicity for the Boussinesq fluid merely by noting whether there is a slope to the buoyancy surfaces relative to the horizontal, such as in Figure 36.11. That is, a sloping buoyancy surface acts as a vorticity source for the Boussinesq fluid by providing a spin to fluid elements (see Section 36.5).

Baroclinicity in a Boussinesq fluid has no direct impact on the vertical component to absolute vorticity since

$$\hat{\mathbf{z}} \cdot \mathbf{B}_{\text{bouss}} = \hat{\mathbf{z}} \cdot (\nabla b \wedge \hat{\mathbf{z}}) = 0. \quad (36.102)$$

Hence, baroclinicity in a Boussinesq fluid only acts directly as a source for horizontal vorticity. We can understand this property by considering a buoyancy configuration such as shown in Figure 36.11. Focusing on a region where $\nabla b = |\nabla b| \hat{\mathbf{y}}$, we have $\mathbf{B}_{\text{bouss}} = |\nabla b| \hat{\mathbf{x}}$, so that baroclinicity generates a zonally oriented vorticity, $\boldsymbol{\omega} \cdot \hat{\mathbf{x}} > 0$, in this region.

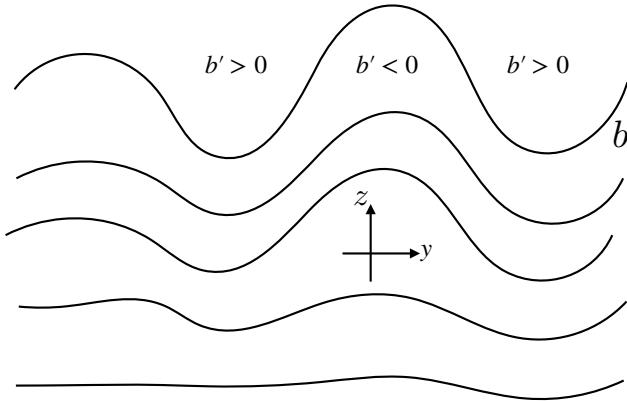


FIGURE 36.11: Baroclinicity in a Boussinesq fluid is manifest by nonzero horizontal gradients in the buoyancy field. Here we depict a region of relatively strong baroclinicity above a region of weaker baroclinicity. A sloping buoyancy surface is therefore synonymous with a nontrivial baroclinic structure. We label anomalously positive ($b' > 0$) and negative buoyancy ($b' < 0$), where the prime denotes anomalies relative to a horizontal average. Furthermore, as per equation (36.102), baroclinicity in a Boussinesq fluid only acts as a source to horizontal vorticity.

36.8.3 Vertical vorticity equation

The vertical component of the vorticity equation (36.99) takes the form

$$\frac{D\zeta_a}{Dt} = (\boldsymbol{\omega}_a^{\text{hy}} \cdot \nabla) w + \hat{\mathbf{z}} \cdot (\nabla \wedge \mathbf{F}), \quad (36.103)$$

with the absence of baroclinicity noted above in Section 36.8.2. The stretching, tilting, and friction curl appearing on the right hand side provide vorticity sources that affect the left hand side's material time evolution. We see this evolution more fully by expanding the terms to render

$$\frac{\partial \zeta_a}{\partial t} + (\mathbf{v} \cdot \nabla) \zeta + \beta v = \hat{\mathbf{z}} \cdot (\partial_z \mathbf{u} \wedge \nabla_z w) + (\zeta + f) \partial_z w + \hat{\mathbf{z}} \cdot (\nabla \wedge \mathbf{F}). \quad (36.104)$$

Planetary geostrophic limit

The linearized, inviscid, and steady version of the vorticity equation (36.104) leads to the linear vorticity balance studied in Section 26.5.4 and Chapter 41

$$\beta v = f \partial_z w \quad (36.105)$$

with this relation comprising the inviscid vorticity equation for a flow respecting planetary geostrophic scaling. It represents a kinematic balance since no forces are exposed here to explicitly cause motion, though such forces do appear in the momentum equation. Reading the balance from right to left indicates that any process generating vorticity via vortex stretching must be balanced by meridional motion. That is, the fluid responds to vortex stretching by moving meridionally through the planet's vorticity field. Since the vorticity of a planetary geostrophic fluid is solely determined by planetary vorticity, meridional movement is the only means for the fluid to balance vortex sources. Conversely, reading the equality from left to right reveals that any meridional motion itself must be balanced by vortex stretching.

Vorticity flux vector

We can write the vorticity equation (36.104) in an alternative form by making use of $\nabla \cdot \mathbf{v} = \nabla \cdot \boldsymbol{\omega}_a^{hy} = 0$ to yield

$$\frac{\partial \zeta_a}{\partial t} = -\nabla \cdot (\mathbf{v} \zeta_a - w \boldsymbol{\omega}_a^{hy}) + \hat{\mathbf{z}} \cdot (\nabla \wedge \mathbf{F}). \quad (36.106)$$

Furthermore, we can write the friction curl in the form

$$\hat{\mathbf{z}} \cdot (\nabla \wedge \mathbf{F}) = \nabla z \cdot (\nabla \wedge \mathbf{F}) \quad (36.107a)$$

$$= \nabla \cdot [z \nabla \wedge \mathbf{F}] \quad (36.107b)$$

$$= \nabla \cdot [\nabla \wedge (z \mathbf{F}) - \nabla z \wedge \mathbf{F}] \quad (36.107c)$$

$$= -\nabla \cdot (\hat{\mathbf{z}} \wedge \mathbf{F}). \quad (36.107d)$$

Hence, the vertical component of the Boussinesq vorticity evolves according to the convergence of the vorticity flux

$$\frac{\partial \zeta_a}{\partial t} = -\nabla \cdot \mathbf{J}_\zeta \quad \text{with} \quad \mathbf{J}_\zeta = \mathbf{v} \zeta_a - w \boldsymbol{\omega}_a^{hy} + \hat{\mathbf{z}} \wedge \mathbf{F}. \quad (36.108)$$

The identity $\boldsymbol{\omega}_a^{hy} = \hat{\mathbf{z}} \wedge \partial_z \mathbf{u} + \hat{\mathbf{z}} \zeta_a$ allows us to write

$$\mathbf{v} \zeta_a - w \boldsymbol{\omega}_a^{hy} = \mathbf{u} \zeta_a - w \hat{\mathbf{z}} \wedge \partial_z \mathbf{u}, \quad (36.109)$$

which is a horizontal vector. Furthermore, note that $\hat{\mathbf{z}} \wedge \mathbf{F}$ is a horizontal vector, which then means that there is no vertical contribution to the vorticity flux vector, $\mathbf{J}_\zeta \cdot \hat{\mathbf{z}} = 0$. Even so, as seen in the discussion of depth integrated vorticity in Section 36.9, boundary conditions are more seamlessly handled by retaining the full form of this flux given by equation (36.108).

36.9 Depth integral of the vertical vorticity equation

In this section we study the depth integral of the vorticity equation (36.108) for a Boussinesq and hydrostatic fluid. The depth integral is taken over the full depth from the ocean bottom at $z = \eta_b(x, y)$ to the ocean surface at $z = \eta(x, y, t)$ (see Figure 31.1). Studies of the depth integrated vorticity equation allow us to focus on the two dimensional budgets with particular attention to how boundary torques alter the budget. This section anticipates our analysis of the depth integrated planetary geostrophic vorticity equation in Section 41.4, with that analysis of use for understanding the role of topography in forcing the large-scale ocean circulation.

36.9.1 Leibniz rule expressions

The necessary manipulations are typical for the analysis of depth integrated budgets, such as considered for the depth integrated momentum in Section 36.10. For vorticity we are interested in the following equation

$$\int_{\eta_b}^{\eta} \frac{\partial \zeta_a}{\partial t} dz = - \int_{\eta_b}^{\eta} \nabla \cdot \mathbf{J}_\zeta dz, \quad (36.110)$$

making use of Leibniz's rule (Section 16.3.4) to move the time and space derivatives from inside the integrals to outside

$$\int_{\eta_b}^{\eta} \frac{\partial \zeta_a}{\partial t} dz = -[\zeta_a \partial_t \eta]_{z=\eta} + \frac{\partial}{\partial t} \int_{\eta_b}^{\eta} \zeta_a dz \quad (36.111)$$

$$- \int_{\eta_b}^{\eta} \nabla_z \cdot \mathbf{J}_\zeta dz = [\nabla_z \eta \cdot \mathbf{J}_\zeta]_{z=\eta} - [\nabla_z \eta_b \cdot \mathbf{J}_\zeta]_{z=\eta_b} - \nabla_z \cdot \int_{\eta_b}^{\eta} \mathbf{J}_\zeta dz \quad (36.112)$$

$$- \int_{\eta_b}^{\eta} \frac{\partial (\hat{\mathbf{z}} \cdot \mathbf{J}_\zeta)}{\partial z} dz = -[\hat{\mathbf{z}} \cdot \mathbf{J}_\zeta]_{z=\eta} + [\hat{\mathbf{z}} \cdot \mathbf{J}_\zeta]_{z=\eta_b}. \quad (36.113)$$

These results then lead to

$$\frac{\partial}{\partial t} \int_{\eta_b}^{\eta} \zeta_a dz + \nabla_z \cdot \int_{\eta_b}^{\eta} \mathbf{J}_\zeta dz = [\zeta_a \partial_t \eta - \nabla(z - \eta) \cdot \mathbf{J}_\zeta]_{z=\eta} + [\nabla(z - \eta_b) \cdot \mathbf{J}_\zeta]_{z=\eta_b}. \quad (36.114)$$

The time tendency for the depth integrated absolute vorticity for a fluid column (left hand side) is determined by the convergence of the depth integrated vorticity flux (first term on right hand side) plus a suite of boundary contributions due to vortex stretching and friction. We next massage the boundary contributions to help understand their physical content.

36.9.2 Bottom boundary contribution

The bottom boundary contribution to the vorticity equation (36.114) takes on the form

$$\nabla(z - \eta_b) \cdot \mathbf{J}_\zeta = \nabla(z - \eta_b) \cdot [\mathbf{v} \zeta_a - w \omega_a^{hy} + \hat{\mathbf{z}} \wedge \mathbf{F}]_{z=\eta_b} \quad (36.115a)$$

$$= -|\nabla(z - \eta_b)| \hat{\mathbf{n}} \cdot [-w \omega_a^{hy} + \hat{\mathbf{z}} \wedge \mathbf{F}]_{z=\eta_b}, \quad (36.115b)$$

where we made use of the no-normal flow bottom kinematic boundary condition, $\hat{\mathbf{n}} \cdot \mathbf{v} = 0$ (Section 15.4.1), and where

$$\hat{\mathbf{n}} = - \left[\frac{\nabla(z - \eta_b)}{|\nabla(z - \eta_b)|} \right] = - \left[\frac{\hat{\mathbf{z}} - \nabla_z \eta_b}{\sqrt{1 + \nabla_z \eta_b \cdot \nabla_z \eta_b}} \right] \quad (36.116)$$

is the outward normal at the bottom.

The first term in the bottom boundary flux (36.115b) provides an inviscid vertical transport of the normal component of the absolute vorticity at the boundary. This term contributes through the action of vertical motion next to a sloping bottom, thus providing a vertical transfer of the vorticity component that is perpendicular to the bottom. This motion provides a form of vortex stretching that vanishes for a flat bottom, in which case $w(\eta_b) = 0$. It also vanishes for flow that parallels the bottom, whereby $\mathbf{u} \cdot \nabla \eta_b = 0$ so that $w(\eta_b) = 0$ according to the bottom kinematic boundary condition (15.41).

The second term in the boundary flux (36.115b) is the contribution from friction along the bottom, and it can be written in the form

$$- \hat{\mathbf{n}} \cdot (\hat{\mathbf{z}} \wedge \mathbf{F}) = -\mathbf{F} \cdot (\hat{\mathbf{n}} \wedge \hat{\mathbf{z}}) \equiv -\mathbf{F} \cdot \mathbf{t} \quad \text{with} \quad \mathbf{t} = \hat{\mathbf{n}} \wedge \hat{\mathbf{z}} = \left[\frac{\nabla \eta_b \wedge \hat{\mathbf{z}}}{\sqrt{1 + \nabla \eta_b \cdot \nabla \eta_b}} \right]. \quad (36.117)$$

The vector \mathbf{t} is horizontal and it points along isolines of constant topography in a direction with land to the left pointing in the direction of \mathbf{t} , as depicted in Figure 36.12. Since $\hat{\mathbf{z}}$ and $\hat{\mathbf{n}}$ are not orthogonal, \mathbf{t} is not normalized so that it is not adorned with a hat. To further our understanding of how bottom friction contributes to the depth integrated vorticity, consider bottom frictional

acceleration to be directly proportional to the velocity. The simplest case is with a Rayleigh drag so that $\mathbf{F} = -\gamma \mathbf{u}$ and

$$-\mathbf{F} \cdot \mathbf{t} = \gamma \mathbf{u} \cdot \mathbf{t}, \quad (36.118)$$

with γ an inverse time scale. If the flow is oriented with shallow water to the right; e.g., into the page in Figure 36.12, then $-\mathbf{F} \cdot \mathbf{t} < 0$, thus contributing a negative vorticity tendency. In general, the bottom friction acts to damp the depth integrated vorticity, which is expected since bottom friction does not spontaneously spin-up the flow.

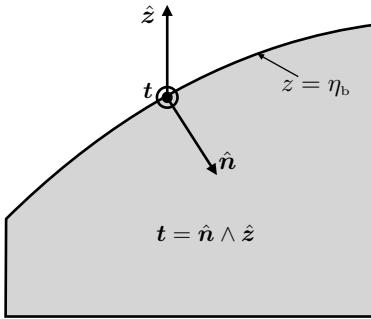


FIGURE 36.12: Orientation of the unit vectors next to the bottom of the fluid. The vertical unit vector, \hat{z} , points vertically upward and the outward unit vector, \hat{n} , points downward into the rock (shaded gray). The along-topography unit vector, $\hat{t} = \hat{n} \wedge \hat{z}$, points along lines of constant topography with land to the left when facing in the direction of \hat{t} ; in this figure it points out from the page.

36.9.3 Surface boundary contribution

The surface boundary contribution to the vorticity equation (36.114) takes on a similar form to the bottom, with the new feature that the free surface is both moving and permeable (Section 15.4.3). This boundary term takes on the form

$$\zeta_a \partial_t \eta - \nabla(z - \eta) \cdot \mathbf{J}_\zeta = \zeta_a \partial_t \eta - \nabla(z - \eta) \cdot [\mathbf{v} \zeta_a - w \omega_a^{\text{hy}} + \hat{z} \wedge \mathbf{F}]_{z=\eta} \quad (36.119a)$$

$$= \zeta_a [\partial_t \eta + \mathbf{u} \cdot \nabla_z \eta - w]_{z=\eta} + \nabla(z - \eta) \cdot [w \omega_a^{\text{hy}} - \hat{z} \wedge \mathbf{F}]_{z=\eta} \quad (36.119b)$$

$$= \zeta_a Q_m / \rho_0 + |\nabla(z - \eta)| \hat{n} \cdot [w \omega_a^{\text{hy}} - \hat{z} \wedge \mathbf{F}]_{z=\eta}, \quad (36.119c)$$

where we made use of the surface kinematic boundary condition (15.79) to introduce the surface mass flux Q_m , and where

$$\hat{n} = \frac{\nabla(z - \eta)}{|\nabla(z - \eta)|} = \frac{\nabla_z \eta + \hat{z}}{\sqrt{1 + \nabla_z \eta \cdot \nabla_z \eta}} \quad (36.120)$$

is the outward normal at the surface. The first term in the surface boundary flux (36.119c) provides transport of boundary vorticity due to the transfer of mass across the boundary. The second term provides an inviscid vertical transport of absolute vorticity at the surface boundary, thus acting as a vortex stretching contribution. The third term provides the corresponding contribution from the friction along the upper surface, with friction acting to reduce the magnitude of the surface boundary vorticity. For a rigid lid surface, $w(0) = 0$ and $Q_m = 0$ so that the only surface boundary contribution arises from friction.

36.10 Vorticity for the depth integrated horizontal flow

An understanding of vorticity mechanics offers many insights into the nature of fluid flow and how that flow is constrained. Hence, vorticity is a central feature in nearly all theories of fluid mechanics, with particular importance for geophysical fluids. In developing many of these theories, it is of interest to move beyond the vorticity of a fluid element as defined by the curl of the velocity field. For example, as a means to summarize much of the vorticity contained within a three-dimensional fluid, we can study the depth integral of the vorticity equation in Section 36.9 for a hydrostatic and Boussinesq fluid and in Section 41.4 for a planetary geostrophic flow. This analysis has its most common application to the ocean where vertical columns extend from the bottom at $z = \eta_b$ to the surface at $z = \eta$.

Relatedly, there are occasions to study vorticity of the depth integrated or depth averaged flow. Such studies emphasize the importance of boundary forces and their curls (“torques”) for the vorticity of a fluid column. In this section we develop dynamical equations for vorticity of the depth integrated flow in a hydrostatic fluid. A compelling application of these ideas comes from the study of large-scale ocean circulation. The leading order impacts from bottom pressure torques has emerged from research during recent decades, thus pointing to the fundamental role of bottom topography and flows next to sloping bottom (rather than vertical sidewalls) in affecting the ocean circulation. This recognition contrasts to traditional theories whereby the wind stress curl balances meridional motion through the beta effect. In particular, numerical model studies reveal that wind stress curl is sub-dominant in any region with nontrivial bottom velocities.⁶ In this section we introduce the basics and provide more discussion in Sections 41.5 and 41.6 when studying planetary geostrophic equations.

36.10.1 Comparing the two vorticities

In Section 36.9 we derived the evolution equation for the depth integral of the vertical component to the absolute vorticity,

$$\int_{\eta_b}^{\eta} \zeta_a dz = \int_{\eta_b}^{\eta} (f + \hat{z} \cdot \nabla \wedge \mathbf{u}) dz. \quad (36.121)$$

In this section we study the evolution equation for the relative vorticity in the depth integrated horizontal flow

$$\hat{z} \cdot \nabla \wedge \mathbf{U}^\rho = \hat{z} \cdot \nabla \wedge \int_{\eta_b}^{\eta} \mathbf{u} \rho dz, \quad (36.122)$$

where we introduced the depth integrated horizontal mass flux

$$\mathbf{U}^\rho = \int_{\eta_b}^{\eta} \rho \mathbf{u} dz. \quad (36.123)$$

For a Boussinesq fluid we set the density to a constant, in which case the difference between the two relative vorticities is

$$\hat{z} \cdot \nabla \wedge \left[\int_{\eta_b}^{\eta} \mathbf{u} dz \right] - \int_{\eta_b}^{\eta} \hat{z} \cdot \nabla \wedge \mathbf{u} dz = \hat{z} \cdot [\nabla \eta \wedge \mathbf{u}(\eta) - \nabla \eta_b \wedge \mathbf{u}(\eta_b)]. \quad (36.124)$$

⁶The natural ocean has no distinction between side and bottom. Rather, the ocean has a sloping bottom that reaches to the surface along its boundary at the “beach.” [Hallberg and Rhines \(1996\)](#), [Hughes and de Cueves \(2001\)](#), and many subsequent studies emphasize that theoretical and numerical models using vertical sides and a flat bottom exhibit distinct dynamical balances from those models with sloping bottoms.

Flows along boundaries generally have a nontrivial projection in the direction parallel to boundary isosurfaces, in which case the cross products are nonzero thus leading to differences in the two relative vorticities.

36.10.2 Evolution of the depth integrated momentum

The bulk of the derivation in this section consists of deriving the momentum equation for the depth integrated flow. For that purpose we start from the flux-form Eulerian momentum equation derived in Section (19.5). Specializing the expression (19.58) to a hydrostatic fluid leads to the horizontal momentum equation

$$\frac{\partial(\rho \mathbf{u})}{\partial t} + \nabla \cdot [\rho \mathbf{v} \otimes \mathbf{u}] + f \hat{\mathbf{z}} \wedge (\rho \mathbf{u}) = -\nabla_z p + \rho \mathbf{F} \quad (36.125)$$

where $\rho \mathbf{F}$ is the horizontal friction vector as in Section 36.8, and we assumed the simple form for the geopotential, $\Phi = g z$. Finally, we have the outer product defining the components to the kinetic stress (Section 19.5)

$$\rho [\mathbf{v} \otimes \mathbf{u}]_{mi} = \rho v_m u_i = -\mathbb{T}_{mi}^{\text{kinetic}}, \quad (36.126)$$

with $m = 1, 2, 3$ extending over the full range and $i = 1, 2$ extending just over the horizontal range.

Leibniz's rule for the inertial and Coriolis accelerations

Leibniz's rule (Section 16.3.4) renders the following expressions for the depth integrated inertial acceleration and Coriolis acceleration

$$\int_{\eta_b}^{\eta} \frac{\partial(\rho \mathbf{u})}{\partial t} dz = \partial_t \mathbf{U}^\rho - [\rho \mathbf{u} \partial_t \eta]_{z=\eta} \quad (36.127a)$$

$$\int_{\eta_b}^{\eta} \frac{\partial(w \rho \mathbf{u})}{\partial z} dz = [w \rho \mathbf{u}]_{z=\eta} - [w \rho \mathbf{u}]_{z=\eta_b} \quad (36.127b)$$

$$\int_{\eta_b}^{\eta} \nabla_z \cdot [\rho \mathbf{u} \otimes \mathbf{u}] dz = \nabla_z \cdot \left[\int_{\eta_b}^{\eta} \rho \mathbf{u} \otimes \mathbf{u} dz \right] - [\mathbf{u} \cdot \nabla \eta (\rho \mathbf{u})]_{z=\eta} + [\mathbf{u} \cdot \nabla \eta_b (\rho \mathbf{u})]_{z=\eta_b} \quad (36.127c)$$

$$\int_{\eta_b}^{\eta} f \hat{\mathbf{z}} \wedge (\rho \mathbf{u}) dz = f \hat{\mathbf{z}} \wedge \mathbf{U}^\rho. \quad (36.127d)$$

Use of the surface and bottom kinematic boundary conditions from Section 15.4

$$\partial_t \eta + \mathbf{u} \cdot \nabla \eta = w + \rho^{-1} Q_m \quad \text{for } z = \eta \quad (36.128a)$$

$$\mathbf{u} \cdot \nabla \eta_b = w \quad \text{for } z = \eta_b \quad (36.128b)$$

lead to the depth integrated inertial and Coriolis accelerations

$$\begin{aligned} \int_{\eta_b}^{\eta} [\partial_t(\rho \mathbf{u}) + \nabla \cdot (\mathbf{v} \otimes (\rho \mathbf{u})) + f \hat{\mathbf{z}} \wedge (\rho \mathbf{u})] dz \\ = (\partial_t + f \hat{\mathbf{z}} \wedge) \mathbf{U}^\rho - \mathbf{u}(\eta) Q_m - \nabla_z \cdot \int_{\eta_b}^{\eta} \mathbb{T}_{\text{hor}}^{\text{kinetic}} dz, \end{aligned} \quad (36.129)$$

where we introduced the horizontal kinetic stress

$$\mathbb{T}_{\text{hor}}^{\text{kinetic}} = -\rho \mathbf{u} \otimes \mathbf{u}. \quad (36.130)$$

Decomposing the kinetic stress into external and internal velocities

Following our treatment for a stacked shallow water model in Section 32.4.9, it is sometimes useful to decompose the kinetic stress by introducing the density weighted depth averaging operator

$$\bar{\Phi} = \frac{\int_{\eta_b}^{\eta} \rho \Phi dz}{\int_{\eta_b}^{\eta} \rho dz} = \frac{\int_{\eta_b}^{\eta} \rho \Phi dz}{(p_b - p_a)/g} \quad (36.131)$$

so that the density weighted depth averaged horizontal velocity is

$$\bar{\mathbf{u}} = \frac{\int_{\eta_b}^{\eta} \rho \mathbf{u} dz}{\int_{\eta_b}^{\eta} \rho dz} = \frac{\mathbf{U}^\rho}{(p_b - p_a)/g}. \quad (36.132)$$

The velocity

$$\mathbf{u}' = \mathbf{u} - \bar{\mathbf{u}} \quad (36.133)$$

is the deviation from the depth average and we refer to it as the *internal velocity*, whereas the depth averaged velocity, $\bar{\mathbf{u}}$, is the *external velocity*.⁷ Making use of $\mathbf{u} = \bar{\mathbf{u}} + \mathbf{u}'$ yields the depth integrated kinetic stress

$$\int_{\eta_b}^{\eta} \mathbb{T}_{\text{hor}}^{\text{kinetic}} dz = -g^{-1} (p_b - p_a) [\bar{\mathbf{u}} \otimes \bar{\mathbf{u}} + \overline{\mathbf{u}' \otimes \mathbf{u}' }]. \quad (36.134)$$

By construction, there are no cross-terms (i.e., no internal-external terms) appearing in the depth integrated stress (36.134). In this manner we have separated the contributions from kinetic stresses due to depth averaged horizontal velocities from those arising from depth-dependent horizontal velocities.

Depth integrated horizontal pressure gradient

It is of use to decompose the depth integrated horizontal pressure gradient by writing

$$\int_{\eta_b}^{\eta} p dz = \int_{\eta_b}^{\eta} [d(pz) - z dp] = p_a \eta - p_b \eta_b + \mathcal{P}, \quad (36.135)$$

where we used the hydrostatic balance for a vertical fluid column to write⁸ $dp = -g \rho dz$, and introduced the potential energy per horizontal area of a fluid column

$$\mathcal{P} = \int_{\eta_b}^{\eta} g \rho z dz. \quad (36.136)$$

These results then lead to the depth integrated horizontal pressure gradient

$$\int_{\eta_b}^{\eta} \nabla_z p dz = \nabla_z \left[\int_{\eta_b}^{\eta} p dz \right] - p_a \nabla_z \eta + p_b \nabla_z \eta_b \quad (36.137a)$$

$$= \nabla_z [p_a \eta - p_b \eta_b + \mathcal{P}] - p_a \nabla_z \eta + p_b \nabla_z \eta_b \quad (36.137b)$$

$$= \eta \nabla_z p_a - \eta_b \nabla_z p_b + \nabla_z \mathcal{P}. \quad (36.137c)$$

⁷It is also common in the oceanography literature to refer to \mathbf{u}' as the *baroclinic velocity* and $\bar{\mathbf{u}}$ as the *barotropic velocity*.

⁸Specializing to a vertical column allows for the simple transformation, $dp = -g \rho dz$, between pressure and geopotential in an approximate hydrostatic fluid. When also allowing for horizontal deviations then $dp = \nabla p \cdot d\mathbf{x}$.

The depth integrated momentum equation

Bringing the pieces together leads to the depth integrated horizontal momentum equation for a hydrostatic fluid

$$(\partial_t + f \hat{\mathbf{z}} \wedge) \mathbf{U}^\rho = \mathbf{u}(\eta) Q_m - \eta \nabla_z p_a + \eta_b \nabla_z p_b - \nabla_z \mathcal{P} + \nabla_z \cdot \left[\int_{\eta_b}^{\eta} \mathbb{T}_{\text{hor}}^{\text{kinetic}} dz \right] + \int_{\eta_b}^{\eta} \rho \mathbf{F} dz. \quad (36.138)$$

For many applications we focus on the vertical divergence of horizontal frictional stress, which has a vertical integral given by

$$\int_{\eta_b}^{\eta} \rho \mathbf{F} dz = \int_{\eta_b}^{\eta} \partial_z \boldsymbol{\tau} dz = \boldsymbol{\tau}^\eta - \boldsymbol{\tau}^{\eta_b} \equiv \Delta \boldsymbol{\tau}, \quad (36.139)$$

where $\boldsymbol{\tau}^\eta$ is the horizontal stress at the surface, and $\boldsymbol{\tau}^{\eta_b}$ is the horizontal stress at the bottom.

36.10.3 Evolution of vorticity for the depth integrated horizontal flow

Operating with $\hat{\mathbf{z}} \cdot \nabla \wedge$ onto the momentum equation (36.138) leads to

$$\hat{\mathbf{z}} \cdot \partial_t (\nabla \wedge \mathbf{U}^\rho) = -\nabla \cdot (f \mathbf{U}^\rho) + \hat{\mathbf{z}} \cdot \nabla \wedge \left[\mathbf{u}(\eta) Q_m - \eta \nabla_z p_a + \eta_b \nabla_z p_b + \Delta \boldsymbol{\tau} + \nabla_z \cdot \left(\int_{\eta_b}^{\eta} \mathbb{T}_{\text{hor}}^{\text{kinetic}} dz \right) \right], \quad (36.140)$$

where we wrote friction in the form (36.139), and used the vector identity

$$\hat{\mathbf{z}} \cdot \nabla \wedge (f \hat{\mathbf{z}} \wedge \mathbf{U}^\rho) = \nabla \cdot (f \mathbf{U}^\rho), \quad (36.141)$$

as well as $\nabla \wedge \nabla_z \mathcal{P} = 0$.

Beta effect

The first term on the right hand side of the vorticity equation (36.140) arises from the convergence of mass within a fluid column due to depth integrated horizontal flow. Weighting by the Coriolis parameter means that mass convergence at higher latitudes has more impact on vorticity than lower latitudes; i.e., the convergence is more aligned with the planetary rotation at the high latitudes. Performing the product rule renders

$$\nabla \cdot (f \mathbf{U}^\rho) = f \nabla \cdot \mathbf{U}^\rho + \beta V^\rho, \quad (36.142)$$

where the contribution from βV^ρ arises from the beta effect as discussed in Section 36.7.2.

Mass transfer, turbulent momentum transfer, and nonlinear effects

The term $\nabla \wedge [\mathbf{u}(\eta) Q_m]$ accounts for vorticity crossing the ocean surface as affected by the mass flux. The term $\nabla \wedge \Delta \boldsymbol{\tau}$ in equation (36.143) is the torque from turbulent stresses at the ocean surface and bottom, and the final term arises from the nonlinear kinetic stresses, $\mathbb{T}_{\text{hor}}^{\text{kinetic}}$, that account for curls in the self-advection operator.

36.10.4 Boundary pressure torques

The pressure terms in equation⁹ (36.140)

$$\hat{z} \cdot \nabla \wedge (-\eta \nabla p_a + \eta_b \nabla p_b) = \hat{z} \cdot \nabla \wedge (p_a \nabla \eta - p_b \nabla \eta_b) \quad (36.143)$$

arise from curls of the pressure form stresses (see Chapter 20) at the ocean surface and bottom, and these contributions are referred to as *pressure torques*. When the fluid is a column of ocean water, then the surface pressure contribution is the *atmospheric pressure torque* and the bottom pressure term is the *bottom pressure torque*.

Geometry of boundary pressure torques

Geometrically, there is a nonzero atmospheric pressure torque when the applied pressure changes when moving along contours of constant free surface. Likewise, there is a nonzero bottom pressure torque when bottom pressure changes along contours of constant bottom topography. Mathematically, we reveal these properties through use of either Exercise 5.1 or 36.10. For example, the bottom pressure torque along an isobath (contour of constant η_b) can be written

$$\hat{z} \cdot \nabla \eta_b \wedge \nabla p_b = -(\hat{n} \cdot \nabla \eta_b)(\hat{t} \cdot \nabla p_b), \quad (36.144)$$

where \hat{t} is a unit tangent vector directed along the isobath, and \hat{n} points to the left of \hat{t} (see Figure 36.13). Hence, $\hat{n} \cdot \nabla \eta_b$ measures the slope of the bottom topography in the direction normal to an isobath, and $\hat{t} \cdot \nabla p_b$ measures the change of the bottom pressure along the isobath. There is a nonzero bottom pressure torque along an isobath so long as there is a slope to the bottom along the isobath, and there is a change in bottom pressure moving along the isobath.

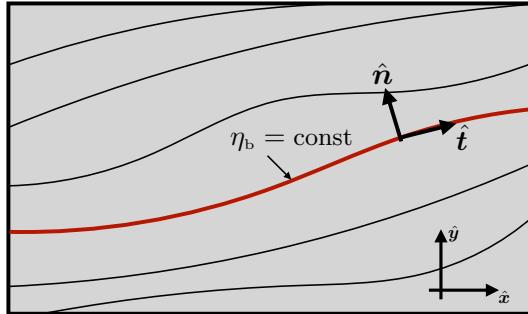


FIGURE 36.13: Geometry depicting a contour along a particular line of constant topography (i.e., an isobath), $\eta_b(x, y)$. The along-contour direction is $\hat{t} = dx/ds$, with s the arc length along the contour. The unit direction pointing to the left of \hat{t} is written \hat{n} , with $\hat{n} \cdot \hat{t} = 0$ and $\hat{t} \wedge \hat{n} = \hat{z}$. Both \hat{n} and \hat{t} are horizontal unit vectors. There is a nonzero bottom pressure torque if bottom pressure changes when following an isobath.

Geostrophic velocity associated with the bottom pressure torque

To further our understanding of the pressure torques in equation (36.143), focus on the bottom pressure and introduce a geostrophic velocity

$$\rho_0 f \mathbf{u}_g = \hat{z} \wedge \nabla p_b, \quad (36.145)$$

⁹Recall that to reduce notational clutter, we can write ∇ rather than ∇_z when operating on functions that are independent of z , such as η , p_a , η_b and p_b .

so that the bottom pressure torque takes the form

$$\hat{z} \cdot \nabla \wedge (\eta_b \nabla p_b) = \hat{z} \cdot \nabla \eta_b \wedge \nabla p_b = \rho_0 f \mathbf{u}_g \cdot \nabla \eta_b. \quad (36.146)$$

This equation is merely a replacement of the bottom pressure gradient with a corresponding geostrophic velocity. However, if this geostrophic velocity is assumed to satisfy the kinematic boundary condition (15.41), then we reach the equality¹⁰

$$\hat{z} \cdot \nabla \eta_b \wedge \nabla p_b = \rho_0 f w_g, \quad (36.147)$$

which links the bottom pressure torque to vertical vortex stretching by the vertical component to the geostrophic velocity.

The equality (36.147) is sometimes used to infer the bottom pressure torque by diagnosing the bottom vertical velocity, $w(\eta_b)$. Although this diagnostic is suitable for some studies, there are important caveats. Namely, the bottom vertical velocity is generally affected by bottom frictional effects and thus can have a nontrivial Ekman component.¹¹ Nonlinear effects can also be important especially when considering motions with sizable Rossby numbers. Neither the Ekman component nor nonlinear terms are directly related to the bottom pressure torque. We thus expect $\rho_0 f w(\eta_b)$ to be distinct from $\hat{z} \cdot \nabla \eta_b \wedge \nabla p_b$ in regions of sizable deep flows where bottom friction and/or nonlinear effects are of leading order importance.¹² The studies from [Gula et al. \(2015\)](#) and [LeCorre et al. \(2020\)](#) illustrate these points from numerical simulations of the subpolar North Atlantic circulation.

36.10.5 Steady state vorticity budget

The steady state form of the vorticity budget (36.140) leads to the balance

$$\beta V^\rho = -f \nabla \cdot \mathbf{U}^\rho + \hat{z} \cdot \nabla \wedge \left[\mathbf{u}(\eta) Q_m - \eta \nabla_z p_a + \eta_b \nabla_z p_b + \Delta \boldsymbol{\tau} + \nabla_z \cdot \left(\int_{\eta_b}^{\eta} \mathbb{T}_{\text{hor}}^{\text{kinetic}} dz \right) \right]. \quad (36.148)$$

Writing the balance in this manner reveals how the beta effect affords a steady meridional mass transport as a balance with the variety of terms on the right hand side. Let us further specialize to the case with a uniform mass atmosphere so that p_a is a constant. Furthermore, assume zero boundary mass transport so that $Q_m = 0$ and, correspondingly, the steady depth integrated mass budget (15.88) means that $\nabla \cdot \mathbf{U}^\rho = 0$ when $Q_m = 0$. These simplifications bring the balance (36.148) to the form

$$\beta V^\rho = \hat{z} \cdot \nabla \wedge \left[\eta_b \nabla_z p_b + \Delta \boldsymbol{\tau} + \nabla_z \cdot \left(\int_{\eta_b}^{\eta} \mathbb{T}_{\text{hor}}^{\text{kinetic}} dz \right) \right] \quad (36.149a)$$

$$\text{MERIDIONAL TRANSPORT} = \text{BOTTOM PRESS TORQUE} + \text{BOUNDARY TURBULENT STRESS} + \text{NONLINEAR}. \quad (36.149b)$$

This steady balance reveals distinct flow regimes depending on which of the terms dominate, and as such it serves as a useful framework for analysis.

¹⁰The velocity, \mathbf{v} , satisfies the kinematic boundary condition (15.41), in which $\mathbf{v}(\eta_b) \cdot \hat{\mathbf{n}} = 0$. Decomposing the velocity into its geostrophic and ageostrophic components, $\mathbf{v} = \mathbf{v}_g + \mathbf{v}_a$, does not generally imply that \mathbf{v}_g and \mathbf{v}_a separately satisfies the kinematic boundary condition. Rather, we must make that assumption in order to reach the equality (36.147).

¹¹Recall our discussion of Ekman boundary layers in Chapter 28.

¹²In addition to the Ekman and nonlinear effects noted here, diagnosing $w(\eta_b)$ in a numerical model can be fraught with difficulties related to the discrete grid stencil given that grids can be quite coarse in the deep ocean with many ocean model configurations.

36.10.6 Budget in terms with the vector invariant velocity formulation

In formulating the budget for vorticity of the depth integrated flow, we started by formulating the budget for the depth integrated momentum in Section 36.10.2. However, many numerical models are formulated using the vector invariant form of the horizontal velocity equation (36.95), here written in the equivalent form for a Boussinesq fluid

$$\underbrace{\frac{\partial \mathbf{u}}{\partial t} + f \hat{\mathbf{z}} \wedge \mathbf{u} + \nabla_z p / \rho_0 - \mathbf{F}}_{\text{linear terms plus friction}} = \underbrace{-\zeta \hat{\mathbf{z}} \wedge \mathbf{u} - w \partial_z \mathbf{u} - \nabla_z \mathbf{u}^2 / 2}_{\text{Magnus + vertical advection + kinetic energy}}. \quad (36.150)$$

The three nonlinear terms on the right hand side arise from expanding the nonlinear self-advection term, $(\mathbf{v} \cdot \nabla) \mathbf{u}$, following the manipulations in Section 36.8.1. We now take the vertical integral of this equation

$$\int_{\eta_b}^{\eta} [\partial_t \mathbf{u} + f \hat{\mathbf{z}} \wedge \mathbf{u} + \nabla_z p / \rho_0 - \mathbf{F}] dz = - \int_{\eta_b}^{\eta} [\zeta \hat{\mathbf{z}} \wedge \mathbf{u} - w \partial_z \mathbf{u} - \nabla_z \mathbf{u}^2 / 2] dz. \quad (36.151)$$

The steady state vorticity balance thus takes the form

$$\beta V = -f \nabla \cdot \mathbf{U} + \hat{\mathbf{z}} \cdot \nabla \wedge \int_{\eta_b}^{\eta} (-\nabla_z p / \rho_0 + \mathbf{F} - \zeta \hat{\mathbf{z}} \wedge \mathbf{u} - w \partial_z \mathbf{u} - \nabla_z \mathbf{u}^2 / 2) dz. \quad (36.152)$$

Note the absence of any Leibniz rule manipulations on the nonlinear terms, which contrasts to the approach pursued in Section 36.10.2. However, for the pressure gradient we might choose to write it in the form

$$\int_{\eta_b}^{\eta} \nabla_z p dz = \nabla_z \int_{\eta_b}^{\eta} p dz - p_a \nabla \eta + p_b \nabla \eta_b, \quad (36.153)$$

thus revealing the boundary pressure torques as in Section 36.10.3.

36.10.7 Vorticity of the depth averaged flow

The vorticity of the depth averaged flow is given by $\hat{\mathbf{z}} \cdot \nabla \wedge \bar{\mathbf{u}}$, where $\bar{\mathbf{u}}$ is given by equation (36.132) for a compressible non-Boussinesq fluid, and which takes on the following form for a Boussinesq fluid

$$\bar{\mathbf{u}} = \frac{\int_{\eta_b}^{\eta} \mathbf{u} dz}{\eta - \eta_b} = \frac{\mathbf{U}}{\eta - \eta_b}. \quad (36.154)$$

The difference is given by

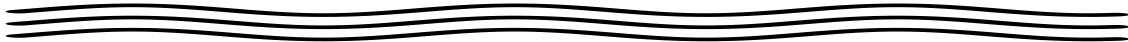
$$\nabla \wedge \mathbf{U} - (\eta - \eta_b) \nabla \wedge \bar{\mathbf{u}} = \nabla(\eta - \eta_b) \wedge \bar{\mathbf{u}}, \quad (36.155)$$

so that the two vorticities are the same in the special case of a depth averaged flow that is parallel to $\nabla(\eta - \eta_b)$. Quite trivially, $\nabla(\eta - \eta_b) \wedge \bar{\mathbf{u}} = 0$ occurs for a rigid lid and flat bottom ocean, in which $\nabla \eta = \nabla \eta_b = 0$. More generally, $\nabla(\eta - \eta_b) \wedge \bar{\mathbf{u}} \neq 0$, particularly in the presence of topography. We further study the budgets for these two vorticities, for planetary geostrophic flow, in Sections 41.5 and 41.6.

36.10.8 Comments and further study

The diagnostic budgets derived in this section have appeared in many studies of ocean vorticity. When diagnosing the budget terms in a numerical model, the choice for how to mathematically formulate the diagnostic balances is largely driven by physical transparency as well as by numerical precision. Concerning numerical precision, it is useful to note that vorticity, as the derivative of velocity, has far more power at the high spatial wave numbers. In a numerical model, such power can manifest as noise. It is thus of use to perform much of the calculation online to enable the most precise diagnostic. Further spatial smoothing is generally required, especially in realistic models, to extract physically clear signals.

In Sections 41.4, 41.5, and 41.6, we consider further aspects of the vorticity for the depth integrated and depth averaged flow with particularly emphasis on a planetary geostrophic fluid. In that discussion, we drop the contributions from $\nabla \wedge [\mathbf{u}(\eta) Q_m]$, drop the nonlinear contributions from the kinetic stress, and drop the time tendency. In this case, the vorticity balance (36.149b) is between boundary frictional stresses, boundary pressure torques, and the beta effect.



36.11 Exercises

EXERCISE 36.1: BAROCLINICITY WITH $\rho_0(z)$

Recall the discussion of the Boussinesq momentum equation in Section 24.1.2. The form given by equation (24.8) is written with the reference density, $\rho_0 = \rho_0(z)$. We then stated that the form of the baroclinicity vector appearing in the Boussinesq vorticity equation is greatly simplified by setting ρ_0 to a global constant, and thus dropping the z dependence. Derive the second term in the baroclinicity

$$\mathbf{B} = \nabla \left[b - \frac{\delta p}{\rho_0^2} \frac{d\rho_0}{dz} \right] \wedge \hat{\mathbf{z}}, \quad (36.156)$$

so that $\mathbf{B} = \nabla b \wedge \hat{\mathbf{z}}$ when ρ_0 is assumed to be a global constant. Hint: write the vector-invariant form of equation (24.8) with $\rho_0(z)$. Then take the curl.

EXERCISE 36.2: FRICTION IN THE VORTICITY EQUATION

Add a viscous term of the form

$$\mathbf{F} = \nu \nabla^2 \mathbf{v}, \quad (36.157)$$

with ν a constant molecular kinematic viscosity. How is the vorticity equation modified?

EXERCISE 36.3: FRICTION FOR NON-DIVERGENT FLOWS

Consider a non-divergent Boussinesq fluid with a Laplacian frictional acceleration

$$\mathbf{F} = \nu \nabla^2 \mathbf{v} \quad \text{with} \quad \nabla \cdot \mathbf{v} = 0, \quad (36.158)$$

with ν a constant molecular kinematic viscosity. Write this expression in terms of the vorticity. Comment.

EXERCISE 36.4: GENERATION OF VORTICITY BY BAROCLINICITY

Consider an initially resting body of water with a flat bottom and rigid sides. Let the top surface be at $z = 0$ and bottom at $z = -H$, and assume zero pressure applied at the top surface. Let the density have a horizontal structure given by

$$\rho(x) = \rho_0 (1 - \gamma x) \quad (36.159)$$

where ρ_0 and γ are positive constants (with dimensions of density and inverse length, respectively). We furthermore assume that $\gamma|x| \ll 1$ so that the density is strictly positive. Note that a study of Figure 36.5 will help with this exercise.

- (a) Compute the density gradient $\nabla\rho$ and draw a schematic.
- (b) Compute the pressure gradient, ∇p , assuming approximate hydrostatic balance so that $\partial p/\partial z = -\rho g$. Draw a schematic at $x = 0$.
- (c) Compute the baroclinicity/solenoidal vector $\mathbf{B} = \rho^{-2} (\nabla\rho \wedge \nabla p)$. Draw a schematic.
- (d) Describe the vorticity induced by the baroclinicity vector.

EXERCISE 36.5: GENERATION OF CIRCULATION BY BAROCLINICITY IN AN IDEAL GAS

In this exercise we examine the baroclinicity vector for a simple ideal gas, which is described by the equation of state (21.115)

$$\rho = \frac{p M_{\text{mole}}}{T R_g} \equiv \frac{p}{T R^M}, \quad (36.160)$$

where R_g is the universal gas constant and R^M is the specific gas constant. We also assume the atmosphere is in approximate hydrostatic balance (Section 23.3), and we ignore rotation (relatively small lateral region of the atmosphere). For further hints to this exercise, see Section 4.1 of Holton (1992) or Section 2.4.3 of Markowski and Richardson (2010), where they discuss circulation generated by differences in land-sea temperatures, thus leading to a sea breeze.

- (a) Express the baroclinicity vector, \mathbf{B} , in terms of pressure and temperature gradients.
- (b) Express the baroclinicity vector in terms of pressure and potential temperature gradients.
Hint: see Section 21.9.11 for potential temperature in an ideal gas.
- (c) Consider an ideal gas atmosphere straddling the ocean and flat land as in Figure 36.14. Let the daytime air be relatively cool over the ocean and relatively warm over the land. Furthermore, assume the sea level pressure is the same value over land and ocean. Ignoring rotation, draw isolines of constant temperature and constant pressure. Assume the horizontal temperature gradient is constant with height. Here are some hints.
 - Temperature decreases from land to ocean and decreases when ascending into the atmosphere.
 - Pressure is assumed to be horizontally constant at sea level and it decreases upward. Use the ideal gas law to determine the sense for the horizontal pressure gradient as one ascends. Consult the discussion in Section 21.9.10 for geopotentials in an ideal gas atmosphere.
 - We are only concerned with a qualitative sense for the isolines in the lower atmosphere and over a horizontal region small enough that rotation can be ignored.
- (d) Describe the sense for the circulation induced by the baroclinicity. Does circulation correspond to your experience at a sunny beach day as the air warms over the land faster than over the adjacent ocean? What force causes air to rise and to fall?

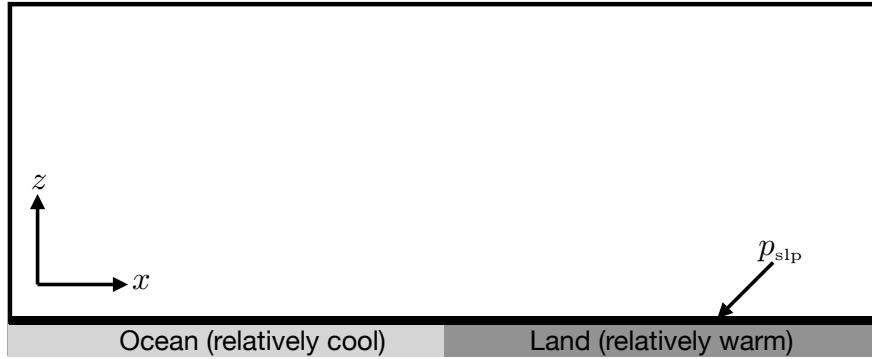


FIGURE 36.14: Setup for the sea breeze Exercise 36.5. We here depict a vertical-zonal cross-section of the atmosphere where the lower boundary straddles the ocean and land. The atmosphere over the ocean is assumed to be cooler than the atmosphere over the land, as typically occurs on a sunny afternoon with solar radiation warming land faster than the ocean.

EXERCISE 36.6: CIRCULATION WITH ISLANDS

Our discussion of Stokes' theorem has been thus far restricted to a simply connected domain, in which

$$\oint_{\partial S} \mathbf{v} \cdot d\mathbf{r} = \int_S \boldsymbol{\omega} \cdot \hat{\mathbf{n}} dS. \quad (36.161)$$

For a simply connected domain, the closed contour can be shrunk to a point without leaving the domain.

A more general topology consists of a region with holes, whereby closed contours cannot in general be shrunk to a point without leaving the region. In an oceanographic context, the “holes” are islands or continents and the circulation is that for the depth integrated flow. Figure 36.15 shows a region of the ocean containing three arbitrarily shaped impenetrable islands, with the three islands surrounded by a contour. The contour cannot be shrunk to a point without crossing over the islands, thus making this region of the ocean multiply-connected. The presence of islands thus adds a level of complexity to the World Ocean that is absent in an AquaPlanet or the global atmosphere.

Derive the following expression for the circulation in multiply-connected regions

$$\oint_{\partial S} \mathbf{v} \cdot d\mathbf{r} = \sum_{n=1}^N \left(\oint_{\partial S_n} \mathbf{v} \cdot d\mathbf{r} \right) + \int_S \boldsymbol{\omega} \cdot \hat{\mathbf{n}} dS, \quad (36.162)$$

where N is the number of islands, S_n is the contour surrounding each island, and S is the region of water that excludes the islands. In words, this result says that the circulation around a region equals to the circulation around the islands within the region, plus the normal component of the vorticity integrated over the area within the fluid region. Removing the islands allows the island contours to be shrunk to zero size, in which case we recover the simply connected result (36.161).

EXERCISE 36.7: EVOLUTION OF CIRCULATION AROUND ISLANDS

The momentum equation for a homogeneous layer of inviscid shallow water on a tangent plane is given by

$$\frac{\partial \mathbf{u}}{\partial t} + (\mathbf{u} \cdot \nabla) \mathbf{u} + f \hat{\mathbf{z}} \wedge \mathbf{u} = -g \nabla \eta. \quad (36.163)$$

In this equation, $\mathbf{u} = (u, v)$ is the horizontal velocity, f is the Coriolis parameter, g is the effective gravitational acceleration, and η is the deviation of the free surface from its horizontal resting

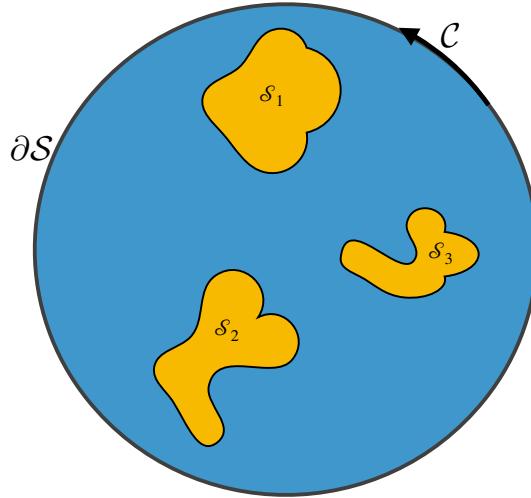


FIGURE 36.15: A region of the ocean consisting of three islands, \mathcal{S}_1 , \mathcal{S}_2 , and \mathcal{S}_3 , each with boundaries $\partial\mathcal{S}_n$ and with the closed contour, $\partial\mathcal{S}$, drawn around the three islands. The contour $\partial\mathcal{S}$ cannot be shrunk to a point without crossing over the islands, thus indicating that the domain is multiply connected. Exercise 36.6 is concerned with deriving an expression for the circulation of the depth-integrated flow as defined along the closed contour, $\partial\mathcal{S}$.

position. All spatial derivatives are horizontal, so that

$$\mathbf{u} \cdot \nabla = u \partial_x + v \partial_y. \quad (36.164)$$

Use of a vector identity allows us to write

$$\frac{\partial \mathbf{u}}{\partial t} + (f + \zeta) \hat{\mathbf{z}} \wedge \mathbf{u} = -\nabla (\mathbf{u}^2/2 + g \eta), \quad (36.165)$$

where

$$\zeta = \hat{\mathbf{z}} \cdot (\nabla \wedge \mathbf{u}) \quad (36.166)$$

is the vorticity of the shallow water fluid.

Consider an island, such as one shown in Figure 36.15. Each island is static and impenetrable to fluid flow, which means that

$$\mathbf{u} \cdot \hat{\mathbf{n}} = 0 \quad (36.167)$$

where $\hat{\mathbf{n}}$ is the outward normal on an island boundary. This no-normal flow constraint means that the velocity just next to an island is parallel to the island¹³

$$\mathbf{u} \wedge d\mathbf{r} = 0. \quad (36.168)$$

Equivalently, the island represents a solid material boundary across which no flow passes. Show that the inviscid shallow-water circulation around an island remains constant in time

$$\frac{d}{dt} \oint_I \mathbf{u} \cdot d\mathbf{r} = 0. \quad (36.169)$$

Recall that Kelvin's circulation theorem is formulated for a material contour in an inviscid fluid. This exercise shows that the circulation theorem also holds for a material contour enclosing a static solid boundary.

¹³This boundary condition is valid only for inviscid fluids such as that considered here. For a real fluid with nonzero viscosity, all components of the velocity vector vanish at solid boundaries due to the no-slip condition.

EXERCISE 36.8: HELICITY FOR A PERFECT BAROTROPIC FLUID

Consider a closed material volume, \mathcal{R} , of a perfect non-rotating homogenous barotropic fluid. Let this material volume have a boundary that is always tangent to the fluid vorticity, ω . Hence, the outward normal to the region boundary is orthogonal to the vorticity,

$$\hat{\mathbf{n}} \cdot \boldsymbol{\omega} = 0. \quad (36.170)$$

Such volumes define closed vortex tubes, such as a smoke ring or linked smoke rings. The *helicity* of the fluid within the vortex tube volume is defined as the integration of the helicity density, $\mathbf{v} \cdot \boldsymbol{\omega}$, over the closed volume

$$\mathbb{H} = \int_{\mathcal{R}(\mathbf{v})} \mathbf{v} \cdot \boldsymbol{\omega} \, dV, \quad (36.171)$$

where the volume $\mathcal{R}(\mathbf{v})$ is material. In Cartesian coordinates, the helicity density takes the form

$$\mathbf{v} \cdot \boldsymbol{\omega} = u (\partial_y w - \partial_z v) + v (\partial_z u - \partial_x w) + w (\partial_x v - \partial_y u). \quad (36.172)$$

Although the helicity density vanishes for some common examples, such as for a fluid in solid-body rotation, it need not vanish in general.

- (a) Show that helicity is materially constant following the material volume

$$\frac{d\mathbb{H}}{dt} = 0. \quad (36.173)$$

- (b) Discuss why helicity is not defined for a shallow water fluid.

Use the following hints.

- Make use of Φ_p that satisfies equation (36.17).
- The shallow water fluid model is based on the small aspect ratio limit, in which the fluid depth is much smaller than its lateral extent. In this limit, only the vertical component of vorticity is nontrivial.

EXERCISE 36.9: DISCRETE CALCULATION OF BOTTOM PRESSURE TORQUE

In many diagnostic studies with numerical models it is of interest to compute pressure torques affecting vorticity. One particularly common diagnostic concerns the bottom pressure torque arising in equation (36.143). Derive a discrete expression for the area averaged bottom pressure torque

$$\overline{\text{BPT}} = A^{-1} \int_{\mathcal{S}} \hat{\mathbf{z}} \cdot \nabla \wedge (\eta_b \nabla p_b) \, d\mathcal{S} = A^{-1} \oint_{\partial\mathcal{S}} \eta_b \nabla p_b \cdot \hat{\mathbf{t}} \, d\ell = -A^{-1} \oint_{\partial\mathcal{S}} p_b \nabla \eta_b \cdot \hat{\mathbf{t}} \, d\ell, \quad (36.174)$$

over the shaded region depicted in Figure 36.16, with $A = \int_{\mathcal{S}} d\mathcal{S}$ the horizontal area of this region. Hint: this exercise shares much with the area averaged vorticity in Exercise 33.6, although the final result is distinct. Note: given that the bottom pressure torque is generally the small difference between large numbers, it is very useful to perform the diagnostic calculation online so that full computational precision can be maintained.

EXERCISE 36.10: DYNAMICAL PORTION OF THE TOPOGRAPHIC FORM STRESS

We discussed bottom topographic form stress in Chapter 20, with its curl leading to the bottom pressure torque in equation (36.143). As noted in Section 20.1.3, the dominant portion of the bottom topographic form stress acting on the ocean has little to do with fluid motion. Rather, it merely holds the ocean fluid within the basin, much as water is held within a drinking container through pressure imparted by the container sides.

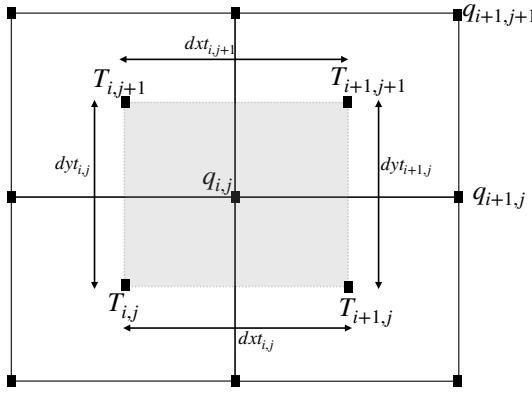


FIGURE 36.16: Discrete grid layout for variables needed to compute the bottom pressure torque as averaged over the shaded vorticity region. The bottom topography and bottom pressure are both known at the tracer points.

- (a) To help isolate the dynamically relevant portion of the bottom bottom pressure, show that we can write the horizontal gradient of the bottom pressure for a hydrostatic fluid according to

$$\nabla_z p_b = g [\rho(\eta) \nabla_z \eta - \rho(\eta_b) \nabla_z \eta_b] + g \int_{\eta_b}^{\eta} \nabla_z \rho dz. \quad (36.175)$$

where we ignore the applied surface pressure, p_a , for simplicity.

- (b) Hence, show that the bottom pressure torque takes the form

$$\hat{z} \cdot \nabla_z \eta_b \wedge \nabla_z p_b = g \hat{z} \cdot \nabla_z \eta_b \wedge \left[\rho(\eta) \nabla_z \eta + \int_{\eta_b}^{\eta} \nabla_z \rho dz \right] \equiv \hat{z} \cdot \nabla \eta_b \wedge \nabla p_b^*, \quad (36.176)$$

where

$$\nabla_z p_b^* = g \rho(\eta) \nabla_z \eta + g \int_{\eta_b}^{\eta} \nabla_z \rho dz. \quad (36.177)$$

Note how $\nabla_z p_b^*$ has no contribution from the potentially very large term, $-g \rho(\eta_b) \nabla_z \eta_b$, arising from gradients in the bottom topography.

- (c) [Molemaker et al. \(2015\)](#) and [Gula et al. \(2015\)](#) assume knowledge of the bottom pressure torque along contours of constant topography. Given that knowledge they then make use of the following diagnostic expression for p_b^*

$$p_b^*(s) - p_b^*(s_0) = - \int_{s_0}^s \frac{\hat{z} \cdot (\nabla \eta_b \wedge \nabla p_b)}{\hat{n} \cdot \nabla \eta_b} ds, \quad (36.178)$$

with $p_b^*(s_0)$ the value at the arbitrary starting point for the contour. Derive equation (36.178), with the following information of possible use.

- As depicted in Figure 36.13, s is the arc length along the chosen contour of constant η_b , with s increasing in the tangent direction, \hat{t} . Likewise, \hat{n} is a unit vector pointing to the left of the contour so that $\hat{n} \cdot \hat{t} = 0$ and $\hat{t} \wedge \hat{n} = \hat{z}$.
- Along any contour of constant $\eta_b(x, y)$ we have

$$0 = d\eta_b = \nabla \eta_b \cdot d\mathbf{x} = \nabla \eta_b \cdot \frac{d\mathbf{x}}{ds} ds = \nabla \eta_b \cdot \hat{t} ds. \quad (36.179)$$

- The main mathematics of this exercise are contained in Exercise 5.1.

Equation (36.178) provides a means to compute the anomalous p_b^* [anomalous relative to $p_b^*(s_0)$] along a constant topography contour. Mapping $p_b^*(s) - p_b^*(s_0)$ for a suite of contours then provides the means to determine the dynamically relevant portion of the bottom pressure and then, when multiplying by the bottom slope, compute the dynamically relevant portion of the form stress.



Angular momentum, vorticity, and strain

As noted in Section 33.5.4, fluid flow in the presence of a free vortex (Section 33.4) has zero vorticity for all points except the origin of the vortex. However, the same points also have a constant angular momentum relative to the origin, and they experience a nonzero strain. In contrast, solid-body fluid flow (Section 33.5) has a nonzero vorticity, nonzero angular momentum, yet a zero strain. In this chapter we detail the formal connection between vorticity, strain, and angular momentum in a fluid flow.

READER'S GUIDE TO THIS CHAPTER

We assume an understanding of vorticity as given in Chapter 36, and make use of Cartesian tensors as presented in Chapter 1. No subsequent chapter makes use of the results here. Rather, this brief chapter serves only to satisfy the curiosity of interested readers.

37.1	Loose threads	871
37.2	A resume of point particle mechanics	871
37.2.1	Angular velocity and moment of inertia	872
37.2.2	Relating angular velocity to velocity	873
37.3	Linear momentum for material fluid regions	873
37.4	Angular momentum for material fluid regions	874
37.4.1	Taylor expanding the velocity	875
37.4.2	Strain and vorticity	875
37.4.3	Relating angular momentum to strain and vorticity	876
37.4.4	Comments	877

37.1 Loose threads

- Section 2.3.1 of [Davidson \(2015\)](#) has nice discussion about vorticity and angular momentum.
- Bring material from Section 37.2 into the particle mechanics Section 10.1.

37.2 A resume of point particle mechanics

Much of this section follows from our earlier treatment of particle mechanics in Part II. We revisit salient points to emphasize elements of interest for the present chapter.

The linear momentum of a point particle is written¹

$$\mathbf{P} = M \mathbf{V}, \quad (37.1)$$

where M is the particle's mass, which is a measure of the particle's inertia. The velocity, \mathbf{V} , is the time change of the particle position,

$$\mathbf{V} = \frac{d\mathbf{X}}{dt}. \quad (37.2)$$

The corresponding angular momentum is given by

$$\mathbf{L} = \mathbf{X} \wedge \mathbf{P} = M (\mathbf{X} \wedge \mathbf{V}). \quad (37.3)$$

The angular momentum is a function of the origin of the chosen coordinate system. The utility and relevance of angular momentum stems from its conservation for systems exhibiting rotational symmetry about special points or special directions. For example, motion on a smooth sphere exhibits rotational symmetry with respect to the center of the sphere. Consequently, all components of angular momentum for a particle are constant in the absence of externally applied torques. Likewise, for motion on a smooth rotating sphere, we showed in Section 11.7 that the component of angular momentum about the rotation axis is a constant of the motion.

37.2.1 Angular velocity and moment of inertia

Whereas linear momentum has physical dimensions of

$$[\mathbf{P}] \equiv \text{mass} \times \text{length} \times \text{time}^{-1}, \quad (37.4)$$

angular momentum has dimensions of

$$[\mathbf{L}] \equiv \text{mass} \times \text{length}^2 \times \text{time}^{-1}. \quad (37.5)$$

We can pursue the analog by introducing the angular velocity

$$\boldsymbol{\Omega} = \frac{\mathbf{X} \wedge \mathbf{V}}{|\mathbf{X}|^2}. \quad (37.6)$$

The angular velocity has physical dimensions of inverse time, and it is defined with respect to the chosen coordinate origin. Furthermore, by construction the angular velocity vector is orthogonal to both the velocity and to the position

$$\boldsymbol{\Omega} \cdot \mathbf{X} = \boldsymbol{\Omega} \cdot \mathbf{V} = 0. \quad (37.7)$$

The angular velocity is not defined at the origin since $|\mathbf{X}| = 0$.

Inserting the definition of the angular velocity (37.6) into the angular momentum (37.3) renders

$$\mathbf{L} = M (\mathbf{X} \wedge \mathbf{V}) \quad (37.8a)$$

$$= M |\mathbf{X}|^2 \boldsymbol{\Omega} \quad (37.8b)$$

$$\equiv I \boldsymbol{\Omega}. \quad (37.8c)$$

In the final equality we introduced the moment of inertia for a point particle

$$I = M |\mathbf{X}|^2. \quad (37.9)$$

The moment of inertia measures the inertia appropriate for determining angular momentum relative to a chosen coordinate origin. The moment of inertia scalar, I , generalizes to the moment of inertia tensor, I_{mn} , when considering angular momentum for extended matter, such as a rigid body or a material fluid region (Section 37.4.3).

¹We use capital letters to accord with our usage of Lagrangian fluid particle trajectories.

37.2.2 Relating angular velocity to velocity

The cross product of the position vector with the angular velocity (37.6) is given by

$$\boldsymbol{\Omega} \wedge \mathbf{X} = \mathbf{V} - \mathbf{X} \frac{\mathbf{V} \cdot \mathbf{X}}{|\mathbf{X}|^2}. \quad (37.10)$$

For the special case of velocity \mathbf{V} orthogonal to the position vector, \mathbf{X} , we have

$$\mathbf{V} = \boldsymbol{\Omega} \wedge \mathbf{X} \quad \text{when } \mathbf{V} \cdot \mathbf{X} = 0. \quad (37.11)$$

In particular, for circular motion the velocity is orthogonal to the position.

37.3 Linear momentum for material fluid regions

We now consider the velocity and linear momentum of a connected material fluid region denoted by $\mathcal{R}(\mathbf{v})$, with each point of the region moving with the local fluid velocity $\mathbf{v}(\mathbf{x}, t)$. Let an arbitrary fluid parcel within this region be marked with the material label \mathbf{a} so that its position vector is $\mathbf{X}(\mathbf{a}, t)$ and its velocity is

$$\mathbf{V}(\mathbf{a}, t) = \frac{\partial \mathbf{X}(\mathbf{a}, t)}{\partial t}, \quad (37.12)$$

where the time derivative is computed holding the material label fixed. Since the parcel is within a finite material region, we find it useful to decompose the motion of the parcel into the motion of the region's center of mass plus the motion of the parcel relative to the center of mass

$$\mathbf{V}(\mathbf{a}, t) = \frac{\partial \mathbf{X}(\mathbf{a}, t)}{\partial t} \quad (37.13a)$$

$$= \frac{\partial (\bar{\mathbf{X}} + \mathbf{X}')}{\partial t} \quad (37.13b)$$

$$= \bar{\mathbf{V}}(t) + \mathbf{V}'(\mathbf{a}, t). \quad (37.13c)$$

In this equation, we introduced the velocity \mathbf{V}' defined relative to the center of mass of the region. Furthermore, the center of mass velocity is given by

$$\bar{\mathbf{V}} = \frac{d\bar{\mathbf{X}}}{dt} \quad (37.14a)$$

$$= \frac{d}{dt} \left[\frac{\int_{\mathcal{R}(\mathbf{v})} \mathbf{x} \rho dV}{\int_{\mathcal{R}(\mathbf{v})} \rho dV} \right] \quad (37.14b)$$

$$= \frac{1}{M} \int_{\mathcal{R}(\mathbf{v})} \frac{D\mathbf{x}}{Dt} \rho dV \quad (37.14c)$$

$$= \frac{1}{M} \int_{\mathcal{R}(\mathbf{v})} \mathbf{v} \rho dV. \quad (37.14d)$$

The identity (37.14c) follows since the material region maintains a constant mass,

$$M = \int_{\mathcal{R}(\mathbf{v})} \rho dV \implies \frac{dM}{dt} = 0, \quad (37.15)$$

allowing the denominator to come outside the derivative. Additionally, each of the fluid parcels in the region maintains constant mass. As per Reynold's transport theorem (Section 15.3.4), the time

derivative moves across the integral to act materially on the position vector. The final equality, (37.14d), follows since the material time derivative of a parcel trajectory when evaluated at a point, \mathbf{x} , equals to the velocity field at that point

$$\mathbf{v}(\mathbf{x}, t) = \frac{D\mathbf{x}}{Dt}. \quad (37.16)$$

It follows that the linear momentum for the material fluid region is given by

$$\mathbf{P} = \int_{\mathcal{R}(\mathbf{v})} \mathbf{v} \rho dV = M \bar{\mathbf{V}}. \quad (37.17)$$

We conclude that the total linear momentum of an extended body equals to that of a point particle of mass $M = \int_{\mathcal{R}(\mathbf{v})} \rho dV$ moving with the center of mass velocity, $\bar{\mathbf{V}}$.

37.4 Angular momentum for material fluid regions

We here consider angular momentum for a material fluid region, which is determined by the integral over that region of the angular momentum for each fluid parcel

$$\mathbf{L} = \int_{\mathcal{R}(\mathbf{v})} (\mathbf{x} \wedge \mathbf{v}) \rho dV. \quad (37.18)$$

Our goal is to expose how physically distinct aspects of the fluid motion contribute to the angular momentum. To proceed, decompose the position vector of a point within the region into the center of mass position plus a deviation, $\mathbf{x} = \bar{\mathbf{x}} + \mathbf{x}'$, where $\bar{\mathbf{x}} = \bar{\mathbf{X}}$ is the instantaneous position of the moving center of mass. The angular momentum thus takes the form

$$\mathbf{L} = \int_{\mathcal{R}(\mathbf{v})} (\mathbf{x} \wedge \mathbf{v}) \rho dV \quad (37.19a)$$

$$= \int_{\mathcal{R}(\mathbf{v})} [(\bar{\mathbf{x}} + \mathbf{x}') \wedge \mathbf{v}] \rho dV \quad (37.19b)$$

$$= \bar{\mathbf{X}} \wedge \left[\int_{\mathcal{R}(\mathbf{v})} \mathbf{v} \rho dV \right] + \int_{\mathcal{R}(\mathbf{v})} (\mathbf{x}' \wedge \mathbf{v}) \rho dV \quad (37.19c)$$

$$= (\bar{\mathbf{X}} \wedge \mathbf{P}) + \int_{\mathcal{R}(\mathbf{v})} (\mathbf{x}' \wedge \mathbf{v}) \rho dV. \quad (37.19d)$$

The final equality introduced the linear momentum, (37.17), for the fluid region. The first term in equation (37.19d) is the angular momentum of the region with respect to the position of the center of mass. The second term is associated with deviations of parcel positions relative to the center of mass.

We now focus on how the deviation term, $\int_{\mathcal{R}(\mathbf{v})} (\mathbf{x}' \wedge \mathbf{v}) \rho dV$, contributes to the angular momentum (37.18). As we will see, this analysis exposes how angular momentum of the extended material fluid region is affected by vorticity and strain in the fluid flow. To facilitate some of the manipulations, we make use of basic Cartesian tensor analysis from Chapter 1, including the summation convention whereby repeated indices are summed over their range. Additionally, we introduce components to the totally anti-symmetric Levi-Civita tensor, ϵ_{mnp} so that the vector cross product is written (see Section 1.4)

$$(\mathbf{A} \wedge \mathbf{B})_m = \epsilon_{mnp} A_n B_p. \quad (37.20)$$

37.4.1 Taylor expanding the velocity

We now perform a Taylor expansion of the velocity $\mathbf{v}(\mathbf{x})$ around the instantaneous center of mass position, $\bar{\mathbf{x}} = \overline{\mathbf{X}}$, and truncate the expansion to the leading order term²

$$\mathbf{v}(\mathbf{x}) = \mathbf{v}(\bar{\mathbf{x}} + \mathbf{x}') \approx \mathbf{v}(\bar{\mathbf{x}}) + (\mathbf{x}' \cdot \nabla) \mathbf{v}|_{\mathbf{x}=\bar{\mathbf{x}}}. \quad (37.21)$$

We are thus left with

$$\mathbf{L} = (\overline{\mathbf{X}} \wedge \mathbf{P}) + \int_{\mathcal{R}(\mathbf{v})} (\mathbf{x}' \wedge \mathbf{v}) \rho dV \quad (37.22a)$$

$$= (\overline{\mathbf{X}} \wedge \mathbf{P}) + \int_{\mathcal{R}(\mathbf{v})} [\mathbf{x}' \wedge \mathbf{v}(\bar{\mathbf{x}})] \rho dV + \int_{\mathcal{R}(\mathbf{v})} [\mathbf{x}' \wedge (\mathbf{x}' \cdot \nabla) \mathbf{v}(\bar{\mathbf{x}})] \rho dV. \quad (37.22b)$$

The velocity $\mathbf{v}(\bar{\mathbf{x}})$ can be removed from the integration since it is evaluated at the center of mass point. Hence, the second term in equation (37.22b) vanishes

$$\int_{\mathcal{R}(\mathbf{v})} [\mathbf{x}' \wedge \mathbf{v}(\bar{\mathbf{x}})] \rho dV = \left[\int_{\mathcal{R}(\mathbf{v})} \mathbf{x}' \rho dV \right] \wedge \mathbf{v}(\bar{\mathbf{x}}) = 0, \quad (37.23)$$

where $\int_{\mathcal{R}(\mathbf{v})} \mathbf{x}' \rho dV = 0$ by definition of the center of mass. The angular momentum is thus given by the two terms

$$\mathbf{L} = (\overline{\mathbf{X}} \wedge \mathbf{P}) + \int_{\mathcal{R}(\mathbf{v})} [\mathbf{x}' \wedge (\mathbf{x}' \cdot \nabla) \mathbf{v}(\bar{\mathbf{x}})] \rho dV. \quad (37.24)$$

The m' th component of the second term can be written

$$\int_{\mathcal{R}(\mathbf{v})} [\mathbf{x}' \wedge (\mathbf{x}' \cdot \nabla) \mathbf{v}(\bar{\mathbf{x}})]_m \rho dV = \epsilon_{mnp} \int_{\mathcal{R}(\mathbf{v})} x'_n [(\mathbf{x}' \cdot \nabla) \mathbf{v}(\bar{\mathbf{x}})]_p \rho dV \quad (37.25a)$$

$$= \epsilon_{mnp} \int_{\mathcal{R}(\mathbf{v})} x'_n x'_q \partial_q v(\bar{\mathbf{x}})_p \rho dV \quad (37.25b)$$

$$= \epsilon_{mnp} \left[\int_{\mathcal{R}(\mathbf{v})} x'_n x'_q \rho dV \right] \partial_q v(\bar{\mathbf{x}})_p. \quad (37.25c)$$

We removed the velocity derivatives

$$\partial_q v(\bar{\mathbf{x}})_p = \left[\frac{\partial v_p}{\partial x_q} \right]_{\mathbf{x}=\bar{\mathbf{x}}} \quad (37.26)$$

from the integral, since they are evaluated at the center of mass point and so do not participate in the integration.

37.4.2 Strain and vorticity

Following from the discussion in Section 14.2.5, we know that the velocity derivatives $\partial_q v_p$ appearing in equation (37.25c) form the components to a second order tensor. To expose the kinematics of this tensor, decompose it into its symmetric and anti-symmetric components

$$\partial_q v_p = \frac{1}{2}(\partial_q v_p + \partial_p v_q) + \frac{1}{2}(\partial_q v_p - \partial_p v_q) \equiv S_{qp} + A_{qp}. \quad (37.27)$$

²The velocity field evaluated at the center of mass position, $\mathbf{v}(\bar{\mathbf{x}})$, is not equal to the center of mass velocity: $\mathbf{v}(\bar{\mathbf{x}}) \neq \bar{\mathbf{v}}$.

The symmetric tensor

$$S_{qp} = \frac{1}{2}(\partial_q v_p + \partial_p v_q) \quad (37.28)$$

is associated with deformations in the fluid arising from strains; it is therefore called the *deformation* or *rate of strain* tensor. The anti-symmetric tensor can be written as

$$2A_{qp} = \partial_q v_p - \partial_p v_q \quad (37.29a)$$

$$= (\delta_{qm} \delta_{pn} - \delta_{qn} \delta_{pm}) \partial_m v_n \quad (37.29b)$$

$$= \epsilon_{sqp} \epsilon_{smn} \partial_m v_n \quad (37.29c)$$

$$= \epsilon_{sqp} \omega_s, \quad (37.29d)$$

where $\omega_s = \epsilon_{smn} \partial_m v_n$ are components to the vorticity pseudo-vector

$$\boldsymbol{\omega} = \nabla \wedge \mathbf{v}. \quad (37.30)$$

37.4.3 Relating angular momentum to strain and vorticity

Making use of the strain and vorticity brings the angular momentum for a connected material fluid into the form

$$L_m = (\overline{\mathbf{X}} \wedge \mathbf{P})_m + \epsilon_{mnp} \left[\int_{\mathcal{R}(\mathbf{v})} x'_n x'_q \rho dV \right] S_{qp} + \epsilon_{mnp} \left[\int_{\mathcal{R}(\mathbf{v})} x'_n x'_q \rho dV \right] A_{qp} \quad (37.31a)$$

$$= (\overline{\mathbf{X}} \wedge \mathbf{P})_m + \epsilon_{mnp} \left[\int_{\mathcal{R}(\mathbf{v})} x'_n x'_q \rho dV \right] S_{qp} + \frac{1}{2} \epsilon_{mnp} \epsilon_{sqp} \left[\int_{\mathcal{R}(\mathbf{v})} x'_n x'_q \rho dV \right] \omega_s \quad (37.31b)$$

$$= \underbrace{(\overline{\mathbf{X}} \wedge \mathbf{P})_m}_{\text{center of mass}} + \underbrace{\epsilon_{mnp} \left[\int_{\mathcal{R}(\mathbf{v})} x'_n x'_q \rho dV \right] S_{qp}}_{\text{strain contribution}} + \underbrace{\frac{1}{2} \left[\int_{\mathcal{R}(\mathbf{v})} (\mathbf{x}' \cdot \mathbf{x}' \delta_{ms} - x'_m x'_s) \rho dV \right] \omega_s}_{\text{vorticity contribution}}. \quad (37.31c)$$

As each point in the fluid can be considered the center of mass for an arbitrary material region, the decomposition (37.31c) is general.

- **CENTER OF MASS ANGULAR MOMENTUM:** The first term on the right hand side of equation (37.31c) arises from the angular momentum of the material region as measured with respect to the center of mass position. It has the form of that for a point particle (see equation (37.8a)). This term vanishes if the origin of the coordinate system is taken at the center of mass.
- **STRAINS:** The second contribution is proportional to fluid deformations acting to dilate or strain the fluid region (see Section 14.2.5). At each point of the fluid, deformations are measured by the deformation tensor S_{qp} . A rigid body moves by uniform translations and/or solid-body rotations, with the deformation tensor vanishing for rigid body motions (see Section 33.5). It is for this reason that the deformation tensor is so-named, as this tensor measures motions that are deviations or deformations relative to the motion of a rigid body. The contribution from these deformations is weighted by an integral of deviations of parcel position from the center of mass position. A closed form expression for this integral is available only for special shapes.

- **VORTICITY:** The third contributor to angular momentum in equation (37.31c) contains the vorticity as weighted by the moment of inertia tensor

$$I_{ms} \equiv \int_{\mathcal{R}(\mathbf{v})} (\mathbf{x}' \cdot \mathbf{x}' \delta_{ms} - x'_m x'_s) \rho dV. \quad (37.32)$$

Since the material region is evolving and is not rigid, the moment of inertia tensor is time dependent. The contribution

$$L_m^{\text{vorticity}} \equiv \frac{1}{2} I_{ms} \omega_s \quad (37.33)$$

has the same form as angular momentum for a rigid body, with one-half the vorticity playing the role of angular velocity (see equation (37.8c) for the point particle expression). Fluid vorticity hence contributes to angular momentum for a material region via its product with the moment of inertia tensor.

37.4.4 Comments

Angular momentum is computed relative to a chosen origin, whereas vorticity is an intrinsic property measuring the spin of the fluid at a point. So although they both offer measures of the rotational properties of fluid motion, they are quite distinct, especially when the fluid has non-zero strains. It is only for the special case of a solid-body motion that the strain contribution to angular momentum vanishes.

The discussion in this chapter supplements that from [Chatwin \(1973\)](#), as well as online notes “The Vorticity Equation and Conservation of Angular Momentum” from A.J. DeCaria.



Potential vorticity mechanics

The chapter details the foundational properties of potential vorticity (PV) and its evolution. The PV we encounter is sometimes referred to as the *Ertel* PV, which is the most fundamental form of PV arising in geophysical fluids ([Ertel, 1942](#)). The barotropic fluid forms a pedagogically useful starting point for the discussion. However, realistic geophysical flows are baroclinic, which are fluids where “PV thinking” is most powerful. The general method exploited for the construction of PV is to choose a scalar field to strategically orient the absolute vorticity. If the scalar is a material invariant, and it annihilates the baroclinicity vector, then PV is a material invariant. As so constructed, PV is a function of the chosen scalar.

For a barotropic fluid, the choice of scalar field is rather arbitrary, with preference given to one that is materially invariant. However, for a baroclinic fluid we are much more restricted since the scalar must orient vorticity in a direction to annihilate the torque from baroclinicity and, ideally, be itself materially invariant. We can only annihilate baroclinicity under certain restricted cases. Nonetheless, even when PV is not materially invariant, it remains a very important fluid property that partially constrains the fluid motion and provides insights into the interpretation of that motion.

READER’S GUIDE FOR THIS CHAPTER

This chapter requires a firm understanding of vorticity from Chapter [36](#) as well as good skills with vector calculus identities for Cartesian coordinates as detailed in Chapter [2](#). The concepts and methods developed in this chapter are fundamental to the notions of potential vorticity, much of which is encountered in the remainder of this part of the book as well as when studying balanced models in Part [VII](#).

38.1	Material invariance of PV in perfect fluids	880
38.1.1	Perfect barotropic fluid	880
38.1.2	Region between two scalar isosurfaces	880
38.1.3	Material invariance	881
38.1.4	Perfect baroclinic fluid	882
38.1.5	Some general remarks	882
38.2	PV and the seawater equation of state	884
38.2.1	Baroclinicity vector	884
38.2.2	PV based on potential density	884
38.2.3	An example EOS admitting a materially invariant PV	885
38.2.4	Further reading	886
38.3	PV evolution with non-conservative processes	886
38.4	Eulerian flux-form PV budget	887

38.4.1	Deriving the Eulerian flux-form PV budget	888
38.4.2	PV-substance and the PV flux	888
38.5	PV budget for a hydrostatic Boussinesq fluid	889
38.5.1	Potential vorticity	889
38.5.2	Potential vorticity flux vector	890
38.5.3	Kinematic derivation of the PV-substance flux	891
38.5.4	Potential vorticity flux vector	891
38.6	Exercises	893

38.1 Material invariance of PV in perfect fluids

In this section we derive the material invariance of potential vorticity (PV) for a perfect fluid (i.e., fluid without mixing). We make use of Kelvin's circulation theorem for an infinitesimal loop, in which case the primary object of interest is a particular component of the absolute vorticity. The discussion starts with a barotropic fluid, in which baroclinicity vanishes (Sections 36.3 and 36.5), and then we generalize to a baroclinic fluid.

38.1.1 Perfect barotropic fluid

Consider a perfect barotropic fluid. As for the shallow water discussion in Section 35.4.3, we apply Kelvin's circulation theorem (Section 36.3.2) to an infinitesimal material circuit within the fluid

$$\frac{D}{Dt}(\boldsymbol{\omega}_a \cdot \hat{\mathbf{n}} \delta\mathcal{S}) = 0, \quad (38.1)$$

with $\delta\mathcal{S}$ the area of the circuit. The conservation of PV is built from specializing this result. For that purpose, introduce a materially invariant scalar field

$$\frac{DC}{Dt} = 0. \quad (38.2)$$

38.1.2 Region between two scalar isosurfaces

We make use of isosurfaces of C to orient the material circuit and hence to orient the vorticity. In particular, referring to Figure 38.1, let the circuit bound a small cylinder whose two ends sit on two isosurfaces, $C - \delta C/2$ and $C + \delta C/2$. The tube volume is given by

$$\delta V = \delta\mathcal{S} \delta h, \quad (38.3)$$

where δh is the distance between the isosurfaces. The unit normal direction orienting the area $\delta\mathcal{S}$ is given by

$$\hat{\mathbf{n}} = \frac{\nabla C}{|\nabla C|}. \quad (38.4)$$

The separation between the two isosurfaces is related to the increment δC through

$$\delta C = \nabla C \cdot \delta \mathbf{x} = |\nabla C| \hat{\mathbf{n}} \cdot \delta \mathbf{x} = |\nabla C| \delta h. \quad (38.5)$$

This result takes on the equivalent expression

$$\delta C = |\nabla C| \delta h = (\hat{\mathbf{n}} \cdot \nabla C) \delta h = \frac{\delta C}{\delta n} \delta h, \quad (38.6)$$

so that the distance (or thickness) between the two isosurfaces is

$$\delta h = \frac{\delta C}{|\nabla C|}. \quad (38.7)$$

This equation has a straightforward geometric interpretation indicated in Figure 38.1. Namely, the geometric separation between the two isosurfaces is reduced in regions of strong scalar gradients and increased in regions of weak gradients.

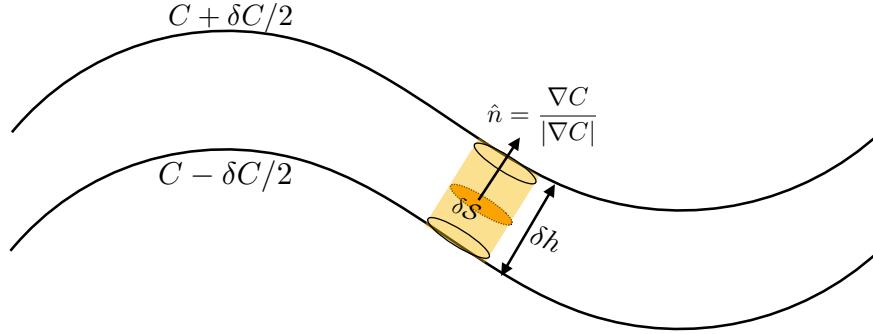


FIGURE 38.1: Illustrating the geometry of a cylindrical region of fluid between two iso-surfaces of a scalar field C . The volume of the region is $\delta V = \delta S \delta h$, with δh the thickness and δS the area. By convention, the normal vector, \hat{n} , points towards larger values of C . Notably, if C is a material invariant so that $DC/Dt = 0$, then so is its infinitesimal increment, $D(\delta C)/Dt = 0$. As per equation (38.7), the geometric thickness between the isosurfaces is related to the scalar field increment by $\delta h = \delta C/|\nabla C|$, so that the larger the magnitude of the gradient in the scalar field, the smaller the layer thickness. For a baroclinic fluid material invariance of PV holds only if we can find a scalar field such that $\hat{n} \cdot \mathbf{B} = 0$, with $\mathbf{B} = (\nabla \rho \wedge \nabla p)/\rho^2$ the baroclinicity.

38.1.3 Material invariance

We now have the necessary pieces in place to write

$$\omega_a \cdot \hat{n} \delta S = \frac{\omega_a \cdot \nabla C}{|\nabla C|} \delta S \quad (38.8a)$$

$$= \frac{\omega_a \cdot \nabla C}{|\nabla C|} \frac{\delta V}{\delta h} \quad (38.8b)$$

$$= (\omega_a \cdot \nabla C) \frac{\delta V}{\delta C} \quad (38.8c)$$

$$= \frac{\omega_a \cdot \nabla C}{\rho} \frac{\rho \delta V}{\delta C}. \quad (38.8d)$$

Mass is materially invariant so that

$$\frac{D(\rho \delta V)}{Dt} = 0. \quad (38.9)$$

Likewise, by assumption C is materially invariant so that the increment between two C isosurfaces is materially invariant

$$\frac{D(\delta C)}{Dt} = 0. \quad (38.10)$$

Bringing these elements into Kelvin's circulation theorem (38.1) leads us to conclude that the potential vorticity, Q , is also materially invariant

$$Q = \frac{\omega_a \cdot \nabla C}{\rho} = \frac{\nabla \cdot (\omega_a C)}{\rho} \implies \frac{DQ}{Dt} = 0. \quad (38.11)$$

This expression for the PV is the most general form and it is often referred to as the *Ertel PV* ([Ertel, 1942](#)). The first expression shows the numerator as the projection of the absolute vorticity into the direction normal to tracer isosurfaces. Conversely, it is a measure of the C stratification in the direction of the absolute vorticity vector. The second expression follows since the absolute vorticity has zero divergence so that the numerator is a total divergence. We make use of this second form when discussing PV budgets in Section 38.4 and layer integrated PV Section 39.3.

38.1.4 Perfect baroclinic fluid

Consider the case of a perfect baroclinic fluid, in which Kelvin's circulation theorem for an infinitesimal circuit takes the form

$$\frac{D}{Dt}(\boldsymbol{\omega}_a \cdot \hat{\mathbf{n}} \delta S) = \mathbf{B} \cdot \hat{\mathbf{n}} \delta S, \quad (38.12)$$

where the source on the right hand side arises from the baroclinicity vector discussed in Sections 36.3 and 36.5.

$$\mathbf{B} = \frac{\nabla \rho \wedge \nabla p}{\rho^2}. \quad (38.13)$$

Now assume there exists a materially invariant scalar, θ , that annihilates the baroclinicity vector as in Figure 38.2, so that

$$\mathbf{B} \cdot \hat{\mathbf{n}} = \frac{\mathbf{B} \cdot \nabla \theta}{|\nabla \theta|} = 0. \quad (38.14)$$

In that case, the barotropic derivation detailed earlier in this section follows for the baroclinic case so that PV remains materially invariant

$$\frac{DQ}{Dt} = 0 \quad \text{where} \quad Q = \frac{\boldsymbol{\omega}_a \cdot \nabla \theta}{\rho}. \quad (38.15)$$

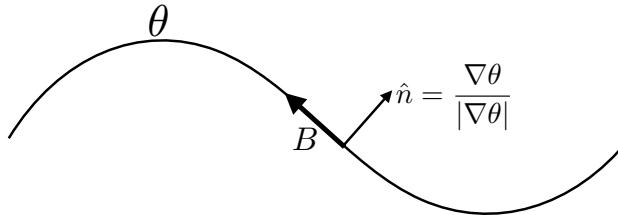


FIGURE 38.2: Material invariance of PV is ensured for perfect fluids that allow for a materially invariant scalar to annihilate baroclinicity, meaning that $\hat{\mathbf{n}} \cdot \mathbf{B} = 0$ where $\hat{\mathbf{n}} = \nabla \theta / |\nabla \theta|$. In this figure we depict the potential temperature field with a baroclinicity vector aligned with isolines of θ .

Existence of a materially invariant PV for perfect baroclinic fluids depends on the existence of a materially invariant scalar that annihilates the baroclinicity vector. The most common choice for this scalar in geophysical fluid applications is the buoyancy, which is typically measured by entropy or potential temperature in the atmosphere and potential density in the ocean. We have more to say on the chosen scalar field in the remainder of this chapter as well as in Section 39.2.

38.1.5 Some general remarks

Perfect fluid PV material invariance \leftrightarrow Kelvin's circulation theorem

Kelvin's circulation theorem from Section 36.3.2 is at the heart of the derivations presented in this section, with the theorem applied to a strategically chosen infinitesimal loop. Because the

loop is tiny, we convert the line integral expression of Kelvin's theorem into a statement about the material evolution of absolute vorticity projected onto the normal direction of the loop, and multiplied by the loop area. We further specialize the theorem to a cylindrical region between two isosurfaces of a materially invariant scalar field. For the perfect barotropic fluid, we merely require that both mass and scalar to be materially invariant to realize a potential vorticity that is also materially invariant. Material invariance of PV for a baroclinic fluid requires a scalar field that is both materially invariant *and* that annihilates the baroclinicity vector. We have more to say in regards to the availability of such scalar fields when considering entropic PV in the remainder of this chapter, whereby potential temperature is the chosen scalar field.

There are numerous forms for PV

The expression (38.11) is, on first glance, quite distinct from the shallow water potential vorticity, $Q = (\zeta + f)/h$, studied in Chapter 35 (see equation (35.37)). However, as shown in Section 46.2, they are closely related for the special case of entropic PV in an incompressible fluid when formulated using isopycnal/isentropic coordinates. Even so, there are a variety of other forms for PV, with the forms (and physical dimensions) depending on the assumptions made in regards to the fluid flow and its thermodynamic properties. We encounter some further forms of PV in the remainder of this chapter, as well as in the oceanic PV discussions of Chapter 46 and in our study of balanced models in Part VII. The review paper by Müller (1995) offers a lucid presentation of PV and its many forms encountered in physical oceanography.

Motivating the adjective “potential”

In Section 35.4.2 we motivated the advective “potential” for the shallow water potential vorticity. We do so here for Ertel’s potential vorticity. For that purpose, write PV in the form

$$Q = \frac{\omega_a \cdot \hat{n}}{\rho} |\nabla\theta| \quad \text{with} \quad \hat{n} = \frac{\nabla\theta}{|\nabla\theta|}. \quad (38.16)$$

As noted in Section 2.1 of Müller (1995), in cases where ρ is roughly a constant, and when Q is materially invariant, the component of the absolute vorticity in the direction parallel to $\nabla\theta$ increases when the fluid parcel moves into a region where $|\nabla\theta|$ decreases. There is hence a “release” of absolute vorticity aligned with $\nabla\theta$ in regions where θ isosurfaces are spread, and a withdrawal of absolute vorticity where θ isosurfaces are tightly packed. We thus conceive of PV as the “potential” for releasing absolute vorticity oriented in the direction parallel to $\nabla\theta$.

PV as a dynamical tracer

We refer to PV as a *dynamical tracer* since it depends directly on the velocity field through the vorticity. It also depends on the thermodynamic tracer, θ , so that PV brings together dynamical and thermodynamical information into a single package. In contrast, material tracers such as salinity, and thermodynamic tracers such as θ , are properties of the fluid whose distribution depends on the velocity field but whose local measurement does not require knowledge of velocity.

PV is the grand unifier for geophysical fluid mechanics

Entropic PV provides a connection between vorticity (mechanics) and stratification (thermodynamics). By connecting these two basic facets of geophysical fluid flows, the study of PV and its conservation properties provides a powerful and unique lens to help rationalize the huge variety of

geophysical flow regimes, and to predict their response to changes in forcing. It is for this reason that PV is sometimes considered the grand unifying concept in geophysical fluid mechanics.

PV as a diagnostic tracer

Suppose we have an initial flow field in which $Q = 0$ so that the absolute vorticity lives within constant θ surfaces. In an inviscid and adiabatic fluid, material conservation of Q means that ω_a remains within constant θ surfaces for all time. Hence, if we know the evolution of θ , then we know the evolution of vortex lines defined by ω_a , which in turn allows for the inference of a number of further flow properties. This particular example offers a hint at the multiple applications of “PV thinking” to understand and predict fluid motion, with [Hoskins \(1991\)](#) providing an elegant survey of such thinking.

38.2 PV and the seawater equation of state

As seen in Section 38.1, material invariance of PV for a perfect fluid requires a materially invariant scalar field to annihilate the baroclinicity vector. In the ocean there is generally no such scalar when there is also a realistic seawater equation of state. Hence, the ocean cannot be considered a perfect fluid even in the absence of friction and diabatic processes. Nonetheless, as discussed in this section, there are important approximate cases that allow for material PV invariance.

38.2.1 Baroclinicity vector

Recall the baroclinicity vector given by (Sections 36.3 and 36.5)

$$\mathbf{B} = \frac{\nabla\rho \wedge \nabla p}{\rho^2}. \quad (38.17)$$

If we take the *in situ* density as the scalar field to define PV, then $\mathbf{B} \cdot \nabla\rho = 0$. However, *in situ* density is not a conserved scalar in the ocean due to pressure effects. Namely, the material time derivative of $\rho = \rho(S, \theta, p)$ is

$$\frac{D\rho}{Dt} = \frac{\partial\rho}{\partial S} \frac{DS}{Dt} + \frac{\partial\rho}{\partial\theta} \frac{D\theta}{Dt} + \frac{\partial\rho}{\partial p} \frac{Dp}{Dt}. \quad (38.18)$$

Even when salinity and potential temperature are materially constant, $DS/Dt = 0$ and $D\theta/Dt = 0$, the *in situ* density has a nonzero material time evolution due to pressure changes, $Dp/Dt \neq 0$. Material changes in the pressure of a fluid element arise even in the absence of irreversible mixing. In general, such *mechanical changes* arise due to the gradients in the pressure field that the fluid element moves through. Given that pressure effects *in situ* density, with such effects occurring even in a perfect fluid, we conclude that *in situ* density is not an appropriate scalar for use in developing a materially invariant PV.

38.2.2 PV based on potential density

Potential density is commonly used in oceanography (see Section 25.3.4), with potential density the *in situ* density referenced to a chosen pressure.¹ We write potential density as

$$\sigma(S, \theta) = \rho(S, \theta, p = p_R). \quad (38.19)$$

¹Oceanographers often choose the reference pressure as the standard atmospheric surface pressure. However, that is not required for the following formalism to hold; any reference pressure is suitable.

Potential density provides a global measure of buoyancy, with accuracy of the measure compromised in the presence of a nonlinear equation of state.

The material time derivative of potential density is

$$\frac{D\sigma}{Dt} = \frac{\partial\sigma}{\partial S} \frac{DS}{Dt} + \frac{\partial\sigma}{\partial\theta} \frac{D\theta}{Dt}, \quad (38.20)$$

which vanishes in the absence of irreversible material changes to salinity and potential temperature. When using potential density as the scalar field for PV, the baroclinicity vector projects onto the diapycnal direction according to

$$\rho^2 \mathbf{B} \cdot \nabla\sigma = (\nabla\rho \wedge \nabla p) \cdot \nabla\sigma \quad (38.21a)$$

$$= (\nabla\sigma \wedge \nabla\rho) \cdot \nabla p \quad (38.21b)$$

$$= [(\sigma_S \nabla S + \sigma_\theta \nabla\theta) \wedge (\rho_S \nabla S + \rho_\theta \nabla\theta + \rho_p \nabla p)] \cdot \nabla p \quad (38.21c)$$

$$= [(\sigma_S \nabla S + \sigma_\theta \nabla\theta) \wedge (\rho_S \nabla S + \rho_\theta \nabla\theta)] \cdot \nabla p \quad (38.21d)$$

$$= [\sigma_S \nabla S \wedge \rho_\theta \nabla\theta + \sigma_\theta \nabla\theta \wedge \rho_S \nabla S] \cdot \nabla p \quad (38.21e)$$

$$= (\sigma_S \rho_\theta - \sigma_\theta \rho_S) (\nabla S \wedge \nabla\theta) \cdot \nabla p, \quad (38.21f)$$

where we used the shorthand notation for partial derivatives

$$\rho_S = \frac{\partial\rho}{\partial S} \quad \sigma_S = \frac{\partial\sigma}{\partial S} \quad (38.22a)$$

$$\rho_\theta = \frac{\partial\rho}{\partial\theta} \quad \sigma_\theta = \frac{\partial\sigma}{\partial\theta}. \quad (38.22b)$$

Note that the triple product, $(\nabla S \wedge \nabla\theta) \cdot \nabla p$, also appears in the discussion of neutral helicity in Section 25.6 (see equation (25.58)). Equation (38.21f) allows us to identify cases where the baroclinicity vector is annihilated, $\mathbf{B} \cdot \nabla\sigma = 0$, thus yielding a materially invariant PV in the absence of irreversible processes.

- UNIFORM SALINITY OR UNIFORM POTENTIAL TEMPERATURE: If salinity or potential temperature are spatially uniform, then $\mathbf{B} \cdot \nabla\sigma = 0$.
- ADDITIVE PRESSURE DEPENDENCE TO THE *in situ* DENSITY: There is an exact PV conservation principle if the thermodynamic pre-factor $\sigma_S \rho_\theta - \sigma_\theta \rho_S$ vanishes. However, the ocean has a pressure dependent equation of state and this pressure dependence generally means that $\mathbf{B} \cdot \nabla\sigma \neq 0$. Nonetheless, the baroclinicity vector is annihilated if the *in situ* density has a pressure dependence that is additive, so that we can write

$$\rho(S, \theta, p) = \sigma(S, \theta) + F(p) - F(p_R) \implies \sigma_S \rho_\theta - \sigma_\theta \rho_S = 0, \quad (38.23)$$

which then leads to a materially invariant PV. Notably, we did not assume the equation of state to be linear; only that it has the special functional form in equation (38.23). For some cases, we may assume F to be a constant, in which case there is no pressure dependence; i.e., the *in situ* density is the same as potential density.

38.2.3 An example EOS admitting a materially invariant PV

An explicit realization of the equation of state (38.23) can be found by taking a Taylor series expansion of the *in situ* density around the reference pressure, and evaluating the derivatives

in the expansion in terms of a chosen reference pressure, reference salinity, and reference potential temperature

$$\rho(S, \theta, p) \approx \sigma(S, \theta) + (p - p_R) \underbrace{\left[\frac{\partial \rho}{\partial p} \right]_{S=S_R, \theta=\theta_R, p=p_R}}_{F(p) - F(p_R)} + H.O.T. \quad (38.24)$$

where *H.O.T.* symbolizes higher order terms. This approach ignores the salinity and potential temperature dependence of terms in the Taylor series expansion. Ignoring this dependence is a rather good approximation for many purposes since the ocean sound speed is not far from a constant

$$c_s^{-2} = \frac{\partial \rho}{\partial p} \approx \text{constant}. \quad (38.25)$$

In this case, the equation of state takes the form

$$\rho(S, \theta, p) \approx \sigma(S, \theta) + \frac{p - p_R}{c_s^2}, \quad (38.26)$$

where again $\sigma(S, \theta) = \rho(S, \theta, p_R)$ is the potential density referenced to $p = p_R$.

38.2.4 Further reading

[Straub \(1999\)](#) presents a discussion of ocean potential vorticity with a focus on the source of potential vorticity arising from a nonzero thermobaricity parameter, $\mathcal{T} = \partial_p(\alpha/\beta)$ (see Section 50.4.4). The presentation given here follows that given in Section 4.5.4 of [Vallis \(2017\)](#).

38.3 PV evolution with non-conservative processes

Thus far we have considered perfect fluids, with the use of Kelvin's circulation theorem a suitable framework to derive the material invariance of PV. In this section we move beyond the perfect fluid by considering a real fluid that contains non-conservative processes. PV is no longer materially invariant when exposed to non-conservative processes. However, as detailed in Chapter 39, there remain some rather remarkable global conservation properties for PV budgets within isentropic layers.

To develop the PV budget in the presence of non-conservative processes, we pursue an algebraic approach that starts from the vorticity equation (36.30)

$$\rho \frac{D(\omega_a / \rho)}{Dt} = (\omega_a \cdot \nabla) v + B + \nabla \wedge F, \quad (38.27)$$

where F is the acceleration from non-conservative forces and B the baroclinicity vector. Furthermore, we introduce a non-conservative scalar field

$$\frac{DC}{Dt} = \dot{C}, \quad (38.28)$$

with \dot{C} arising from diffusion, sources, or other irreversible processes.

As part of the manipulations we make use of the identity

$$(\omega_a \cdot \nabla) \frac{DC}{Dt} = \omega_a \cdot \frac{D(\nabla C)}{Dt} + [(\omega_a \cdot \nabla) v] \cdot \nabla C, \quad (38.29)$$

which is readily proven by expanding terms and assuming Cartesian coordinates. Rearrangement, and use of the scalar equation (38.28), leads to

$$\boldsymbol{\omega}_a \cdot \frac{D(\nabla C)}{Dt} = (\boldsymbol{\omega}_a \cdot \nabla) \dot{C} - [(\boldsymbol{\omega}_a \cdot \nabla) \mathbf{v}] \cdot \nabla C. \quad (38.30)$$

Now project the vorticity equation (38.27) onto the direction normal to the C isosurfaces

$$\rho \nabla C \cdot \frac{D(\boldsymbol{\omega}_a/\rho)}{Dt} = \nabla C \cdot [(\boldsymbol{\omega}_a \cdot \nabla) \mathbf{v}] + \nabla C \cdot (\mathbf{B} + \nabla \wedge \mathbf{F}). \quad (38.31)$$

The sum of equations (38.30) and (38.31) leads to

$$\rho \frac{D(\nabla C \cdot \boldsymbol{\omega}_a/\rho)}{Dt} = (\boldsymbol{\omega}_a \cdot \nabla) \dot{C} + \nabla C \cdot (\mathbf{B} + \nabla \wedge \mathbf{F}). \quad (38.32)$$

This equation is general so that it applies to any scalar field.

To simplify the source terms on the right hand side of equation (38.32), follow the discussion from Section 38.1.4 by assuming there is a special scalar field that annihilates the baroclinicity vector.² We generally denote this scalar field as θ , which we use in this chapter as a general designation for potential temperature, specific entropy, buoyancy, or potential density. We thus have

$$\nabla \theta \cdot \mathbf{B} = 0, \quad (38.33)$$

which in turn leads to the potential vorticity equation in the presence of irreversible processes

$$\rho \frac{DQ}{Dt} = (\boldsymbol{\omega}_a \cdot \nabla) \dot{\theta} + \nabla \theta \cdot (\nabla \wedge \mathbf{F}), \quad (38.34)$$

where the entropic-PV is again given by

$$Q = \frac{\boldsymbol{\omega}_a \cdot \nabla \theta}{\rho}. \quad (38.35)$$

As defined, the material evolution of PV is affected by diabatic processes (heating and cooling) as well as friction. Absent these irreversible processes leaves us with the same PV conservation statement (38.15) derived for the perfect fluid using Kelvin's circulation theorem. In their presence, the PV of a fluid element can be either generated or destroyed depending on details of the irreversible process. Such source/sink regions of potential vorticity are often localized to regions of mixing as well as to boundaries where strong mechanical and/or buoyancy processes are active. The study of how PV is materially modified by irreversible processes forms an important area of research in PV dynamics.

38.4 Eulerian flux-form PV budget

The material invariance of PV is an example of a material or Lagrangian conservation property of perfect fluids, with the material conservation statement $\rho DQ/Dt = 0$ having its Eulerian flux-form expression

$$\frac{\partial(\rho Q)}{\partial t} + \nabla \cdot (\rho \mathbf{v} Q) = 0 \quad \text{perfect fluid.} \quad (38.36)$$

²In Section 39.2 we study what happens when no such scalar exists.

Following the formalism established for material tracers in Section 16.3, the flux-form local conservation law (38.36) leads then to conservation properties over finite regions, which we refer to as *global conservation* laws. In this section we examine the Eulerian flux-form budget in the presence of non-conservative processes. As we will see, the Eulerian evolution of PV will continue to be determined by the convergence of a flux, thus allowing for natural extensions to global conservation laws.

38.4.1 Deriving the Eulerian flux-form PV budget

To transform the material evolution equation (38.34) into a flux-form Eulerian equation we make use of the following identities

$$\frac{D}{Dt} = \frac{\partial}{\partial t} + \mathbf{v} \cdot \nabla \quad \text{relating material and Eulerian time changes} \quad (38.37a)$$

$$\frac{D\rho}{Dt} = -\rho \nabla \cdot \mathbf{v} \quad \text{mass conservation} \quad (38.37b)$$

$$(\boldsymbol{\omega}_a \cdot \nabla) \dot{\theta} = \nabla \cdot (\boldsymbol{\omega}_a \dot{\theta}) \quad \nabla \cdot \boldsymbol{\omega}_a = 0 \quad (38.37c)$$

$$\nabla \theta \cdot (\nabla \wedge \mathbf{F}) = \nabla \cdot (\mathbf{F} \wedge \nabla \theta) \quad \text{divergence of curl vanishes.} \quad (38.37d)$$

The identity (38.37d) follows from

$$\nabla \theta \cdot (\nabla \wedge \mathbf{F}) = \nabla \cdot (\theta \nabla \wedge \mathbf{F}) = \nabla \cdot [\nabla \wedge (\theta \mathbf{F}) - \nabla \theta \wedge \mathbf{F}] = \nabla \cdot (\mathbf{F} \wedge \nabla \theta), \quad (38.38)$$

where a vanishing divergence of a curl is needed to reach the first and third equalities. These identities then lead to the material evolution equation

$$\rho \frac{DQ}{Dt} = \nabla \cdot [\boldsymbol{\omega}_a \dot{\theta} + \mathbf{F} \wedge \nabla \theta]. \quad (38.39)$$

Now converting the material time derivative into its Eulerian expression, and making use of mass conservation, renders the flux-form Eulerian PV budget equation

$$\frac{\partial(\rho Q)}{\partial t} + \nabla \cdot [\rho Q \mathbf{v} - \boldsymbol{\omega}_a \dot{\theta} - \mathbf{F} \wedge \nabla \theta] = 0. \quad (38.40)$$

38.4.2 PV-substance and the PV flux

The budget equation (38.40) says that the density-weighted potential vorticity

$$\rho Q = \boldsymbol{\omega}_a \cdot \nabla \theta \quad (38.41)$$

has a local time tendency determined by the convergence of the PV flux vector

$$\frac{\partial(\rho Q)}{\partial t} = -\nabla \cdot \mathbf{J}_Q \quad \text{with} \quad \mathbf{J}_Q = \rho Q \mathbf{v} - \boldsymbol{\omega}_a \dot{\theta} - \mathbf{F} \wedge \nabla \theta. \quad (38.42)$$

The budget (38.40) follows a form similar to material tracers detailed in Chapter 16, though with some specific terms in the flux vector, \mathbf{J}_Q . The correspondence suggests that one consider equation (38.40) as the local budget for *PV-substance*, with Q the concentration of PV-substance and \mathbf{J}_Q its flux. This interpretation is pursued further in Chapter 39 when exposing the rather novel properties of budgets for PV-substance when integrated over regions bounded by isentropes.

The first term in the PV-substance flux vector (38.42) arises from the advection of PV-substance; the second contribution arises from diabatic processes; and the third from the curl of the friction

vector. Note that there is a nonzero friction contribution only when the frictional acceleration has a component that is not parallel to $\nabla\theta$. In this manner we can think of the friction vector as contributing to a “torque” that rotates the θ isosurfaces as it contributes to the evolution of PV-substance (see Figure 38.3).

The time tendency for PV-substance, $\partial_t(\rho Q)$, is unchanged by adding the curl of a vector to \mathbf{J}_Q . This ambiguity manifests a *gauge freedom* afforded the PV-substance flux vector. We offer a general discussion of gauge freedom in Section 17.5.1 and provide more specifics for PV in Chapter 39.

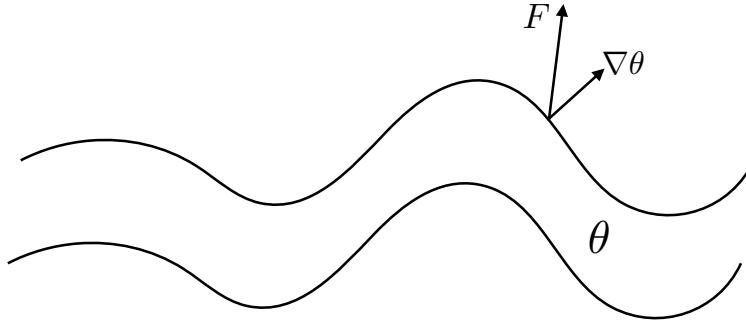


FIGURE 38.3: The contribution from friction to the PV flux is given by $\mathbf{J}_{\text{friction}} = \nabla\theta \wedge \mathbf{F}$, which is nonzero only when the frictional acceleration is not fully aligned with $\nabla\theta$. Friction creates PV by acting to rotate θ isosurfaces, so if friction is aligned with $\nabla\theta$, or when there is no spatial structure to $\nabla\theta \wedge \mathbf{F}$ (i.e., zero divergence), then friction does not contribute to PV evolution. This interpretation is analogous to that given to the effects from baroclinicity on vorticity given in Section 36.5.

38.5 PV budget for a hydrostatic Boussinesq fluid

Building on the vorticity budget developed in Section 36.8, we here develop the potential vorticity substance budgets for a hydrostatic Boussinesq fluid in the presence of diabatic and frictional forcing. For that purpose, we recall the vorticity equation for a hydrostatic and Boussinesq fluid (36.99)

$$\frac{D\omega_a^{\text{hy}}}{Dt} = \underbrace{(\omega_a^{\text{hy}} \cdot \nabla) \mathbf{v}}_{\text{stretching + tilting}} + \underbrace{\nabla \wedge \hat{\mathbf{z}} b}_{\text{baroclinicity}} + \underbrace{\nabla \wedge \mathbf{F}}_{\text{friction curl}}. \quad (38.43)$$

38.5.1 Potential vorticity

Potential vorticity evolves in the absence of baroclinicity, which is eliminated from the vorticity equation (38.43) by projecting the absolute vorticity onto the direction normal to buoyancy surfaces

$$\nabla b \cdot \frac{D\omega_a^{\text{hy}}}{Dt} = \nabla b \cdot [(\omega_a^{\text{hy}} \cdot \nabla) \mathbf{v}] + \nabla b \cdot (\nabla \wedge \mathbf{F}), \quad (38.44)$$

where we used

$$\nabla b \cdot (\nabla \wedge \hat{\mathbf{z}} b) = 0. \quad (38.45)$$

We next make use of the identity

$$\frac{D(\partial b / \partial x^i)}{Dt} = \frac{\partial}{\partial x^i} \left[\frac{\partial b}{\partial t} \right] - \nabla b \cdot \frac{\partial \mathbf{v}}{\partial x^i} = \frac{\partial \dot{b}}{\partial x^i} - \nabla b \cdot \frac{\partial \mathbf{v}}{\partial x^i}, \quad (38.46)$$

so that

$$\boldsymbol{\omega}_a^{hy} \cdot \left[\frac{D \nabla b}{Dt} \right] = \boldsymbol{\omega}_a^{hy} \cdot \nabla \dot{b} - \nabla b \cdot [(\boldsymbol{\omega}_a^{hy} \cdot \nabla) \mathbf{v}]. \quad (38.47)$$

Making use of this result in equation (38.44) renders

$$\nabla b \cdot \frac{D \boldsymbol{\omega}_a^{hy}}{Dt} + \boldsymbol{\omega}_a^{hy} \cdot \frac{D \nabla b}{Dt} = \boldsymbol{\omega}_a^{hy} \cdot \nabla \dot{b} + \nabla b \cdot (\nabla \wedge \mathbf{F}), \quad (38.48)$$

which leads to

$$\frac{DQ}{Dt} = \boldsymbol{\omega}_a^{hy} \cdot \nabla \dot{b} + \nabla b \cdot (\nabla \wedge \mathbf{F}) \quad (38.49)$$

where

$$Q = \boldsymbol{\omega}_a^{hy} \cdot \nabla b = \boldsymbol{\omega}^{hy} \cdot \nabla b + f \frac{\partial b}{\partial z} \quad (38.50)$$

is the potential vorticity for a rotating hydrostatic Boussinesq fluid. Potential vorticity is materially invariant for the inviscid and adiabatic case, in which $\mathbf{F} = 0$ and $\dot{b} = 0$.

It is sometimes useful to split the hydrostatic vorticity into its vertical and horizontal terms as per equation (36.89). In this way, PV takes on the form

$$Q = \frac{\partial u}{\partial z} \frac{\partial b}{\partial y} - \frac{\partial v}{\partial z} \frac{\partial b}{\partial x} + (\zeta + f) \frac{\partial b}{\partial z} = \hat{z} \cdot \left[\frac{\partial \mathbf{u}}{\partial z} \wedge \nabla b \right] + (\zeta + f) \frac{\partial b}{\partial z}. \quad (38.51)$$

38.5.2 Potential vorticity flux vector

The material form of the PV equation (38.49) is converted into its Eulerian flux-form via

$$\frac{\partial Q}{\partial t} + \nabla \cdot (\mathbf{v} Q) = \boldsymbol{\omega}_a^{hy} \cdot \nabla \dot{b} + \nabla b \cdot (\nabla \wedge \mathbf{F}) \quad (38.52a)$$

$$= \nabla \cdot [\dot{b} \boldsymbol{\omega}_a^{hy} + b (\nabla \wedge \mathbf{F})] \quad (38.52b)$$

$$= \nabla \cdot [\dot{b} \boldsymbol{\omega}_a^{hy} + \nabla \wedge (b \mathbf{F}) - \nabla b \wedge \mathbf{F}] \quad (38.52c)$$

$$= \nabla \cdot [\dot{b} \boldsymbol{\omega}_a^{hy} - \nabla b \wedge \mathbf{F}], \quad (38.52d)$$

where we used

$$\nabla \cdot \mathbf{v} = 0 \quad (38.53a)$$

$$\nabla \cdot \boldsymbol{\omega}_a^{hy} = 0 \quad (38.53b)$$

$$\nabla \cdot (\nabla \wedge \mathbf{F}) = 0 \quad (38.53c)$$

$$\nabla \cdot [\nabla \wedge (b \mathbf{F})] = 0. \quad (38.53d)$$

The conservation equation (38.52d) allows us to identify a potential vorticity flux vector

$$\mathbf{J}_Q = \mathbf{v} Q - \dot{b} \boldsymbol{\omega}_a^{hy} + \nabla b \wedge \mathbf{F} + \nabla \wedge \mathbf{A}, \quad (38.54)$$

so that the PV equation takes the form

$$\frac{\partial Q}{\partial t} + \nabla \cdot \mathbf{J}_Q = 0. \quad (38.55)$$

The potential vorticity flux (38.54) is comprised of an advective term

$$\mathbf{J}_{\text{advective}} = \mathbf{v} Q, \quad (38.56)$$

and non-advective terms arising from diabatic and frictional forcing,

$$\mathbf{J}_{\text{non-advective}} = -\dot{b}\boldsymbol{\omega}_{\text{a}}^{\text{hy}} + \nabla b \wedge \mathbf{F}, \quad (38.57)$$

as well as a gauge term,

$$\mathbf{J}_{\text{gauge}} = \nabla \wedge \mathbf{A}, \quad (38.58)$$

whose form remains unspecified.

38.5.3 Kinematic derivation of the PV-substance flux

Following the discussion in Section 39.1.2, we consider a purely kinematic derivation of the PV flux vector. For that purpose write the hydrostatic Boussinesq PV in the form

$$Q = \boldsymbol{\omega}_{\text{a}}^{\text{hy}} \cdot \nabla b = \nabla \cdot (b \boldsymbol{\omega}_{\text{a}}^{\text{hy}}), \quad (38.59)$$

which follows since $\nabla \cdot \boldsymbol{\omega}_{\text{a}}^{\text{hy}} = 0$. Taking the Eulerian time derivative then leads to

$$\frac{\partial Q}{\partial t} = \nabla \cdot \left[\frac{\partial b}{\partial t} \boldsymbol{\omega}_{\text{a}}^{\text{hy}} + b \frac{\partial \boldsymbol{\omega}_{\text{a}}^{\text{hy}}}{\partial t} \right] = \nabla \cdot \left[\frac{\partial b}{\partial t} \boldsymbol{\omega}_{\text{a}}^{\text{hy}} + \frac{\partial \mathbf{u}}{\partial t} \wedge \nabla b \right], \quad (38.60)$$

which made use of the identities

$$\frac{\partial \boldsymbol{\omega}_{\text{a}}^{\text{hy}}}{\partial t} = \frac{\partial \boldsymbol{\omega}^{\text{hy}}}{\partial t} = \frac{\partial (\nabla \wedge \mathbf{u})}{\partial t} \quad \text{and} \quad \nabla \cdot [\nabla \wedge (b \mathbf{u})] = 0. \quad (38.61)$$

The PV thus evolves according to

$$\frac{\partial Q}{\partial t} = -\nabla \cdot \tilde{\mathbf{J}}_Q \quad (38.62)$$

where the PV flux is given by

$$\tilde{\mathbf{J}}_Q = -\frac{\partial b}{\partial t} \boldsymbol{\omega}_{\text{a}}^{\text{hy}} - \frac{\partial \mathbf{u}}{\partial t} \wedge \nabla b. \quad (38.63)$$

This form of the PV flux manifests the impermeability property of Chapter 39 since

$$\hat{\mathbf{n}} \cdot (\tilde{\mathbf{J}}_Q / Q) = -\frac{1}{|\nabla b|} \frac{\partial b}{\partial t} = \hat{\mathbf{n}} \cdot \mathbf{v}_{b\perp}, \quad (38.64)$$

where $\hat{\mathbf{n}} = \nabla b / |\nabla b|^{-1}$ is the unit normal for a buoyancy surface, and with $\mathbf{v}_{b\perp}$ the velocity of a point on the isopycnal that satisfies

$$\frac{\partial b}{\partial t} + \mathbf{v}_{b\perp} \cdot \nabla b = 0. \quad (38.65)$$

38.5.4 Potential vorticity flux vector

We introduced the Bernoulli potential and Bernoulli theorem in Section 22.6.2. Schär (1993) provided a generalization of Bernoulli's Theorem, with Marshall (2000), Marshall et al. (2001), and Polton and Marshall (2007) applying this theorem to oceanic contexts. In particular, Marshall (2000) and Polton and Marshall (2007) work with a Boussinesq and hydrostatic ocean. We here consider the details of their analysis, with some of the material in this section building from the more general discussion in Section 39.2.1.

Momentum equation

Start by exposing the Boussinesq form of the Bernoulli potential from the vector-invariant velocity equation. For this purpose, return to the horizontal momentum equation (36.94), and expand the expressions for the perturbation pressure and the buoyancy

$$\frac{\partial \mathbf{u}}{\partial t} + (\mathbf{f} + \boldsymbol{\omega}^{\text{hy}}) \wedge \mathbf{v} = -\nabla(\varphi + |\mathbf{u}|^2/2) + \hat{\mathbf{z}} b + \mathbf{F} \quad (38.66\text{a})$$

$$= -\nabla(|\mathbf{u}|^2/2) - \frac{1}{\rho_0} \nabla(p - p_0) - \hat{\mathbf{z}} \left[\frac{g(\rho - \rho_0)}{\rho_0} \right] + \mathbf{F} \quad (38.66\text{b})$$

$$= -\nabla(|\mathbf{u}|^2/2) - \frac{1}{\rho_0} \nabla p - \hat{\mathbf{z}} \frac{g\rho}{\rho_0} + \mathbf{F} \quad (38.66\text{c})$$

$$= -\nabla \left[\frac{|\mathbf{u}|^2}{2} + \frac{p}{\rho_0} \right] - \hat{\mathbf{z}} \left[\frac{g\rho - g\rho_o + g\rho_o}{\rho_0} \right] + \mathbf{F} \quad (38.66\text{d})$$

$$= -\nabla \left[\frac{|\mathbf{u}|^2}{2} + \frac{p}{\rho_0} + g z \right] - \hat{\mathbf{z}} \left[\frac{g(\rho - \rho_o)}{\rho_0} \right] + \mathbf{F} \quad (38.66\text{e})$$

$$= -\nabla \mathcal{B} + \hat{\mathbf{z}} b + \mathbf{F}, \quad (38.66\text{f})$$

where we introduced the Bernoulli potential for a hydrostatic and Boussinesq fluid³

$$\mathcal{B} = \frac{|\mathbf{u}|^2}{2} + \frac{p}{\rho_0} + g z. \quad (38.67)$$

Potential vorticity flux

The flux-form potential vorticity conservation statement remains as given by equation (38.52d), and the PV-substance flux is given by equation (38.54). However, we can make use of the gauge invariance of the PV-substance flux to write it in a manner conducive to analyzing steady state conditions. For this purpose, operate with $\nabla b \wedge$ on the velocity equation (38.66f) to have

$$\nabla b \wedge \frac{\partial \mathbf{u}}{\partial t} + \nabla b \wedge (\boldsymbol{\omega}_a^{\text{hy}} \wedge \mathbf{v}) = -\nabla b \wedge \nabla \mathcal{B} + \nabla b \wedge \hat{\mathbf{z}} b + \nabla b \wedge \mathbf{F}, \quad (38.68)$$

and then make use of the identity

$$\nabla b \wedge (\boldsymbol{\omega}_a^{\text{hy}} \wedge \mathbf{v}) = (\nabla b \cdot \mathbf{v}) \boldsymbol{\omega}_a^{\text{hy}} - (\boldsymbol{\omega}_a^{\text{hy}} \cdot \nabla b) \mathbf{v} \quad (38.69)$$

in equation (38.68) to render

$$(\mathbf{v} \cdot \nabla b) \boldsymbol{\omega}_a^{\text{hy}} - (\boldsymbol{\omega}_a^{\text{hy}} \cdot \nabla b) \mathbf{v} = -\nabla b \wedge \frac{\partial \mathbf{u}}{\partial t} - \nabla b \wedge \nabla \mathcal{B} + \nabla b \wedge \hat{\mathbf{z}} b + \nabla b \wedge \mathbf{F}. \quad (38.70)$$

³See Section 22.6.1 for the non-Boussinesq Bernoulli potential.

Now write the flux given by equation (38.54) in the form

$$\mathbf{J}_Q = \mathbf{v} Q - \dot{b} \boldsymbol{\omega}_a^{hy} + \nabla b \wedge \mathbf{F} + \nabla \wedge \mathbf{A} \quad (38.71a)$$

$$= \mathbf{v} (\boldsymbol{\omega}_a^{hy} \cdot \nabla b) - \left[\frac{\partial b}{\partial t} + \mathbf{v} \cdot \nabla b \right] \boldsymbol{\omega}_a^{hy} + \nabla b \wedge \mathbf{F} + \nabla \wedge \mathbf{A} \quad (38.71b)$$

$$= [\mathbf{v} (\boldsymbol{\omega}_a^{hy} \cdot \nabla b) - (\mathbf{v} \cdot \nabla b) \boldsymbol{\omega}_a^{hy}] - \frac{\partial b}{\partial t} \boldsymbol{\omega}_a^{hy} + \nabla b \wedge \mathbf{F} + \nabla \wedge \mathbf{A} \quad (38.71c)$$

$$= \left[\nabla b \wedge \frac{\partial \mathbf{u}}{\partial t} + \nabla b \wedge \nabla \mathcal{B} - \nabla b \wedge \hat{\mathbf{z}} b - \nabla b \wedge \mathbf{F} \right] - \frac{\partial b}{\partial t} \boldsymbol{\omega}_a^{hy} + \nabla b \wedge \mathbf{F} + \nabla \wedge \mathbf{A} \quad (38.71d)$$

$$= \nabla b \wedge \frac{\partial \mathbf{u}}{\partial t} - \frac{\partial b}{\partial t} \boldsymbol{\omega}_a^{hy} + \nabla b \wedge \nabla \mathcal{B} - \nabla b \wedge \hat{\mathbf{z}} b + \nabla \wedge \mathbf{A}. \quad (38.71e)$$

$$= \nabla b \wedge \left[\frac{\partial \mathbf{u}}{\partial t} + \nabla \mathcal{B} \right] - \frac{\partial b}{\partial t} \boldsymbol{\omega}_a^{hy} + \nabla \wedge (\mathbf{A} - \hat{\mathbf{z}} b^2/2). \quad (38.71f)$$

Choosing the gauge function according to

$$\mathbf{A} = \hat{\mathbf{z}} (b^2/2) \quad (38.72)$$

renders the flux vector

$$\mathbf{J}_Q = \nabla b \wedge \left(\frac{\partial \mathbf{u}}{\partial t} + \nabla \mathcal{B} \right) - \frac{\partial b}{\partial t} \boldsymbol{\omega}_a^{hy} \quad (38.73a)$$

$$= \mathbf{v} Q - \dot{b} \boldsymbol{\omega}_a^{hy} + \nabla b \wedge \mathbf{F}. \quad (38.73b)$$

The second equality reintroduced the first form of the PV flux given by equation (38.54); made use of the gauge choice (38.72); and furthermore dropped the non-divergent term

$$\nabla b \wedge b \hat{\mathbf{z}} = (1/2) \nabla b^2 \wedge \hat{\mathbf{z}} \quad \text{with} \quad \nabla \cdot (\nabla b \wedge b \hat{\mathbf{z}}) = 0. \quad (38.74)$$

In the steady state, the PV flux (38.73a) reduces to

$$\mathbf{J}_Q^{ss} = \nabla b \wedge \nabla \mathcal{B}. \quad (38.75)$$

Hence, the steady state potential vorticity flux is aligned with the intersection of surfaces of constant buoyancy and Bernoulli potential

$$\nabla b \cdot \mathbf{J}_Q^{ss} = 0 \quad \text{and} \quad \nabla \mathcal{B} \cdot \mathbf{J}_Q^{ss} = 0. \quad (38.76)$$

This result is the Boussinesq/hydrostatic form of the more general result derived in Section 39.2.3.

38.6 Exercises

EXERCISE 38.1: PV FOR A PERFECT NON-HYDROSTATIC BOUSSINESQ FLUID

Consider a perfect rotating Boussinesq fluid whose governing equations are given by

$$\frac{D\mathbf{v}}{Dt} + f(\hat{\mathbf{z}} \wedge \mathbf{v}) = -\nabla \varphi + b \hat{\mathbf{z}} \quad (38.77)$$

$$\nabla \cdot \mathbf{v} = 0 \quad (38.78)$$

$$\frac{Db}{Dt} = 0 \quad (38.79)$$

$$b = -\frac{g(\rho - \rho_0)}{\rho_0} = g \alpha \theta, \quad (38.80)$$

where θ is the potential temperature and $\alpha > 0$ is a constant thermal expansion coefficient. Further details are provided in Section 24.1.6. Some of this exercise follows the hydrostatic Boussinesq discussion in Section 38.5, though they differ in important places so be careful!

- Derive the equation for the material time evolution of potential vorticity in this fluid system. Show all steps in the derivation.
- The baroclinicity vector appearing in the Boussinesq vorticity equation is $\mathbf{B} = \nabla \wedge \hat{\mathbf{z}} b = \nabla b \wedge \hat{\mathbf{z}}$ (see Section 36.8.1 to check the derivation in part (a) above). Show that this vector results upon making the Boussinesq approximation to the non-Boussinesq expression $\mathbf{B} = (\nabla \rho \wedge \nabla p)/\rho^2$. Hint: drop the $\delta\rho$ and δp product given that it is a higher order quantity.
- Show that the vertical portion of Q^{bouss} can be written

$$Q_{\text{vert}}^{\text{bouss}} = (\zeta + f) N^2 \quad (38.81)$$

where $\zeta = \partial_x v - \partial_y u$ is the vertical component to the relative vorticity and $N^2 = \partial b / \partial z$ is the squared buoyancy frequency (Section 25.5.4). Hint: this is not tough to show; there is no trick here.

- If flow maintains hydrostatic and geostrophic balance, show that the horizontal portion of Q^{bouss} can be written

$$Q_{\text{horz}}^{\text{bouss}} = \boldsymbol{\omega} \cdot \nabla_z b \approx -f^{-1} |\nabla_z b|^2. \quad (38.82)$$

Hint: recall that for hydrostatic and geostrophic flow, the vertical velocity is much smaller than horizontal.

EXERCISE 38.2: PV FOR DIABATIC AND FRICTIONAL NON-HYDROSTATIC BOUSSINESQ FLOW

Reconsider Exercise 38.1 in the presence of irreversible friction and buoyancy sources so that the governing equations are

$$\frac{D\mathbf{v}}{Dt} + f(\hat{\mathbf{z}} \wedge \mathbf{v}) = -\nabla \varphi + b \hat{\mathbf{z}} + \mathbf{F} \quad (38.83)$$

$$\nabla \cdot \mathbf{v} = 0 \quad (38.84)$$

$$\frac{Db}{Dt} = \dot{b} \quad (38.85)$$

$$b = -\frac{g(\rho - \rho_0)}{\rho_0} = g \alpha \theta, \quad (38.86)$$

where $\dot{\theta}$ is a diabatic heating source/sink, α is a constant thermal expansion coefficient, and \mathbf{F} is a non-conservative acceleration such as from friction or boundary stresses.

- Derive the equation for the material time evolution of potential vorticity in this fluid system, including the irreversible contributions from friction and heating.
- Derive an equation for the potential vorticity time tendency (i.e., Eulerian time derivative), written in the form

$$\frac{\partial Q}{\partial t} = -\nabla \cdot \mathbf{J}_Q. \quad (38.87)$$

What is the PV flux \mathbf{J}_Q ? Note that your answer is unique up to the curl of an arbitrary vector (gauge symmetry). Also note that for a Boussinesq flow we drop the constant reference density in the definition of \mathbf{J}_Q .

- (c) A common diabatic process is written in the form of a damping source

$$\dot{b} = -\mu(b - b^*), \quad (38.88)$$

where μ is a constant Newtonian damping coefficient (dimensions of inverse time), and b^* is a specified buoyancy profile. This form of a buoyancy source acts to damp the buoyancy towards a specified profile b^* . Show that Newtonian damping of buoyancy corresponds to potential vorticity damping towards $Q^* = \boldsymbol{\omega}_a \cdot \nabla b^*$.

- (d) A form for the friction operator is given by Rayleigh drag

$$\mathbf{F} = -\gamma \mathbf{v}, \quad (38.89)$$

with γ a constant Rayleigh damping coefficient with dimension of inverse time. Show that Rayleigh drag in the momentum equation, which acts to damp velocity towards zero, corresponds to a damping of potential vorticity towards its planetary geostrophic form, $Q^{pg} = f N^2$, where $N^2 = \partial b / \partial z$ is the squared buoyancy frequency.

- (e) Discuss the balance needed between forcing terms in \mathbf{J}_Q to arrive at a steady state (i.e., zero Eulerian time tendency). Continue to assume the friction is in the form of Rayleigh drag and heating is in the form of Newtonian damping.



Potential vorticity budgets

The chapter examines a particularly striking property of the flux of potential vorticity substance. Namely, there are forms for this flux (differing by gauge choices) that have identically zero flux of PV-substance crossing an isentrope. This *impermeability theorem* holds even when there are mass and entropy fluxes crossing the isentrope. Isentropes are thus semi-permeable membranes in the sense that they are open to the transport of matter and thermodynamic properties but closed to the transport of PV-substance. This kinematic result offers further understanding for why, as developed in Section 39.3, ρQ remains constant within an isentrope unless the isentrope intersects a boundary. These notions are fundamental to how we perform a budget of potential vorticity substance over a region.

READER'S GUIDE FOR THIS CHAPTER

This chapter builds from the potential vorticity mechanics introduced in Chapter 38. It is an essential read for those interested in potential vorticity theory and potential vorticity budgets.

39.1	Variations on the impermeability theorem	898
39.1.1	Impermeability for the Haynes-McIntyre PV flux	898
39.1.2	A kinematic derivation of impermeability	899
39.1.3	Impermeability for all components of absolute vorticity	900
39.1.4	Comments	901
39.2	Impermeability theorem for a realistic seawater EOS	901
39.2.1	Ocean PV in terms of potential density	901
39.2.2	A modified PV-substance flux	902
39.2.3	Integral constraints for steady state	903
39.2.4	Further study	904
39.3	Integrated PV	904
39.3.1	Region bounded by a buoyancy surface	904
39.3.2	Region bounded by two buoyancy surfaces	905
39.3.3	Region bounded by land and a buoyancy surface	906
39.3.4	An ocean layer that outcrops	907
39.4	Layer integrated PV budget	907
39.4.1	Layer integrated budget	908
39.4.2	Impermeability across interior layer interfaces	908
39.4.3	PV flux at the land-sea boundary	909
39.4.4	PV flux at the air-sea boundary	910
39.4.5	Thought experiments	911
39.4.6	Is there a preferred form of the PV-substance flux?	911

39.1 Variations on the impermeability theorem

In this section we derive the impermeability theorem by making use of a variety of forms for the flux vector of potential vorticity substance. The forms differ by a gauge function, so that the divergence of all the flux forms are the same.

39.1.1 Impermeability for the Haynes-McIntyre PV flux

Recall the Eulerian flux-form evolution equation for PV-substance (38.40)

$$\frac{\partial(\rho Q)}{\partial t} + \nabla \cdot \mathbf{J}_Q = 0 \quad \text{with} \quad \mathbf{J}_Q = \rho \mathbf{v} Q - \dot{\theta} \boldsymbol{\omega}_a + \nabla \theta \wedge \mathbf{F}, \quad (39.1)$$

with the PV-substance flux vector, \mathbf{J}_Q , here given in the form examined in [Haynes and McIntyre \(1987\)](#). Now decompose the velocity into two components, one oriented parallel to isentropes and one oriented perpendicular

$$\mathbf{v}_{\parallel} = \mathbf{v} - \hat{\mathbf{n}} (\mathbf{v} \cdot \hat{\mathbf{n}}) \quad \text{and} \quad \mathbf{v}_{\perp} = -\frac{\hat{\mathbf{n}} \partial \theta / \partial t}{|\nabla \theta|} \quad \Rightarrow \mathbf{v} = \mathbf{v}_{\parallel} + \mathbf{v}_{\perp} + \frac{\hat{\mathbf{n}} \dot{\theta}}{|\nabla \theta|} \quad (39.2)$$

where $\hat{\mathbf{n}} = \nabla \theta / |\nabla \theta|$ is the normal vector on an isentrope. The velocity \mathbf{v}_{\perp} , by construction, satisfies

$$\frac{\partial \theta}{\partial t} + \mathbf{v}_{\perp} \cdot \nabla \theta = 0. \quad (39.3)$$

Hence, according to the kinematics detailed in Section 15.4.2,

$$\mathbf{v}_{\perp} \cdot \hat{\mathbf{n}} = \mathbf{v}_{\theta} \cdot \hat{\mathbf{n}}, \quad (39.4)$$

where \mathbf{v}_{θ} is the velocity of a point fixed on the θ isosurface. We make use of this key identity below.

With the velocity decomposition (39.2), the PV-substance flux vector takes the form

$$\mathbf{J}_Q = \rho \mathbf{v} Q - \dot{\theta} \boldsymbol{\omega}_a + \nabla \theta \wedge \mathbf{F} \quad (39.5a)$$

$$= \left[\mathbf{v}_{\parallel} + \mathbf{v}_{\perp} + \frac{\dot{\theta} \nabla \theta}{|\nabla \theta|^2} \right] \rho Q - \dot{\theta} \boldsymbol{\omega}_a + \nabla \theta \wedge \mathbf{F} \quad (39.5b)$$

$$= (\mathbf{v}_{\parallel} + \mathbf{v}_{\perp}) \rho Q - \dot{\theta} [\boldsymbol{\omega}_a - (\boldsymbol{\omega}_a \cdot \hat{\mathbf{n}}) \hat{\mathbf{n}}] + \nabla \theta \wedge \mathbf{F} \quad (39.5c)$$

$$= (\mathbf{v}_{\parallel} + \mathbf{v}_{\perp}) \rho Q - \dot{\theta} (\boldsymbol{\omega}_a)_{\parallel} + \nabla \theta \wedge \mathbf{F} \quad (39.5d)$$

$$= \mathbf{v}_{\perp} \rho Q + \left[\rho Q \mathbf{v}_{\parallel} - \dot{\theta} (\boldsymbol{\omega}_a)_{\parallel} \right] + \nabla \theta \wedge \mathbf{F} \quad (39.5e)$$

$$\equiv \mathbf{J}_{\perp} + \mathbf{J}_{\parallel}, \quad (39.5f)$$

where

$$(\boldsymbol{\omega}_a)_{\parallel} = \boldsymbol{\omega}_a - (\boldsymbol{\omega}_a \cdot \hat{\mathbf{n}}) \hat{\mathbf{n}} = \boldsymbol{\omega}_a - \left[\frac{\boldsymbol{\omega}_a \cdot \nabla \theta}{|\nabla \theta|^2} \right] \nabla \theta = \boldsymbol{\omega}_a - \frac{\rho Q}{|\nabla \theta|} \hat{\mathbf{n}}. \quad (39.6)$$

The above results motivate us to write the PV-substance budget equation (39.1) in the alternative form

$$\frac{\partial(\rho Q)}{\partial t} + \nabla \cdot (\mathbf{v}_Q \rho Q) = 0, \quad (39.7)$$

where

$$\mathbf{v}_Q = \frac{\mathbf{J}_Q}{\rho Q} = \mathbf{v}_{\perp} + \mathbf{v}_{\parallel} - \frac{\dot{\theta} (\boldsymbol{\omega}_a)_{\parallel} + \mathbf{F} \wedge \nabla \theta}{\rho Q} \quad (39.8)$$

is an effective velocity that advects the PV-substance through the fluid. A direct calculation shows that \mathbf{v}_Q satisfies the following key property

$$\mathbf{v}_Q \cdot \hat{\mathbf{n}} = \mathbf{v}_\perp \cdot \hat{\mathbf{n}} = \mathbf{v}_\theta \cdot \hat{\mathbf{n}}, \quad (39.9)$$

where the final equality made use of the identity (39.4). As a result, the velocity \mathbf{v}_Q , which advects PV-substance, has a normal component that is identical to that of the velocity of a point fixed on the isentrope

$$\frac{\partial \theta}{\partial t} + \mathbf{v}_Q \cdot \nabla \theta = 0. \quad (39.10)$$

We depict this result in Figure 39.1, whereby the PV-substance flux never crosses the isentrope, even as the isentrope moves and even in the presence of irreversible processes that allow for matter and heat to cross the isentrope. This result holds since the θ -isosurface moves in a way to precisely track the PV-substance flux. In the presence of non-conservative processes, isentropes are permeable to matter and heat but impermeable to PV-substance. This is a rather remarkable kinematic result that has important implications for budgets of PV-substance, as shown by the discussion in the remainder of this chapter as well as when discussing ocean potential vorticity in Chapter 46.

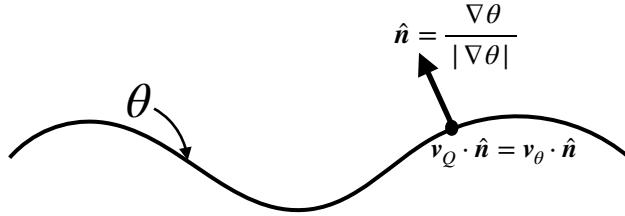


FIGURE 39.1: The flux, \mathbf{J}_Q , of PV-substance does not penetrate a surface of constant potential temperature (isentrope). This kinematic result follows since the effective velocity of PV-substance, $\mathbf{v}_Q = \mathbf{J}_Q^{\text{HM}}/(\rho Q)$, has the same normal component as a point fixed on a potential temperature iso-surface, $\mathbf{v}_Q \cdot \hat{\mathbf{n}} = \mathbf{v}_\theta \cdot \hat{\mathbf{n}}$. Consequently, the potential temperature surface moves in a manner so that no flux of PV-substance crosses the surface, even in the presence of irreversible processes. This result is known as the impermeability theorem since potential temperature surfaces are impermeable to the flux of PV-substance.

39.1.2 A kinematic derivation of impermeability

We here offer a second derivation of the impermeability theorem that emphasizes its kinematic origins. That is, we make no use of the vorticity equation or the potential temperature equation. Instead, we merely make use of the definition of potential vorticity and, critically, the non-divergence of absolute vorticity.

The key identity we need is the following

$$\rho Q = \boldsymbol{\omega}_a \cdot \nabla \theta = \nabla \cdot (\boldsymbol{\omega}_a \theta), \quad (39.11)$$

so that ρQ is a pure divergence. Taking the Eulerian time derivative then leads to

$$\frac{\partial(\rho Q)}{\partial t} = -\nabla \cdot \mathbf{J}_Q^{\text{kin}}, \quad (39.12)$$

where

$$\mathbf{J}_Q^{\text{kin}} = -\frac{\partial(\boldsymbol{\omega}_a \theta)}{\partial t} \quad (39.13)$$

is the kinematic form of the PV-substance flux. By construction, this flux vanishes in the steady state

$$\mathbf{J}_Q^{\text{kin}} = 0 \quad \text{in steady state,} \quad (39.14)$$

which contrasts to the Haynes-McIntyre form given by equation (39.1). We return to this point in Section 39.2 where we introduce a strategically chosen gauge term (a total curl) to render a nonzero flux in the steady state.

Introducing the velocity seen from an inertial reference frame (also called the absolute velocity; Section 10.9.1)

$$\mathbf{v}_a = \mathbf{v} + \boldsymbol{\Omega} \wedge \mathbf{x} \quad (39.15)$$

leads to

$$-\frac{\partial \boldsymbol{\omega}_a}{\partial t} \cdot \theta = -\frac{\partial(\nabla \wedge \mathbf{v}_a)}{\partial t} \cdot \theta = -\left[\nabla \wedge \frac{\partial \mathbf{v}_a}{\partial t} \right] \theta = -\nabla \wedge \left[\frac{\partial \mathbf{v}_a}{\partial t} \theta \right] + \nabla \theta \wedge \frac{\partial \mathbf{v}_a}{\partial t}. \quad (39.16)$$

Dropping the total curl yields the modified kinematic form for the PV-substance flux

$$\tilde{\mathbf{J}}_Q^{\text{kin}} = -\frac{\partial \mathbf{v}_a}{\partial t} \wedge \nabla \theta - \boldsymbol{\omega}_a \frac{\partial \theta}{\partial t}. \quad (39.17)$$

It follows that

$$\tilde{\mathbf{v}}_Q^{\text{kin}} \cdot \hat{\mathbf{n}} = \frac{\tilde{\mathbf{J}}_Q^{\text{kin}} \cdot \hat{\mathbf{n}}}{\rho Q} = -\frac{\boldsymbol{\omega}_a \cdot \nabla \theta}{\rho Q} \frac{\partial \theta}{\partial t} \frac{1}{|\nabla \theta|} = -\frac{\partial \theta}{\partial t} \frac{1}{|\nabla \theta|} = \mathbf{v}_\theta \cdot \hat{\mathbf{n}}, \quad (39.18)$$

which then leads to the impermeability theorem for $\tilde{\mathbf{J}}_Q^{\text{kin}}$. We again emphasize that there has been no use of the dynamical equations for vorticity or potential temperature. Instead, this expression of impermeability only used the definition of potential vorticity along with the central property $\nabla \cdot \boldsymbol{\omega}_a = 0$.

39.1.3 Impermeability for all components of absolute vorticity

The above kinematic result suggests that the impermeability theorem is both remarkable and trivial. We support that dual characterization by here showing that all components of absolute vorticity satisfy an impermeability theorem with respect to the direction along which the components are defined. For that purpose, consider an arbitrary smooth scalar field λ and use it to project out that component of absolute vorticity normal to λ isosurfaces

$$\rho Q_\lambda \equiv \boldsymbol{\omega}_a \cdot \nabla \lambda = \nabla \cdot (\boldsymbol{\omega}_a \lambda). \quad (39.19)$$

Following the derivation given in Section 39.1.2 leads to

$$\frac{\partial(\rho Q_\lambda)}{\partial t} = \nabla \cdot \left[\frac{\partial \mathbf{v}_a}{\partial t} \wedge \nabla \lambda + \boldsymbol{\omega}_a \frac{\partial \lambda}{\partial t} \right] \equiv -\nabla \cdot \mathbf{J}_{Q_\lambda}. \quad (39.20)$$

It follows that

$$\frac{\mathbf{J}_{Q_\lambda} \cdot \hat{\mathbf{n}}_\lambda}{\rho Q_\lambda} = -\frac{\boldsymbol{\omega}_a \cdot \nabla \lambda}{\rho Q_\lambda} \frac{\partial \lambda}{\partial t} \frac{1}{|\nabla \lambda|} = -\frac{\partial \lambda}{\partial t} \frac{1}{|\nabla \lambda|} = \mathbf{v}_\lambda \cdot \hat{\mathbf{n}}_\lambda, \quad (39.21)$$

where

$$\hat{\mathbf{n}}_\lambda = \frac{\nabla \lambda}{|\nabla \lambda|} \quad \text{and} \quad \frac{\partial \lambda}{\partial t} + \mathbf{v}_\lambda \cdot \nabla \lambda = 0. \quad (39.22)$$

We are thus led to an impermeability theorem for the λ -surface. Hence, *any* scalar field used to project out a component of the absolute vorticity has its iso-surfaces impenetrable to the flux of the corresponding component of absolute vorticity. This result trivializes the impermeability theorem from a mathematical perspective, with Section 5 of [Haynes and McIntyre \(1987\)](#) offering further mathematical insights into the inevitability of the result based on the structure of the vorticity equation. Even so, it does not reduce the importance of the entropic PV impermeability theorem for studies of stratified flows, since entropic PV is a special component of the absolute vorticity that has direct connection to dynamics and thermodynamics.

39.1.4 Comments

The impermeability theorem was introduced by [Haynes and McIntyre \(1987\)](#), with their paper met by some confusion that prompted [Haynes and McIntyre \(1990\)](#). The impermeability theorem emphasizes the importance of boundary processes when studying how potential vorticity substance changes within a domain bounded by isentropes. For example, the emphasis on boundary processes is a focus of oceanographic studies of submesoscale instabilities such as [Thomas et al. \(2008\)](#).

Besides exposing the purely kinematic aspects of the impermeability theorem, the presentation in this section reveals that there are multiple PV flux vectors that satisfy the impermeability theorem, with the vectors differing by a gauge transformation. Which vector is preferred depends on the application, with [Bretherton and Schär \(1993\)](#), [Davies-Jones \(2003a\)](#), and [Marshall et al. \(2001\)](#) proposing criteria favoring one form over another. We pursue an example of such considerations in Section 39.2.

39.2 Impermeability theorem for a realistic seawater EOS

As seen from Section 39.1.3, impermeability holds for any component of vorticity and the corresponding scalar isosurface. In contrast, material invariance of potential vorticity requires a materially conserved scalar to annihilate the baroclinicity vector (e.g., Section 38.1.4). As shown in Section 38.2, there is no such materially invariant scalar for an ocean with a realistic nonlinear equation of state (EOS). Hence, there is no materially invariant potential vorticity for the ocean even in the absence of irreversible processes. Nevertheless, one can define an ocean potential vorticity according to any scalar field, such as potential density, and still make use of the impermeability theorem when performing a PV budget.

39.2.1 Ocean PV in terms of potential density

The distinction between impermeability and material invariance was emphasized by [Marshall et al. \(2001\)](#), and we here summarize features from that paper. For that purpose, introduce an ocean potential vorticity field according to

$$Q^{\text{ocn}} = \frac{\nabla b \cdot \omega_a}{\rho}, \quad (39.23)$$

where the globally defined buoyancy field, b , is approximated by a chosen potential density. As shown in Section 38.2, a globally defined buoyancy does not annihilate the baroclinicity vector for a realistic seawater equation of state

$$\mathbf{B} \cdot \nabla b = [-\nabla(1/\rho) \wedge \nabla p] \cdot \nabla b \neq 0. \quad (39.24)$$

Consequently, $DQ^{\text{ocn}}/Dt \neq 0$ even in the absence of irreversible processes. Nonetheless, the Eulerian budget for PV-substance satisfies

$$\frac{\partial(\rho Q^{\text{ocn}})}{\partial t} = -\nabla \cdot \tilde{\mathbf{J}}_Q^{\text{ocn}}, \quad (39.25)$$

and $\tilde{\mathbf{J}}_Q^{\text{ocn}}$ satisfies the impermeability theorem for b -surfaces. A flux-form budget equation greatly facilitates the study of budgets for PV-substance even within an ocean with a realistic equation of state. Derivation of the flux-form equation (39.25) follows from the discussion in Section 39.1.2, where we know that the kinematic flux

$$\tilde{\mathbf{J}}_Q^{\text{ocn}} = -\frac{\partial \mathbf{v}_a}{\partial t} \wedge \nabla b - \boldsymbol{\omega}_a \frac{\partial b}{\partial t} = -\frac{\partial \mathbf{v}}{\partial t} \wedge \nabla b - \boldsymbol{\omega}_a \frac{\partial b}{\partial t} \quad (39.26)$$

satisfies the impermeability theorem for b -surfaces and whose convergence drives the time tendency for the PV-substance. Note that the second equality in equation (39.26) follows since

$$\frac{\partial \mathbf{v}_a}{\partial t} = \frac{\partial(\mathbf{v} + \boldsymbol{\Omega} \wedge \mathbf{x})}{\partial t} = \frac{\partial \mathbf{v}}{\partial t}, \quad (39.27)$$

given that the Eulerian time derivative is computed at a fixed position, \mathbf{x} , and the planetary rotation is assumed constant.

39.2.2 A modified PV-substance flux

The kinematic flux (39.26) vanishes in the steady state. We here motivate a gauge transformed flux that leads to the same flux divergence yet that renders a more physically interesting steady state flux. For this purpose, make use of the vector-invariant velocity equation (equation (36.21))

$$\frac{\partial \mathbf{v}}{\partial t} + \boldsymbol{\omega}_a \wedge \mathbf{v} = -\rho^{-1} \nabla p - \nabla m + \mathbf{F}, \quad (39.28)$$

where

$$m = \frac{1}{2} \mathbf{v} \cdot \mathbf{v} + \Phi \quad (39.29)$$

is the mechanical energy per mass of a fluid element. Bringing the pressure term inside of the gradient operator leads to

$$\frac{\partial \mathbf{v}}{\partial t} + \boldsymbol{\omega}_a \wedge \mathbf{v} = p \nabla(1/\rho) - \nabla(m + p/\rho) + \mathbf{F}. \quad (39.30)$$

In the following, we write

$$B = m + p/\rho, \quad (39.31)$$

which is a Bernoulli function.¹ The vector-invariant velocity equation (39.30) leads to the cross-product

$$\frac{\partial \mathbf{v}}{\partial t} \wedge \nabla b = -(\boldsymbol{\omega}_a \wedge \mathbf{v}) \wedge \nabla b + [p \nabla(1/\rho) - \nabla(m + p/\rho) + \mathbf{F}] \wedge \nabla b \quad (39.32a)$$

$$= -(\nabla b \cdot \boldsymbol{\omega}_a) \mathbf{v} + (\nabla b \cdot \mathbf{v}) \boldsymbol{\omega}_a + [p \nabla(1/\rho) - \nabla(m + p/\rho) + \mathbf{F}] \wedge \nabla b \quad (39.32b)$$

$$= -\mathbf{v} \rho Q^{\text{ocn}} + (\dot{b} - \partial_t b) \boldsymbol{\omega}_a + [p \nabla(1/\rho) - \nabla(m + p/\rho) + \mathbf{F}] \wedge \nabla b. \quad (39.32c)$$

¹The Bernoulli potential, \mathcal{B} , arises from an analysis of the total energy budget as in equation (22.78), where we see that the Bernoulli potential in a compressible (non-Boussinesq) fluid, $\mathcal{B} = m + p/\rho + \mathcal{I}$, also includes the internal energy per mass, \mathcal{I} . However, the internal energy is missing from equation (39.31), thus motivating our use of the terminology “a Bernoulli function” rather than “the Bernoulli potential”.

Use of this result leads to the flux (39.26)

$$\tilde{\mathbf{J}}_Q^{\text{ocn}} = -\frac{\partial \mathbf{v}}{\partial t} \wedge \nabla b - \boldsymbol{\omega}_a \frac{\partial b}{\partial t} \quad (39.33a)$$

$$= \mathbf{v} \rho Q^{\text{ocn}} - \dot{b} \boldsymbol{\omega}_a - \mathbf{F} \wedge \nabla b + [\nabla(m + p/\rho) - p \nabla(1/\rho)] \wedge \nabla b \quad (39.33b)$$

$$= \mathbf{J}_Q + [\nabla(m + p/\rho) - p \nabla(1/\rho)] \wedge \nabla b, \quad (39.33c)$$

where \mathbf{J}_Q is the Haynes-McIntyre form of the PV-substance flux given by equation (39.1). The term

$$\nabla(m + p/\rho) \wedge \nabla b = \nabla \wedge [(m + p/\rho) \nabla b] \quad (39.34)$$

is a total curl and as such it can be moved around without altering the evolution of PV-substance; i.e., it is a “do-nothing” flux. Furthermore, since it is parallel to buoyancy isosurfaces it does not alter the impermeability properties of the PV-substance flux. [Marshall et al. \(2001\)](#) focused attention on the flux

$$\mathbf{J}_Q^{\text{marshall}} = \tilde{\mathbf{J}}_Q^{\text{ocn}} - \nabla(m + p/\rho) \wedge \nabla b = -\left[\frac{\partial \mathbf{v}}{\partial t} + \nabla(m + p/\rho)\right] \wedge \nabla b - \boldsymbol{\omega}_a \frac{\partial b}{\partial t}. \quad (39.35)$$

Diagnostically desirable features of $\mathbf{J}_Q^{\text{marshall}}$ include the following

- $\nabla b \cdot \mathbf{J}_Q^{\text{marshall}} / (\rho Q) = \partial_t b$, thus satisfying the impermeability theorem.
- $\mathbf{J}_Q^{\text{marshall}}$ has no explicit reference to irreversible processes. Consequently, it can in some cases be simpler to diagnose than the Haynes-McIntyre flux, \mathbf{J}_Q .
- In a steady state, the flux is given by

$$\mathbf{J}_Q^{\text{marshall}} = \nabla b \wedge \nabla(m + p/\rho) = \nabla \wedge [b \nabla(m + p/\rho)]. \quad (39.36)$$

Consequently, $m + p/\rho$ provides a streamfunction for the steady state flux on buoyancy surfaces. As emphasized by [Schär \(1993\)](#), this result holds even when there are irreversible processes, thus providing useful diagnostics even in the presence of dissipation.

39.2.3 Integral constraints for steady state

The steady state PV-substance flux in the form (39.36) can be used to develop some integral constraints on the steady flow. For this purpose consider the steady form of $\mathbf{J}_Q^{\text{marshall}}$ and integrate over an arbitrary simply connected area making use of Stokes' theorem

$$\int_S \nabla \wedge [b \nabla B] \cdot \hat{\mathbf{n}} dS = \oint_{\partial S} b \nabla B \cdot d\mathbf{r} = \oint_{\partial S} b dB = - \oint_{\partial S} B db. \quad (39.37a)$$

The first equality made use of Stokes' theorem; the second make use of the identity for exact differentials

$$\nabla B \cdot d\mathbf{r} = dB; \quad (39.38)$$

and the final equality made use of

$$b dB = d(b B) - B db \quad (39.39)$$

and noted that

$$\oint_{\partial S} d(B b) = 0. \quad (39.40)$$

If we can find a closed contour where either B is a constant ($\mathrm{d}B = 0$), or the buoyancy is a constant ($\mathrm{d}b = 0$), then we have the steady state constraint

$$\int_S \mathbf{J}_Q^{\text{marshall}} \cdot \hat{\mathbf{n}} \, \mathrm{d}S = 0 \quad \text{area enclosed by contour with } m + p/\rho \text{ constant or } b \text{ constant.} \quad (39.41)$$

In regions where there are such closed contours, this constraint offers useful insight into the steady state balances. [Marshall \(2000\)](#) and [Polton and Marshall \(2007\)](#) make particular use of closed B contours on constant depth surfaces (so that $\hat{\mathbf{n}} = \hat{\mathbf{z}}$) in a Boussinesq fluid.

39.2.4 Further study

[Marshall et al. \(2001\)](#) builds from the generalized Bernoulli theorem of [Schär \(1993\)](#) and [Bretherton and Schär \(1993\)](#). We also consider these topics for a Boussinesq hydrostatic fluid in Section 38.5.4.

39.3 Integrated PV

In this section we derive some kinematic properties of integrated potential vorticity, with these properties relying solely on the definition of PV. We write potential vorticity using a global buoyancy field, b , as in our discussion of ocean potential vorticity in Section 39.2

$$Q = \frac{\boldsymbol{\omega}_a \cdot \nabla b}{\rho} = \frac{\nabla \cdot (\boldsymbol{\omega}_a b)}{\rho}. \quad (39.42)$$

Notably, the properties exhibited in this section hold even when there is no materially invariant potential vorticity, since they rely solely on the non-divergent nature of the absolute vorticity.

In this section we consider Q to be an intensive fluid property measuring the amount of PV-substance per unit mass (i.e., the concentration of PV-substance), and correspondingly with ρQ the amount of PV-substance per volume.² With this interpretation, the amount of PV substance within an arbitrary finite region is determined by the volume integral of ρQ

$$\mathcal{I} = \int_{\mathcal{R}} Q \rho \, \mathrm{d}V = \int_{\mathcal{R}} \nabla \cdot (\boldsymbol{\omega}_a b) \, \mathrm{d}V = \oint_{\partial\mathcal{R}} \boldsymbol{\omega}_a b \cdot \hat{\mathbf{n}} \, \mathrm{d}S, \quad (39.43)$$

where the final equality used Gauss's divergence theorem. Hence, the volume integrated PV substance in a region is determined solely by values on the region boundary. This property is strikingly distinct from material tracers. We next explore some implications of this result.

39.3.1 Region bounded by a buoyancy surface

Consider a volume of fluid bounded by a single buoyancy surface as shown in the bubble-like region in the left panel of Figure 39.2. Since the outer boundary of the region is set by a constant b -surface, we can pull b outside of the surface integral in equation (39.43) so that

$$\mathcal{I} = \oint_{\partial\mathcal{R}} \boldsymbol{\omega}_a b \cdot \hat{\mathbf{n}} \, \mathrm{d}S = b_2 \oint_{\partial\mathcal{R}} \boldsymbol{\omega}_a \cdot \hat{\mathbf{n}} \, \mathrm{d}S. \quad (39.44)$$

We can now use the divergence theorem to return to the volume integral, only now with b outside of the integral

$$\mathcal{I} = b_2 \int_{\mathcal{R}} \nabla \cdot \boldsymbol{\omega}_a \, \mathrm{d}V = 0, \quad (39.45)$$

²Recall our discussion of extensive and intensive fluid properties in Section 16.3.1.

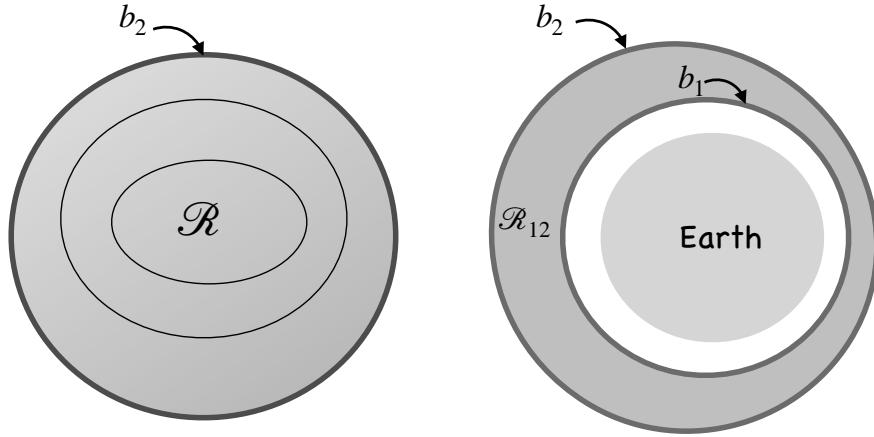


FIGURE 39.2: Integrating PV over regions bounded by potential temperature surfaces that do not intersect the ground. The left panel considers a single buoyancy surface, $b = b_2$, bounding the bubble-like fluid region \mathcal{R} . Notably, the region inside the bubble generally has nontrivial buoyancy distribution, as depicted here by exhibiting a couple of contours inside the b_2 -bounded domain. The only assumption is that it is wholly contained inside the $b = b_2$ contour. The right panel considers a buoyancy layer or shell, \mathcal{R}_{12} , bounded by two buoyancy isosurfaces, $b_1 < b_2$, surrounding the earth, with neither surface intersecting the ground. There is identically zero domain integrated PV in both \mathcal{R} and \mathcal{R}_{12} . Hence, if there is any nontrivial distribution of PV in either domain, there must be as much integrated positive values as there are negative.

where $\nabla \cdot \boldsymbol{\omega}_a = 0$ led to the final equality. Equivalently, we can use Stokes' theorem to convert the closed area integral, $\oint_{\partial\mathcal{R}} \boldsymbol{\omega}_a \cdot \hat{\mathbf{n}} d\mathcal{S}$, to a line integral around the boundary. However, there is no boundary for the closed area since it covers the sphere, thus again showing that $\mathcal{G} = 0$ (see Section 2.7.5).

The remarkable identity (39.45) follows from kinematics alone since no dynamical information was used in its derivation. It says that there is zero integrated PV-substance contained within any region bounded solely by a single buoyancy surface. Hence, within the domain there is just as much positive PV-substance as there is negative PV-substance. So if PV changes locally within the domain, then somewhere else it must experience an oppositely signed change so to leave a net integrated value of zero. The result holds whether there are reversible or irreversible processes acting on the buoyancy surface. Our only assumption is that the domain is fully enclosed by a b -surface.

39.3.2 Region bounded by two buoyancy surfaces

The identity (39.45) has a corollary, in which we consider a region bounded by two b -surfaces such as \mathcal{R}_{12} shown in the right panel of Figure 39.2. The above arguments hold for that region as well, since we can decompose the surface integral into two integrals separately over b_1 and b_2

$$\mathcal{G} = \int_{\mathcal{R}_{12}} \nabla \cdot (\boldsymbol{\omega}_a b) dV = \int_{\mathcal{R}_2} \nabla \cdot (\boldsymbol{\omega}_a b) dV - \int_{\mathcal{R}_1} \nabla \cdot (\boldsymbol{\omega}_a b) dV, \quad (39.46)$$

where the domain \mathcal{R}_1 extends from the ground up to b_1 and \mathcal{R}_2 extends from the ground up to b_2 . Integration over the region below b_1 cancels through the subtraction. Indeed, the region below b_1 could be anything without changing the result. So let that region be filled with fluid throughout (i.e., ignore the earth) to allow us to extend both integrals throughout the spherical region just like in the single isentrope domain \mathcal{R} in Figure 39.2. Invoking the single isentrope result we see that

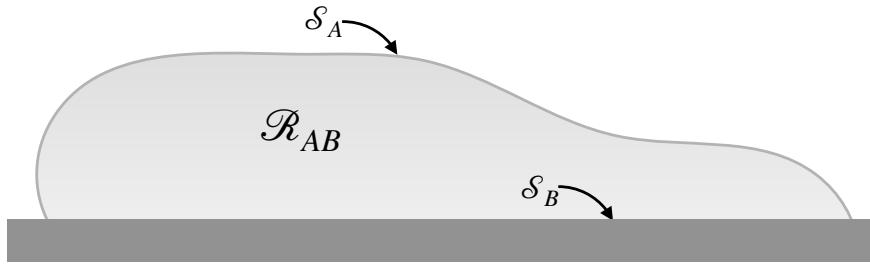


FIGURE 39.3: A fluid region, \mathcal{R}_{AB} , bounded by two surfaces, \mathcal{S}_A and \mathcal{S}_B . The upper surface \mathcal{S}_A is defined by a buoyancy isosurface, $b = b_A$, with this surface intersecting the ground. The lower surface, \mathcal{S}_B , is along the ground and has a buoyancy that is a function of space and time, $b_B(\mathbf{x}, t)$.

both integrals separately vanish. We are thus led to a vanishing integral for the layer

$$\mathcal{I} = \int_{\mathcal{R}_{12}} \nabla \cdot (\boldsymbol{\omega}_a b) dV = 0. \quad (39.47)$$

Again, the key assumption is that no buoyancy surface intersects land, in which case we are able to ignore the presence of land altogether and thus make use of the single isentrope result.

39.3.3 Region bounded by land and a buoyancy surface

Now consider a fluid domain bounded by a buoyancy surface that intersects the ground (atmospheric example) or ocean surface (ocean example), as shown in Figure 39.3. The integrated PV is given by

$$\mathcal{I} = \int_{\mathcal{R}_{AB}} \nabla \cdot (\boldsymbol{\omega}_a b) dV \quad (39.48a)$$

$$= \int_{\mathcal{S}_A} \boldsymbol{\omega}_a b \cdot \hat{\mathbf{n}} d\mathcal{S} + \int_{\mathcal{S}_B} \boldsymbol{\omega}_a b \cdot \hat{\mathbf{n}} d\mathcal{S} \quad (39.48b)$$

$$= b_A \int_{\mathcal{S}_A} \boldsymbol{\omega}_a \cdot \hat{\mathbf{n}} d\mathcal{S} + \int_{\mathcal{S}_B} b_B \boldsymbol{\omega}_a \cdot \hat{\mathbf{n}} d\mathcal{S} \quad (39.48c)$$

$$= b_A \left[\int_{\mathcal{S}_A} \boldsymbol{\omega}_a \cdot \hat{\mathbf{n}} d\mathcal{S} + \int_{\mathcal{S}_B} \boldsymbol{\omega}_a \cdot \hat{\mathbf{n}} d\mathcal{S} \right] + \int_{\mathcal{S}_B} (b_B - b_A) \boldsymbol{\omega}_a \cdot \hat{\mathbf{n}} d\mathcal{S} \quad (39.48d)$$

$$= b_A \int_{\mathcal{R}_{AB}} \nabla \cdot \boldsymbol{\omega}_a dV + \int_{\mathcal{S}_B} (b_B - b_A) \boldsymbol{\omega}_a \cdot \hat{\mathbf{n}} d\mathcal{S} \quad (39.48e)$$

$$= \int_{\mathcal{S}_B} (b_B - b_A) \boldsymbol{\omega}_a \cdot \hat{\mathbf{n}} d\mathcal{S}, \quad (39.48f)$$

where we made use $\nabla \cdot \boldsymbol{\omega}_a = 0$ to reach the final equality. Consequently, the PV substance in a region enclosed by buoyancy surfaces can change only when the buoyancy surfaces intersect a boundary that has a non-constant buoyancy. As both the ground and the ocean surface have buoyancy gradients, they contribute to changes in the PV substance within the region they bound.

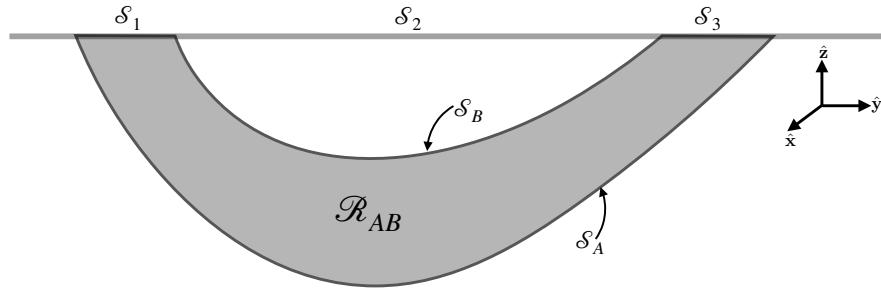


FIGURE 39.4: This figure depicts a buoyancy layer in the ocean that outcrops at both ends of the layer. The bounds the regions are given by the following surfaces. Surfaces \mathcal{S}_A and \mathcal{S}_B are defined by buoyancy isosurfaces with $b_A < b_B$. The sea surface is decomposed into three regions, \mathcal{S}_1 , \mathcal{S}_2 , and \mathcal{S}_3 according to the outcrop locations of \mathcal{S}_A and \mathcal{S}_B .

39.3.4 An ocean layer that outcrops

Figure 39.4 depicts a buoyancy layer in the ocean that outcrops at the ocean surface at both of its ends. Following the derivation in Section 39.3.3 we have the following integrated PV contents

$$\mathcal{I}_A \equiv \int_{\mathcal{S}_A} b \boldsymbol{\omega}_a \cdot \hat{\mathbf{n}} d\mathcal{S} + \int_{\mathcal{S}_1 + \mathcal{S}_2 + \mathcal{S}_3} b \boldsymbol{\omega}_a \cdot \hat{\mathbf{n}} d\mathcal{S} = \int_{\mathcal{S}_1 + \mathcal{S}_2 + \mathcal{S}_3} (b - b_A) \boldsymbol{\omega}_a \cdot \hat{\mathbf{n}} d\mathcal{S} \quad (39.49a)$$

$$\mathcal{I}_B \equiv \int_{\mathcal{S}_B} b \boldsymbol{\omega}_a \cdot \hat{\mathbf{n}} d\mathcal{S} + \int_{\mathcal{S}_2} b \boldsymbol{\omega}_a \cdot \hat{\mathbf{n}} d\mathcal{S} = \int_{\mathcal{S}_2} (b - b_B) \boldsymbol{\omega}_a \cdot \hat{\mathbf{n}} d\mathcal{S}. \quad (39.49b)$$

Taking the difference in the integrals leads to the integrated PV within the layer \mathcal{R}_{AB}

$$\mathcal{I}_{AB} \equiv \int_{\mathcal{R}_{AB}} \nabla \cdot (\boldsymbol{\omega}_a b) dV \quad (39.50a)$$

$$= \mathcal{I}_A - \mathcal{I}_B \quad (39.50b)$$

$$= \int_{\mathcal{S}_1} (b - b_A) \boldsymbol{\omega}_a \cdot \hat{\mathbf{n}} d\mathcal{S} + (b_B - b_A) \int_{\mathcal{S}_2} \boldsymbol{\omega}_a \cdot \hat{\mathbf{n}} d\mathcal{S} + \int_{\mathcal{S}_3} (b - b_A) \boldsymbol{\omega}_a \cdot \hat{\mathbf{n}} d\mathcal{S}. \quad (39.50c)$$

39.4 Layer integrated PV budget

In Section 39.3 we developed expressions for the PV-substance integrated over a selection of volumes determined by buoyancy surfaces. That discussion illustrated how the volume integrated potential vorticity has contributions only from boundaries; e.g., where an atmospheric region intersects the ground or ocean, and where an oceanic region intersects the ground or the atmosphere. In this section we further illustrate implications of the impermeability theorem of Section 39.1 by considering a buoyancy (isopycnal) layer within the ocean that intersects the bottom on one side and the atmosphere on the other (Figure 39.5). This discussion aims to expose the physical processes affecting changes to the PV-substance within the layer by unpacking the boundary fluxes.

An isopycnal layer generally moves as it expands and contracts due to both reversible and irreversible processes (waves, currents, mixing). The impermeability theorem means that the total potential vorticity substance for the layer changes only through exchanges at the bottom (boundary between the solid earth and ocean) and air-sea boundaries. Removing interior interfaces from the layer PV budget simplifies the budget analysis, as already revealed in Section 39.3. As per the discussion of Section 39.2, the material in this section applies even when there is no materially invariant potential vorticity. All we require is an Eulerian flux-form budget.

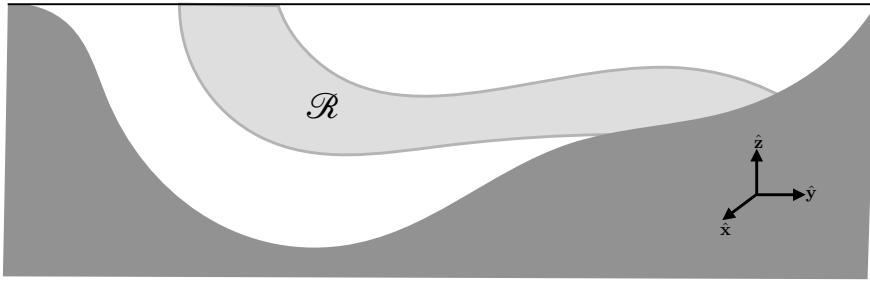


FIGURE 39.5: An isopycnal layer of seawater denoted by \mathcal{R} , with the layer intersecting bottom topography on one side and the atmosphere on the other.

39.4.1 Layer integrated budget

In addition to waves, currents, mixing, and sources affecting the layer interfaces, there is movement of the intersection of the layer with the side boundaries, thus changing the vertical and horizontal extents of these intersections. As a general formulation framework, we derive the isopycnal layer PV budget making use of the Leibniz-Reynolds transport theorem derived in Section 16.3.4. Just as for the layer integrated tracer budget considered in Section 16.5, applying Leibniz-Reynolds to the layer integrated PV budget renders

$$\frac{d}{dt} \left[\int_{\mathcal{R}} \rho Q dV \right] = \int_{\mathcal{R}} \left[\frac{\partial(\rho Q)}{\partial t} + \nabla \cdot (\rho Q \dot{x}) \right] dV, \quad (39.51)$$

where \mathcal{R} is the domain defined by the layer and

$$\dot{x} = \frac{dx}{dt} \quad (39.52)$$

is the velocity for a point on the layer boundary. Making use of the PV equation, $\partial(\rho Q)/\partial t = -\nabla \cdot \mathbf{J}_Q$, and the divergence theorem renders

$$\frac{d}{dt} \left[\int_{\mathcal{R}} \rho Q dV \right] = \int_{\mathcal{R}} \nabla \cdot (-\mathbf{J}_Q + \rho Q \dot{x}) dV = \int_{\partial\mathcal{R}} (-\mathbf{J}_Q + \rho Q \dot{x}) \cdot \hat{n} d\mathcal{S}. \quad (39.53)$$

This result holds around the full domain boundary. Now we decompose that boundary into portions defined by isopycnal layer interfaces and those along the air-sea and land-sea boundaries.

39.4.2 Impermeability across interior layer interfaces

Rather than invoking the impermeability theorem derived in Section 39.1, we rederive it within the present context to further our confidence in its validity. We thus consider the following for interior layer interfaces

$$[-\mathbf{J}_Q + \rho Q \dot{x}] \cdot \hat{n} = [\rho Q (\dot{x} - \mathbf{v}) + \dot{b} \boldsymbol{\omega}_a - \nabla b \wedge \mathbf{F}] \cdot \hat{n} \quad (39.54a)$$

$$= [(\boldsymbol{\omega}_a \cdot \nabla b) (\dot{x} - \mathbf{v}) + \dot{b} \boldsymbol{\omega}_a] \cdot \hat{n} \quad (39.54b)$$

$$= [(\boldsymbol{\omega}_a \cdot \nabla b) (\dot{x} - \mathbf{v}) + (\partial b / \partial t + \mathbf{v} \cdot \nabla b) \boldsymbol{\omega}_a] \cdot \hat{n} \quad (39.54c)$$

$$= [(\boldsymbol{\omega}_a \cdot \nabla b) \dot{x} + (\partial b / \partial t) \boldsymbol{\omega}_a] \cdot \hat{n}, \quad (39.54d)$$

where

$$\hat{n} = \frac{\nabla b}{|\nabla b|} \quad (39.55)$$

is the outward normal direction pointing to regions of higher buoyancy. Now recall that the velocity of a point fixed on an isopycnal interface has a normal component that satisfies equation (39.9) (here applied to isopycnals)

$$\dot{\mathbf{x}} \cdot \hat{\mathbf{n}} = -\frac{\partial b / \partial t}{|\nabla b|}. \quad (39.56)$$

This result then leads to the impermeability statement for isopycnal interfaces in the fluid interior

$$(-\mathbf{J}_Q + \rho Q \dot{\mathbf{x}}) \cdot \hat{\mathbf{n}} = 0. \quad (39.57)$$

We thus conclude that changes to the layer integrated PV occur only via transfer across the land-sea boundary and the air-sea boundary

$$\frac{d}{dt} \left[\int_{\mathcal{R}} \rho Q dV \right] = \int_{\text{land-sea}} [-\mathbf{J}_Q + \rho Q \dot{\mathbf{x}}] \cdot \hat{\mathbf{n}} dS + \int_{\text{air-sea}} [-\mathbf{J}_Q + \rho Q \dot{\mathbf{x}}] \cdot \hat{\mathbf{n}} dS. \quad (39.58)$$

39.4.3 PV flux at the land-sea boundary

We here evaluate the PV flux from equation (39.5a) at the land-sea boundary

$$-\mathbf{J}_Q + \rho Q \dot{\mathbf{x}} = \rho Q (\dot{\mathbf{x}} - \mathbf{v}) + \dot{b} \boldsymbol{\omega}_a + \nabla b \wedge \mathbf{F}. \quad (39.59)$$

At a solid boundary, the no-normal flow boundary condition means that $\hat{\mathbf{n}} \cdot \mathbf{v} = 0$. Likewise, the velocity of a point along the boundary moves along the tangent to the boundary so that $\dot{\mathbf{x}} \cdot \hat{\mathbf{n}} = 0$. The bottom boundary condition is thus affected just by irreversible processes

$$(-\mathbf{J}_Q + \rho Q \dot{\mathbf{x}}) \cdot \hat{\mathbf{n}} = (\dot{b} \boldsymbol{\omega}_a + \nabla b \wedge \mathbf{F}) \cdot \hat{\mathbf{n}}. \quad (39.60)$$

In many parts of the ocean, geothermal heating is negligible so that there is no buoyancy input at the bottom, thus leaving just the contribution from friction

$$\text{no geothermal heating} \implies (-\mathbf{J}_Q + \rho Q \dot{\mathbf{x}}) \cdot \hat{\mathbf{n}} = (\nabla b \wedge \mathbf{F}) \cdot \hat{\mathbf{n}} = (\hat{\mathbf{n}} \wedge \nabla b) \cdot \mathbf{F}. \quad (39.61)$$

Furthermore, in the absence of geothermal heating the buoyancy satisfies a no-flux boundary condition, which can be ensured by having the buoyancy field maintaining

$$\text{no geothermal heating} \implies \hat{\mathbf{n}} \cdot \nabla b = 0. \quad (39.62)$$

Buoyancy isolines thus intersect the bottom parallel to the bottom outward normal, as shown in Figure 39.6. Correspondingly, $(\hat{\mathbf{n}} \wedge \nabla b) \cdot \mathbf{F}$ projects out that component of the friction vector pointing parallel to the bottom with constant geopotential, z . Assuming buoyancy increases upward along the sloping bottom, as per a stably stratified bottom fluid, then $\hat{\mathbf{n}} \wedge \nabla b$ points clockwise around bowls and counter-clockwise around bumps (see Figure 39.6).

Within the bottom boundary layer, quadratic bottom drag is a common parameterization of the acceleration associated with turbulent frictional processes

$$\mathbf{F} = -C_d |\mathbf{u}| \mathbf{u}, \quad (39.63)$$

where C_d is a non-dimensional drag coefficient and \mathbf{u} is the horizontal velocity. In this case the boundary condition for PV takes the form

$$(\nabla b \wedge \mathbf{F}) \cdot \hat{\mathbf{n}} = -C_d |\mathbf{u}| (\hat{\mathbf{n}} \wedge \nabla b) \cdot \mathbf{u}. \quad (39.64)$$

This boundary flux provides a positive PV source in cases where the bottom boundary layer flow is clockwise around abyssal hills and counter-clockwise around abyssal bowls, and conversely for oppositely oriented flow. In contrast, bottom flows that are parallel to ∇b (i.e., flows that are orthogonal to buoyancy isosurfaces at the bottom boundary) provide a zero source for PV since $\nabla b \wedge \mathbf{F} = 0$. This result is expected from the discussion in Section 38.4 and Figure 38.3, where we note that friction changes PV by rotating buoyancy surfaces, with that rotation realized only when friction is not aligned with ∇b .

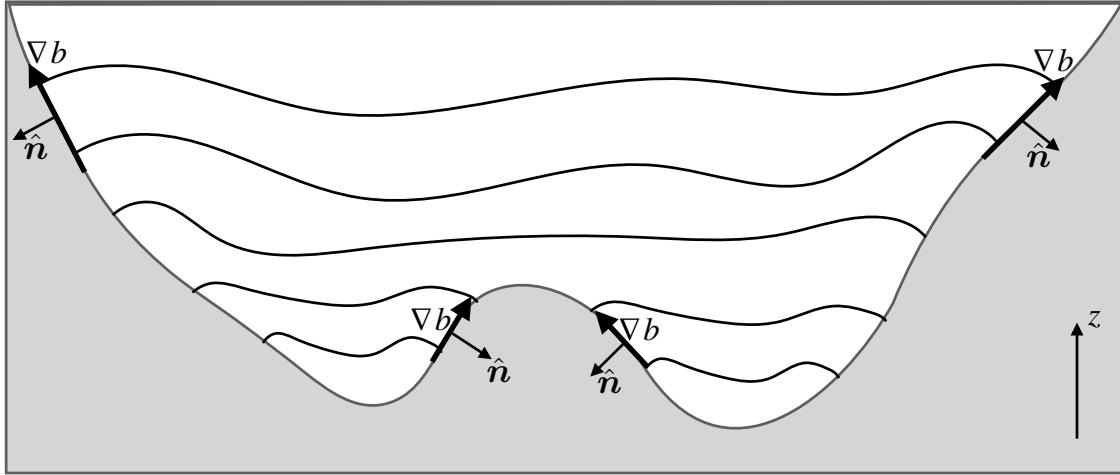


FIGURE 39.6: Buoyancy isosurfaces that intersect the bottom. As discussed in Section 16.5 and depicted in Figure 16.4, in the absence of geothermal heating, a buoyancy isosurface satisfies the no-normal flux bottom boundary condition, $\hat{\mathbf{n}} \cdot \nabla b = 0$. This boundary condition requires buoyancy isosurfaces to be orthogonal to the bottom. Assuming buoyancy increases upward along the sloping bottom, then $\hat{\mathbf{n}} \wedge \nabla b$ points clockwise around bowls and counter-clockwise around bumps (when viewed from above). This structure for the buoyancy surfaces affects how friction impacts on the layer-integrated PV budget, with details provided in Section 39.4.3.

39.4.4 PV flux at the air-sea boundary

We make use of the kinematic boundary condition derived in Section 15.4.3 for the permeable air-sea boundary, where the boundary condition (15.63) leads to

$$\rho \hat{\mathbf{n}} \cdot (\dot{\mathbf{x}} - \mathbf{v}) = \mathcal{Q}_m \quad \text{air-sea boundary} \quad (39.65)$$

with \mathcal{Q}_m the mass per time per surface area crossing the boundary. We are thus led to the air-sea boundary condition

$$(-\mathbf{J}_Q + \rho Q \dot{\mathbf{x}}) \cdot \hat{\mathbf{n}} = Q \mathcal{Q}_m + (\dot{b} \boldsymbol{\omega}_a + \nabla b \wedge \mathbf{F}) \cdot \hat{\mathbf{n}}. \quad (39.66)$$

Besides the irreversible processes, PV is affected at the air-sea interface by the transfer of matter across the boundary via the term $Q \mathcal{Q}_m$. We can think of this term as an advection of PV across the boundary via the boundary mass transport. Note that we might choose to approximate the sea surface as nearly flat, so that

$$\hat{\mathbf{n}} \approx \hat{\mathbf{z}}, \quad (39.67)$$

in which case

$$(-\mathbf{J}_Q + \rho Q \dot{\mathbf{x}}) \cdot \hat{\mathbf{n}} \approx Q \mathcal{Q}_m + \dot{b}(\zeta + f) + (\nabla b \wedge \mathbf{F}) \cdot \hat{\mathbf{z}}. \quad (39.68)$$

The buoyancy source term appears multiplied by the vertical component to the absolute vorticity, so that modifications to PV appear in proportion to the sign of both \dot{b} and $\zeta + f$. We think of this

term as acting to stretch/compress the fluid column so to alter vorticity and hence PV. The friction source arises from the mis-alignment of the horizontal friction vector and the horizontal buoyancy gradient, thus acting like a torque to spin the fluid analogous to the action from baroclinicity described in Section 36.5.

39.4.5 Thought experiments

The surface PV flux (39.68) provides an explicit expression for how surface boundary fluxes affect the PV budget within a buoyancy layer outcropping at the ocean surface. It contains a wealth of physics that can be explored via thought experiments.

PV generation in a fluid with zero initial baroclinicity

Consider a fluid region initially with zero baroclinicity and zero flow so that the initial PV is given by $f N^2$, with N^2 the squared buoyancy frequency. The surface PV flux (39.68) creates PV via the mass flux term and through heating/cooling. If this term alone affected the PV, and it did so uniformly in space, then it would alter PV only via changes in the vertical stratification. More generally, both the mass term and the diabatic term create horizontal buoyancy gradients, which then generate currents and vorticity that generate further contributions to the PV flux.

PV generation in a fluid that is initially homogeneous

Consider an initially homogenous box of seawater with zero PV. In this case it is only the buoyancy term, $\dot{b} f$, that contributes to initial changes in PV. Northern hemisphere ($f > 0$) surface cooling ($\dot{b} < 0$) adds negative PV to the box. Cooling also initiates gravitational instability that mixes the water and in turn spreads the negative PV boundary source throughout the fluid. Cooling adds structure to the buoyancy field by inflating the formerly zero thickness buoyancy layers, with layer inflation originating from the boundary. Once inflated, the impermeability theorem dictates that the layer integrated PV-substance changes only via boundary interactions, whereas stirring and mixing transport PV into the fluid interior. Notably, a region with $f Q < 0$ is locally unstable to symmetric instability, with the generated symmetric instability acting to locally bring the flow towards a zero PV state. However, the constraints from impermeability mean that the net PV-substance remains unchanged within a buoyancy layer, even in the presence of mixing.

39.4.6 Is there a preferred form of the PV-substance flux?

Analysis in this section made use of the Haynes-McIntyre form of the PV-substance flux (equation (39.1))

$$\mathbf{J}_Q = \rho \mathbf{v} Q - \dot{b} \boldsymbol{\omega}_a + \nabla b \wedge \mathbf{F}. \quad (39.69)$$

We could have instead chosen to work with the Marshall form (equation (39.35))

$$\mathbf{J}_Q^{\text{marshall}} = - \left[\frac{\partial \mathbf{v}}{\partial t} + \nabla(m + p/\rho) \right] \wedge \nabla b - \boldsymbol{\omega}_a \frac{\partial b}{\partial t}, \quad (39.70)$$

or the modified kinematic form (equation (39.17))

$$\tilde{\mathbf{J}}_Q^{\text{kin}} = - \frac{\partial \mathbf{v}_a}{\partial t} \wedge \nabla \theta - \boldsymbol{\omega}_a \frac{\partial \theta}{\partial t}. \quad (39.71)$$

These fluxes differ by a gauge choice and yet they each satisfy the impermeability theorem. Subjective choices determine which one is preferred. Importantly, once chosen, we can use only a single form of the flux throughout the budget analysis in order to remain self-consistent.

The PV-substance budget, though invariant to the choice of flux, has distinct physical pictures depending on the choice of the flux. As a particularly clear example consider a steady state budget in which the fluxes take the form

$$\mathbf{J}_Q = \rho \mathbf{v} Q - (\mathbf{v} \cdot \nabla b) \boldsymbol{\omega}_a + \nabla b \wedge \mathbf{F} \quad (39.72)$$

$$\mathbf{J}_Q^{\text{marshall}} = -\nabla(m + p/\rho) \wedge \nabla b \quad (39.73)$$

$$\tilde{\mathbf{J}}_Q^{\text{kin}} = 0. \quad (39.74)$$

The physical picture for $\tilde{\mathbf{J}}_Q^{\text{kin}}$ is rather trivial, whereby the PV-substance stays constant within buoyancy layers and there are zero PV-substance fluxes across *all* boundaries of the layer. In contrast, a steady state budget when working with \mathbf{J}_Q or $\mathbf{J}_Q^{\text{marshall}}$ afford a physical picture of PV-substance entering, leaving, and transported through the buoyancy layers. [Marshall et al. \(2001\)](#) developed a rather elegant analysis framework using $\mathbf{J}_Q^{\text{marshall}}$ for steady budgets, and we explore facets of that approach in Section 38.5 for the special case of a Boussinesq hydrostatic fluid.

Nevertheless, our use of the Haynes-McIntyre PV-substance flux in the present section is motivated by its utility for describing how boundary forcing can change the sign of the PV. Such forcing exposes the flow to a variety of local instabilities (e.g., symmetric, centrifugal, gravitational). [Thomas et al. \(2008\)](#) offer a pedagogical review for the ocean; [Thomas et al. \(2013\)](#) provides a thorough study of the upper reaches of the Gulf Stream; and [Naveira Garabato et al. \(2019\)](#) provide evidence for such boundary forcing in regions of strong abyssal flows. Each of these studies points to the need to further understand details of the boundary PV flux and to furthermore ensure it is properly formulated within numerical models (e.g., [Hallberg and Rhines \(1996\)](#)).



Part VII

Nearly geostrophic balanced flows

Fluid motion dominated by rotation is characterized by a small Rossby number. To zeroth order in an asymptotic expansion in Rossby number, the flow maintains geostrophic balance, which is a balance between the Coriolis acceleration and pressure gradient acceleration. As seen in Chapter 26, the geostrophic balance is diagnostic so that it offers no means to compute the time evolution of the fluid. To obtain a prognostic equation requires going to next order in Rossby number within the asymptotic expansion. The resulting prognostic equation makes use of ageostrophic motions, though only as an intermediate step towards an evolution equation involving just zeroth order geostrophically balanced fields.

The nuts and bolts of this part of the book involve methods of scaling analysis and asymptotic analysis via perturbation series. In Chapter 40, we use these tools to derive equations for planetary geostrophy (PG) and quasi-geostrophy (QG) within the shallow water system, and then PG and QG for continuously stratified flows in Chapters 41 and 42 . PG and QG are useful theoretical models lending insights into ocean and atmosphere fluid mechanics. In particular, PG is commonly used to study features of the large-scale laminar ocean circulation, and QG is ubiquitous in studies of both oceanic and atmospheric flows at or near the deformation radius.

Shallow water planetary and quasi-geostrophy

In this chapter we develop the mechanical equations for planetary geostrophy (PG) and quasi-geostrophy (QG) within the shallow water fluid. The derivations require elements from dimensional analysis and asymptotic methods. Salient details are introduced and the many mathematical steps exposed.

READER'S GUIDE FOR THIS CHAPTER

This chapter is largely technical in nature, aiming to provide the necessary tools from dimensional analysis, scale analysis, and asymptotic methods to derive approximate geostrophic equations. We assume an understanding of the equations for a single layer of shallow water fluid as derived in Chapter 31. We follow this work with further studies of the continuously stratified systems in Chapters 41 and 42. We generally follow the notation in Chapter 5 in [Vallis \(2017\)](#).

40.1	Loose threads	916
40.2	Scaling analysis and the Buckingham-II theorem	916
40.3	Shallow water equations	916
40.3.1	Dimensional scales	917
40.3.2	Physical dimensions	918
40.3.3	Number of non-dimensional parameters	918
40.3.4	Choosing the non-dimensional parameters	918
40.3.5	Assumed values for the non-dimensional parameters	920
40.3.6	Deformation radius and the free surface undulation scale	921
40.3.7	Non-dimensional shallow water equations	922
40.4	Shallow water planetary geostrophy	923
40.5	Shallow water quasi-geostrophy	924
40.5.1	Quasi-geostrophic scaling	924
40.5.2	Outlining the asymptotic method	925
40.5.3	Order Ro^0 asymptotic equations	926
40.5.4	Order Ro^1 asymptotic equations	927
40.6	Physical elements of the theory	928
40.6.1	Dimensional potential vorticity and streamfunction	928
40.6.2	Contributions to the QG potential vorticity	929
40.6.3	Summarizing the equations of quasi-geostrophy	930
40.6.4	Connecting quasi-geostrophic PV to shallow water PV	931
40.6.5	Considering topography to be $\mathcal{O}(\text{Ro}^0)$	932
40.6.6	Two layer quasi-geostrophy	933
40.7	Exercises	934

40.1 Loose threads

- External and internal modes for QG two layers

40.2 Scaling analysis and the Buckingham-II theorem

Scale analysis is ubiquitous in physics, with the Buckingham-II theorem providing a useful framework. The theorem states that the number of dimensionless parameters in a physical system is a function of the number of dimensional parameters or scales K (e.g., scales for the velocity, rotation rate, pressure force, friction force, gravitational acceleration) and the number of physical dimensions R (e.g., time, length, mass). Precisely, Buckingham-II states that the number of dimensionless parameters is

$$N_{\text{dimensionless}} = K - R. \quad (40.1)$$

Different physical systems possessing the same suite of dimensionless parameters are isomorphic. For example, a laboratory study of flow around a cylinder contains two dimensionless parameters: the drag coefficient, C_d , and the Reynolds number, Re . If the problem is scaled up to a building with the same shape, then so long as the values for the dimensionless parameters are the same (e.g., same drag coefficient and same Reynolds number), one can make use of the laboratory analog for determining suitability of the building architecture. Similar isomorphisms exist between flows in a rotating tank and flows in the ocean and atmosphere.

The Buckingham-II theorem does not provide the form of the dimensionless parameters. Nor does the theorem determine their values. This information comes only after introducing physical prejudices surrounding a regime of chosen interest. Additionally, Buckingham-II does not offer information about how the dimensionless parameters might be related. Instead, any such relations are derived from the mechanical equations describing the system.

We focus in this chapter on the regime of large-scale atmospheric and oceanic flow where the shallow water fluid is close to geostrophic balance. That choice then guides the length and time scales, which in turn determines the size of the dimensionless parameters. In many cases, one is able to identify dimensionless parameters that are large or small in particular regimes, which in turn suggests asymptotic analyses to render equations specific to the regime of interest.

40.3 Shallow water equations

A single-layer of inviscid shallow water fluid of thickness h is governed by the velocity and thickness equations (Chapter 31)

$$\frac{\partial \mathbf{u}}{\partial t} + (\mathbf{u} \cdot \nabla) \mathbf{u} + \mathbf{f} \wedge \mathbf{u} = -g \nabla \eta \quad (40.2a)$$

$$\frac{\partial h}{\partial t} + \nabla \cdot (h \mathbf{u}) = 0, \quad (40.2b)$$

where \mathbf{u} is the horizontal velocity that is independent of depth within the layer. The geometry of the layer is specified by the free surface height, $z = \eta(x, y, t)$, and bottom topography, $z = \eta_b(x, y)$ (see Figure 31.1). They are related according to

$$\eta = \eta_b + h = H + \bar{\eta}_b + \Delta\eta = \bar{h} + \bar{\eta}_b + \Delta\eta, \quad (40.3)$$

where $H = \bar{h}$ is the area average layer thickness, $\bar{\eta}_b$ is the area average of the bottom topography, $\Delta\eta = \eta - \bar{\eta}$ is the deviation of the surface height from its area average $\bar{\eta} = \bar{\eta}_b + H$. We will also find occasion to write the layer thickness in the form

$$h = \eta - \eta_b = H + (\bar{\eta}_b - \eta_b) + \Delta\eta = \bar{h} - \Delta\eta_b + \Delta\eta = \bar{h} + \Delta h, \quad (40.4)$$

where $\Delta\eta_b = \eta_b - \bar{\eta}_b$ is the deviation of the bottom topography from its area mean, and we introduced the thickness deviation from the area mean

$$\Delta h = h - \bar{h} = -\Delta\eta - \Delta\eta_b. \quad (40.5)$$

Dexterity with these various geometrical relations is assumed in the following.

40.3.1 Dimensional scales

We identify nine dimensional parameters for the shallow water system.

- LENGTH SCALES

- ★ H = depth scale of the fluid, which we take equal to the area average layer thickness (see Figure 31.1).
- ★ L = horizontal/lateral length scale of motions under consideration. We assume both horizontal directions to have the same length scale. This assumption is not necessarily valid on a rotating planet, where flows in the zonal (east-west) direction can have length scales longer than meridional (north-south) flow scales. Nonetheless, this choice does not preclude the dynamical emergence of anisotropic length scales.
- ★ R = radius of the planet. We include this scale anticipating that for length scales small compared to the earth's radius, the Coriolis parameter may be approximated by a constant (f -plane) or linear function of latitude (β -plane).
- ★ \mathcal{H} = scale for deviations of the free surface height, η , relative to its area average, $\bar{\eta}$, so that $\Delta\eta = \eta - \bar{\eta} \sim \mathcal{H}$.
- ★ \mathcal{B} = scale for deviations of the bottom topography, η_b , relative to its area average, $\bar{\eta}_b$, so that $\Delta\eta_b = \eta_b - \bar{\eta}_b \sim \mathcal{B}$.

- VELOCITY SCALES

- ★ U = velocity scale for horizontal fluid particle motion; i.e., the speed for horizontal currents or winds.
- ★ c = wave speed scale. For the shallow water model, the wave speed scale is given by the shallow water gravity wave (Section 32.8.3)

$$c = \sqrt{g H}. \quad (40.6)$$

We introduce the wave speed anticipating the presence of distinct flow regimes depending on whether the fluid particle speed is larger or smaller than the wave speed. The ratio of the wave speed to particle speed is known as the *Froude number* (Section 40.3.4).

- BODY FORCES: There are two body forces acting on the fluid; one from gravity and one from Coriolis.

- ★ g = gravitational acceleration.
- ★ f = Coriolis frequency.

If we were interested in other forces, such as electromagnetic or frictional forces, then we would have other dimensional parameters corresponding to these forces.

40.3.2 Physical dimensions

There are two physical dimensions in the shallow water system: length, L , and time, T . Notably, there is no mass in the shallow water system. The reason is that the fluid density is assumed uniform within a shallow water layer so that mass is described by area times thickness

$$M = \int \rho dV [\equiv] L^2 H \rho, \quad (40.7)$$

where $[\equiv]$ is read “has dimensions”.

40.3.3 Number of non-dimensional parameters

The Buckingham-II theorem then says there are

$$N_{\text{dimensionless}} = 9 - 2 = 7 \quad (40.8)$$

non-dimensional parameters. What if we incorrectly count the physical dimensions or the dimensional scales/parameters? Fortunately, the process of determining the non-dimensional parameters is largely self-correcting. Namely, in the process of non-dimensionalizing the shallow water equations, the seven non-dimensional parameters will arise as part of the analysis. Hence, making use of Buckingham-II is useful but it is not essential. If one left out a physical dimension or a physical parameter, then it would appear somewhere in the subsequent analysis, often not until near the end where something mathematically or physically inconsistent appears. One must always be cognizant of the need to self-correct when performing dimensional analysis.

40.3.4 Choosing the non-dimensional parameters

There is no unique choice for the non-dimensional parameters. Our choice is guided by experience, interest, and what parameters might be available to experimental control or measurement.

1. VERTICAL TO HORIZONTAL ASPECT RATIO: The ratio of the vertical scale to the horizontal scale defines the aspect ratio

$$\delta_{\text{vertical/horizontal}} = \frac{\text{vertical depth scale}}{\text{horizontal length scale}} = \frac{H}{L}. \quad (40.9)$$

2. RATIO OF HORIZONTAL SCALE TO PLANETARY SCALE: The ratio of the lateral length scale to the planet radius is

$$\delta_{\text{horizontal/planet}} = \frac{\text{lateral length scale}}{\text{planetary length scale}} = \frac{L}{R}. \quad (40.10)$$

3. RATIO OF FREE SURFACE UNDULATION TO DEPTH SCALE: The ratio of the free surface undulation scale to the vertical depth scale is

$$\delta_{\text{free surface/depth}} = \frac{\text{free surface undulation scale}}{\text{vertical depth scale}} = \frac{\mathcal{H}}{H}. \quad (40.11)$$

4. RATIO OF BOTTOM TOPOGRAPHY UNDULATION TO DEPTH: The ratio of the bottom topography undulation scale to the vertical depth scale is

$$\delta_{\text{bottom/depth}} = \frac{\text{bottom topography undulation scale}}{\text{vertical depth scale}} = \frac{\mathcal{B}}{H}. \quad (40.12)$$

5. FROUDE NUMBER: The Froude number is the ratio of the fluid particle speed to the wave speed. For the shallow water system, this ratio is

$$\text{Fr} = \frac{U}{c} = \frac{U}{\sqrt{g H}}. \quad (40.13)$$

6. ROSSBY NUMBER: The Rossby number is the ratio of the fluid particle acceleration scale to the Coriolis acceleration

$$\text{Ro} = \frac{\text{parcel acceleration}}{\text{Coriolis acceleration}}. \quad (40.14)$$

The particle acceleration scale is determined by the local time tendency plus advection

$$\text{particle acceleration} = \frac{\partial \mathbf{u}}{\partial t} + (\mathbf{u} \cdot \nabla) \mathbf{u} \sim \frac{U}{T} + \frac{U^2}{L}. \quad (40.15)$$

The local time tendency and advection generally have distinct scales, thus leading to the potential for two Rossby numbers

$$\text{Ro}^{\text{local}} = \frac{1}{T f} \quad \text{and} \quad \text{Ro}^{\text{adv}} = \frac{U}{L f}. \quad (40.16)$$

In the following we consider the two Rossby numbers to be on the same order so that the advective contributions to material time evolution are comparable to local time changes. We refer to this choice as an *advection time scale*, whereby

$$T = \frac{L}{U} \implies \frac{U^2}{L} = \frac{U}{T}, \quad (40.17)$$

so that there is only one Rossby number

$$\text{Ro} = \frac{1}{f T} = \frac{U}{f L}. \quad (40.18)$$

Another interpretation for the Rossby number is the ratio of the relative vorticity to the planetary vorticity

$$\text{Ro} = \frac{\text{relative vorticity}}{\text{planetary vorticity}} \quad (40.19)$$

With the relative vorticity scaling as U/L and the planetary vorticity scaling as f , we recover the expression (40.18) for the Rossby number.

7. GEOSTROPHIC NUMBER: We define the geostrophic number as the ratio of the Coriolis acceleration to the pressure gradient acceleration

$$\text{Ge} = \frac{\text{Coriolis acceleration}}{\text{pressure gradient acceleration}}. \quad (40.20)$$

The Coriolis acceleration scales as

$$\text{Coriolis acceleration} \sim f U \quad (40.21)$$

whereas the pressure gradient acceleration, $-g \nabla \eta$, scales as

$$\text{pressure gradient acceleration} \sim \frac{g \mathcal{H}}{L}, \quad (40.22)$$

so that

$$Ge = \frac{\text{Coriolis acceleration}}{\text{pressure gradient acceleration}} = \frac{f U}{(g/L) \mathcal{H}}. \quad (40.23)$$

40.3.5 Assumed values for the non-dimensional parameters

We now enumerate the assumed values for the non-dimensional parameters.

1. SMALL VERTICAL TO HORIZONTAL ASPECT RATIO: The aspect ratio is generally small for large-scale atmospheric and oceanic fluid systems

$$\delta_{\text{vertical/horizontal}} = H/L \ll 1. \quad (40.24)$$

This assumption was made when formulating the shallow water system, which is based on hydrostatic balance (see Section 31.2). We thus retain this assumption as we further scale the shallow water system.

2. SMALL OR ORDER ONE RATIO OF HORIZONTAL TO PLANETARY SCALES: The ratio of the lateral length scale to the planet radius is small for quasi-geostrophic systems and order unity for planetary geostrophy

$$\delta_{\text{horizontal/planet}} = L/R \ll 1 \quad \text{quasi-geostrophy} \quad (40.25a)$$

$$\delta_{\text{horizontal/planet}} = L/R \sim 1 \quad \text{planetary geostrophy}. \quad (40.25b)$$

3. RATIO OF FREE SURFACE UNDULATION TO DEPTH: The ratio \mathcal{H}/H will be implied by assuming a unit geostrophic number below.
4. RATIO OF BOTTOM TOPOGRAPHY UNDULATION TO DEPTH: For quasi-geostrophy, we assume that undulations in the bottom topography are small relative to the depth, whereas there is no restriction for planetary geostrophy. “Small” in the present context is determined by the Rossby number, in which case

$$\delta_{\text{bottom/depth}} = \mathcal{B}/H = Ro \quad \text{for quasi-geostrophy}. \quad (40.26)$$

5. FROUDE NUMBER: We will see that the Froude number is implied by scales assumed for the other non-dimensional numbers.
6. SMALL ROSSBY NUMBER: The Rossby number is assumed small

$$Ro = U/(f L) \ll 1, \quad (40.27)$$

which means that the Coriolis acceleration is a leading order term in the horizontal velocity equation (40.2a).

7. UNIT GEOSTROPHIC NUMBER: The geostrophic number is assumed to be order unity

$$\text{Ge} \sim 1. \quad (40.28)$$

This assumption means that the Coriolis acceleration and pressure gradient acceleration scale together

$$f U \sim (g/L) \mathcal{H}. \quad (40.29)$$

Making use of the momentum equation (40.2a), we see that this scaling is consistent only so long as the Rossby number is small, $\text{Ro} \ll 1$. Furthermore, this scaling constrains the scale of the free surface undulation, \mathcal{H} , as we discuss in Section 40.3.6.

40.3.6 Deformation radius and the free surface undulation scale

We determine the scale for the free surface height undulation, \mathcal{H} , by making use of the assumed order unity geostrophy number. For this purpose, start from the geostrophic scaling of Coriolis and pressure gradient accelerations, (40.29), to express the free surface undulation scale according to

$$\Delta\eta \sim \mathcal{H} = \frac{f U L}{g} = \text{Ro} \frac{f^2 L^2}{g} = \text{Ro} H \frac{f^2 L^2}{g H} = \text{Ro} H \left[\frac{L}{L_d} \right]^2. \quad (40.30)$$

In the final equality we introduced the *deformation radius*

$$L_d = f^{-1} \sqrt{g H}. \quad (40.31)$$

We previously encountered the deformation radius when discussing geostrophic adjustment in Section 35.7. In Section 40.6.2 we further discuss its role as a regime boundary (between relative vorticity and vortex stretching) for the quasi-geostrophic potential vorticity. The deformation radius decreases toward the poles, so that rotational effects are felt by smaller scales in the high latitudes than in the tropics.

We can use L_d to rewrite the Froude number as the ratio of the advection speed to the rotational speed

$$\text{Fr} = \frac{U}{\sqrt{g H}} = \frac{U}{f L_d} = \text{Ro} \frac{L}{L_d}. \quad (40.32)$$

Furthermore, the squared ratio of the deformation radius to the lateral length scale is termed the *Burger number*

$$\text{Bu} = \left[\frac{L_d}{L} \right]^2. \quad (40.33)$$

Use of the Burger number allows us to write the Froude number in terms of the Rossby number and Burger number

$$\text{Fr} = \frac{\text{Ro}}{\sqrt{\text{Bu}}}. \quad (40.34)$$

Likewise, the free surface height undulation scale can be written

$$\mathcal{H} = H \text{Ro} \left[\frac{L}{L_d} \right]^2 = H \frac{\text{Ro}}{\text{Bu}} = H \frac{\text{Fr}^2}{\text{Ro}}. \quad (40.35)$$

Hence, the ratio of the free surface undulations to the layer thickness (depth) scale is given by

$$\delta_{\text{free surface/depth}} = \frac{\mathcal{H}}{H} = \text{Ro} \left[\frac{L}{L_d} \right]^2 = \frac{\text{Ro}}{\text{Bu}} = \frac{\text{Fr}^2}{\text{Ro}}. \quad (40.36)$$

Again, this scaling is implied by making the dynamical assumption of a unit geostrophic number, which means that the pressure gradient acceleration scales according to the Coriolis acceleration.

40.3.7 Non-dimensional shallow water equations

To non-dimensionalize the shallow water equations we introduce non-dimensional space, time, velocity, and Coriolis variables, denoted by a widehat¹

$$t = T \widehat{t} \quad (x, y) = L (\widehat{x}, \widehat{y}) \quad \partial_t = \frac{\partial \widehat{t}}{T} \quad \nabla = \frac{\widehat{\nabla}}{L} \quad (u, v) = U (\widehat{u}, \widehat{v}) \quad f = f_0 \widehat{f}, \quad (40.37)$$

where f_0 is the Coriolis parameter at the central latitude for the β -plane approximation (Section 23.2). We also require non-dimensional variables for the surface and bottom undulations

$$\Delta\eta = \mathcal{H} \widehat{\eta} \quad \Delta\eta_b = \mathcal{B} \widehat{\eta}_b \quad h = H + \Delta\eta - \Delta\eta_b = H + \mathcal{H} \widehat{\eta} - \mathcal{B} \widehat{\eta}_b, \quad (40.38)$$

where we used equation (40.4) for the layer thickness. Importantly, we assume that the non-dimensional variables (the widehat variables) are order unity.

Non-dimensional velocity equation

Introducing the above variables into the shallow water velocity equation (40.2a) renders

$$\frac{U}{T} \frac{\partial \widehat{\mathbf{u}}}{\partial \widehat{t}} + \frac{U^2}{L} (\widehat{\mathbf{u}} \cdot \widehat{\nabla}) \widehat{\mathbf{u}} + f_0 U (\widehat{\mathbf{f}} \wedge \widehat{\mathbf{u}}) = - \frac{g \mathcal{H}}{L} \widehat{\nabla} \widehat{\eta}. \quad (40.39)$$

As before, we assume the time scale is given by the advection time

$$T = \frac{L}{U} = \frac{1}{\text{Ro } f_0}, \quad (40.40)$$

so that dividing by $f_0 U$ leads to

$$\text{Ro} \left[\frac{\partial \widehat{\mathbf{u}}}{\partial \widehat{t}} + (\widehat{\mathbf{u}} \cdot \widehat{\nabla}) \widehat{\mathbf{u}} \right] + (\widehat{\mathbf{f}} \wedge \widehat{\mathbf{u}}) = - \left[\frac{g H}{f_0 L U} \text{Ro} \right] \widehat{\nabla} \widehat{\eta}, \quad (40.41)$$

where we set $\mathcal{H} = H (\text{Ro}/\text{Bu})$ according to equation (40.36). We reduce the factor on the right hand side according to

$$\frac{g H}{f_0 L U} \frac{\text{Ro}}{\text{Bu}} = \frac{g H}{f_0 L U} \frac{U}{f_0 L} \frac{L^2}{L_d^2} = \frac{g H}{f_0 L U} \frac{U}{f_0 L} \frac{L^2 f_0^2}{g H} = 1. \quad (40.42)$$

Hence, the non-dimensional inviscid shallow water velocity equation takes on the rather elegant form

$$\text{Ro} \left[\frac{\partial \widehat{\mathbf{u}}}{\partial \widehat{t}} + (\widehat{\mathbf{u}} \cdot \widehat{\nabla}) \widehat{\mathbf{u}} \right] + \widehat{\mathbf{f}} \wedge \widehat{\mathbf{u}} = - \widehat{\nabla} \widehat{\eta}. \quad (40.43)$$

Introducing the non-dimensional material time derivative

$$\frac{D}{Dt} = \frac{\partial}{\partial \widehat{t}} + \widehat{\mathbf{u}} \cdot \widehat{\nabla} \quad (40.44)$$

brings the velocity equation to

$$\text{Ro} \frac{D \widehat{\mathbf{u}}}{Dt} + \widehat{\mathbf{f}} \wedge \widehat{\mathbf{u}} = - \widehat{\nabla} \widehat{\eta}. \quad (40.45)$$

We see that the velocity equation is consistent with a unit geostrophy number (i.e., Coriolis acceleration balances pressure gradient acceleration) if and only if the Rossby number is small, thus eliminating the material acceleration.

¹The L^AT_EX widehat symbol is used for non-dimensional variables, such as the non-dimensional velocity, $\widehat{\mathbf{u}}$. The widehat is distinguished from the hat used for unit vectors, such as for the vertical unit vector, $\widehat{\mathbf{z}}$. We also use widehats for thickness weighted means in Chapter 47, but that usage is completely distinct from the non-dimensionalization usage in the present chapter.

Non-dimensional thickness equation

The thickness (continuity) equation (40.2b) can be written

$$\frac{\partial \Delta\eta}{\partial t} + \nabla \cdot [(H + \Delta\eta - \Delta\eta_b) \mathbf{u}] = 0, \quad (40.46)$$

which takes on the non-dimensional form

$$\frac{\mathcal{H}}{T} \frac{\partial \hat{\eta}}{\partial \hat{t}} + \frac{U H}{L} \hat{\nabla} \cdot [(1 + \hat{\eta}) \mathcal{H}/H - \hat{\eta}_b \mathcal{B}/H] \hat{\mathbf{u}} = 0. \quad (40.47)$$

The advective time scaling $T = L/U$ brings the thickness equation to

$$\frac{\mathcal{H}}{H} \frac{\partial \hat{\eta}}{\partial \hat{t}} + \hat{\nabla} \cdot [(1 + \hat{\eta}) \mathcal{H}/H - \hat{\eta}_b \mathcal{B}/H] \hat{\mathbf{u}} = 0. \quad (40.48)$$

40.4 Shallow water planetary geostrophy

We make use of the non-dimensional equations derived in Section 40.3.7 to derive the dynamical equations for planetary geostrophy. Our presentation is rather brief, since the derivation is rather trivial technically. In Chapter 41 we develop the theory of planetary geostrophic flows, with this theory of particular use for understanding the large-scale ocean circulation.

Planetary geostrophy is realized by dropping the fluid particle acceleration from the momentum equation (40.45), given that it is one order of Rossby number smaller than the Coriolis and pressure gradient accelerations. This assumption means that the velocity equation reduces to the geostrophic balance

$$\hat{\mathbf{f}} \wedge \hat{\mathbf{u}} = -\hat{\nabla} \hat{\eta}. \quad (40.49)$$

We furthermore assume that the Rossby number and Burger numbers scale together

$$\text{Ro} \sim \text{Bu} = (L_d/L)^2 \ll 1, \quad (40.50)$$

so that the horizontal length scale for the planetary geostrophic flow is much larger than the deformation radius

$$L \gg L_d. \quad (40.51)$$

This assumption is consistent with dropping the material acceleration term in the velocity equation. Although the velocity equation is greatly simplified, we make no assumption concerning the thickness equation. Consequently, the free surface and bottom undulations are unconstrained with planetary geostrophic flows, so long as the flow maintains the hydrostatic balance.

In summary, the thickness equation for the planetary geostrophic fluid retains its full unapproximated form, whereas the velocity equation reduces to geostrophy. Reintroducing dimensions leads to the planetary geostrophic equations

$$f \hat{\mathbf{z}} \wedge \mathbf{u} = -g \nabla \eta \quad \text{and} \quad \frac{Dh}{Dt} = -h \nabla \cdot \mathbf{u} \quad \text{and} \quad h = \eta - \eta_b. \quad (40.52)$$

Since the Coriolis parameter retains its spatial dependence, and so includes the beta effect, the horizontal velocity field is divergent

$$f \nabla \cdot \mathbf{u} = -\beta (g/f) \partial_x \eta = -\beta v. \quad (40.53)$$

As shown in Exercise 40.1, the shallow water planetary geostrophic equations are equivalent to

$$f \hat{\mathbf{z}} \wedge \mathbf{u} = -g \nabla \eta \quad \text{and} \quad \frac{DQ}{Dt} = 0 \quad \text{with} \quad Q = \frac{f}{h}. \quad (40.54)$$

We encountered this form for the potential vorticity in Section 35.4, in which the shallow water potential vorticity, $(f + \zeta)/h$, is approximated by f/h in the small Rossby number limit.

40.5 Shallow water quasi-geostrophy

In this section we develop the quasi-geostrophic equations for a single shallow water fluid layer. Doing so requires more work than for the planetary geostrophic equations. For that purpose, we use rudimentary of asymptotic methods with the Rossby number the small parameter.

40.5.1 Quasi-geostrophic scaling

Quasi-geostrophic scaling is based on the following assumptions, with the first and second shared with planetary geostrophy whereas the remaining are distinct to quasi-geostrophy.

1. SMALL ROSSBY NUMBER: $\text{Ro} \ll 1$, which is fundamental to geostrophic scaling.
2. ADVECTIVE TIME SCALE: $T \sim L/U$; that is, the time scale is determined by advection, which is how time has scaled throughout this chapter.
3. ORDER ROSSBY NUMBER BETA EFFECT: $|\beta L| \ll |f_0|$, which means that the Coriolis frequency does not vary much from its central value. To incorporate this assumption into the asymptotics, we expand the non-dimensional Coriolis parameter in terms of the Rossby number²

$$\hat{\mathbf{f}} = \mathbf{f}/f_0 = (1 + \beta y/f_0) \hat{\mathbf{z}} \equiv (\hat{f}_0 + \text{Ro} \hat{\beta} \hat{y}) \hat{\mathbf{z}}. \quad (40.55)$$

In this equation, $\hat{\mathbf{z}}$ is the unit vector in the vertical, the time scaling in equation (40.40) renders³

$$\hat{\beta} \hat{y} = \beta y/(\text{Ro} f_0) = T \beta y \quad \text{and} \quad \hat{f}_0 = f_0/f_0 = 1. \quad (40.56)$$

The Coriolis expression (40.55) and the scaling (40.56) are motivated by assuming the horizontal scales of motion are on the same order as the deformation radius, and that the Coriolis frequency does not vary much from its central value. Quasi-geostrophy is thus formulated within the beta plane approximation discussed in Section 23.2.

4. BURGER NUMBER ORDER ONE: $Bu \sim 1$, which means that the horizontal scales of motion for the quasi-geostrophic flows are on the order as the deformation radius, $L \sim L_d$.
5. ORDER ROSSBY NUMBER FREE SURFACE UNDULATIONS: From equation (40.36), an order unity Burger number means that undulations of the free surface height scale according to the Rossby number: $\mathcal{H}/H = \text{Ro}$, so that free surface height undulations are small.
6. ORDER ROSSBY NUMBER BOTTOM TOPOGRAPHY UNDULATIONS: As seen in Section 40.6.4, for the quasi-geostrophic potential vorticity to correspond to the small Rossby number version of the shallow water potential vorticity requires the topography undulations to scale as $\mathcal{B}/H = \text{Ro}$. The assumed scaling for the bottom topography undulation pairs with that for the free surface, so that the layer thickness undulations, $\Delta h = \Delta\eta - \Delta\eta_b$, also scale as Ro. We are thus able to take a sensible Ro expansion of the $1/h$ factor appearing in the shallow water potential vorticity (Section 40.6.4). It is useful to maintain a direct connection to the shallow water model as doing so helps to ensure that the resulting asymptotic theory is self-consistent.

²One could conceive of another small parameter that scales the beta effect, but the resulting asymptotics would be more difficult to manage given the need to keep track of two small parameters.

³Although $\hat{f}_0 = 1$ in equation (40.56), it is useful to retain this term as a placeholder in the manipulations to follow.

In particular, it ensures that quasi-geostrophic energetics are physically sensible since their is a direct lineage to the shallow water energetics.

The $\mathcal{B}/H = \text{Ro}$ scaling is consistent with the assumption that the planetary beta effect is small. Together, the two scalings $\mathcal{B}/H = \text{Ro}$ and $\mathcal{H}/H = \text{Ro}$ mean that the effective beta effect (arising from planetary and topographic variations; see Section 40.6.1 below) are small.

40.5.2 Outlining the asymptotic method

To derive the quasi-geostrophic shallow water model, we employ an asymptotic series method with the Rossby number as the small parameter and stop at the first nontrivial order. For this purpose, recall the non-dimensional shallow water equatoins from Section 40.3.7, and make use of the assumed $Bu \sim 1$ scaling

$$\text{Ro} \frac{D\hat{\mathbf{u}}}{Dt} + (\hat{\mathbf{f}} \wedge \hat{\mathbf{u}}) = -\hat{\nabla}\hat{\eta} \quad (40.57a)$$

$$\text{Ro} \left[\frac{\partial(\hat{\eta} - \hat{\eta}_b)}{\partial t} + \hat{\nabla} \cdot [(\hat{\eta} - \hat{\eta}_b) \hat{\mathbf{u}}] \right] = -\hat{\nabla} \cdot \hat{\mathbf{u}}. \quad (40.57b)$$

We brought the time independent bottom topography, $\hat{\eta}_b$, into the time derivative for the thickness equation to remain symmetric with $\hat{\eta}$.

Asymptotic expansion of prognostic fields

Asymptotic methods are ideally suited for non-dimensional equations since we can unambiguously determine scales via the size of non-dimensional parameters. We here assume the Rossby number to be small, in which case we are led to perform an asymptotic expansion of the prognostic fields in terms of the Rossby number. There are three prognostic fields, $\hat{u}, \hat{v}, \hat{\eta}$, and corresponding vertical velocity, \hat{w} , which we assume can be written as an asymptotic series

$$\hat{u} = \hat{u}_0 + \text{Ro} \hat{u}_1 + \text{Ro}^2 \hat{u}_2 + \dots \quad (40.58a)$$

$$\hat{v} = \hat{v}_0 + \text{Ro} \hat{v}_1 + \text{Ro}^2 \hat{v}_2 + \dots \quad (40.58b)$$

$$\hat{w} = \hat{w}_0 + \text{Ro} \hat{w}_1 + \text{Ro}^2 \hat{w}_2 + \dots \quad (40.58c)$$

$$\hat{\eta} = \hat{\eta}_0 + \text{Ro} \hat{\eta}_1 + \text{Ro}^2 \hat{\eta}_2 + \dots \quad (40.58d)$$

We thus refer to the zeroth, first, second, etc. “order” of the asymptotic expansion. The velocity field is non-divergent at each order of Rossby number, so that

$$\nabla \cdot \hat{\mathbf{v}}_n = 0 \quad \forall n. \quad (40.59)$$

Practical goals

The practical goal of asymptotic analysis is to develop a closed set of prognostic equations for functions appearing in the asymptotic expansions (40.58a)-(40.55). For our purposes, we are content to stop at the lowest nontrivial order, meaning the point at which there is a prognostic equation that provides a means to move the system forward in time. Motivation for asymptotic analysis is to produce an equation set offering a means to focus analysis on dynamics most active under the regime determined by the chosen non-dimensional parameters. Each higher order in asymptotic expansion generally requires more complex algebraic manipulations. Hence, pursuit of higher order expansions should be motivated by first determining that the lower order equation set remains physically lacking.

Enabling the machinery

At this point we enable the machinery by “turning the crank”. To do so, we insert the asymptotic expansions (40.58a)-(40.58d) into the non-dimensional partial differential equations (40.57a) and (40.57b). Since Ro is arbitrarily small, and all non-dimensional fields are order unity regardless their order, the only means to maintain self-consistency is for terms to balance at equal order in Rossby number. Hence, we do not mix terms from different orders of Rossby number as we give care while organizing terms. This observation is basic to asymptotic methods.

Again, our goal is to establish a set of prognostic equations that allows us to evolve a state that is arbitrarily close to geostrophic balance. We anticipate that at zeroth order, the asymptotic method will offer us just the geostrophic balance, which has no prognostic value. Hence, we need to go at least to order Ro^1 , and hopefully no further as the algebraic tedium increases with order. With that anticipation and hope, we only keep track of terms of order Ro^0 and Ro^1 , in which the momentum and continuity equations become

$$\text{Ro} \frac{D\hat{\mathbf{u}}_0}{Dt} + (\hat{f}_0 + \text{Ro} \hat{\beta} \hat{y}) \hat{\mathbf{z}} \wedge (\hat{\mathbf{u}}_0 + \text{Ro} \hat{\mathbf{u}}_1) = -\hat{\nabla}(\hat{\eta}_0 + \text{Ro} \hat{\eta}_1) \quad (40.60\text{a})$$

$$\text{Ro} \left[\frac{\partial \hat{\eta}_0}{\partial \hat{t}} + \hat{\nabla} \cdot [(\hat{\eta}_0 - \hat{\eta}_b) \hat{\mathbf{u}}_0] + \hat{\nabla} \cdot \hat{\mathbf{u}}_1 \right] = -\hat{\nabla} \cdot \hat{\mathbf{u}}_0. \quad (40.60\text{b})$$

40.5.3 Order Ro^0 asymptotic equations

Terms in equations (40.60a) and (40.60b) balancing at order Ro^0 are given by

$$\hat{f}_0 \wedge \hat{\mathbf{u}}_0 = -\hat{\nabla} \hat{\eta}_0 \quad (40.61\text{a})$$

$$\hat{\nabla} \cdot \hat{\mathbf{u}}_0 = 0. \quad (40.61\text{b})$$

The zeroth order velocity equation (40.61a) is the *f*-plane geostrophic balance. The curl of this equation leads to the horizontal non-divergence condition

$$\hat{\nabla} \cdot \hat{\mathbf{u}}_0 = 0, \quad (40.62)$$

which is identical to the zeroth order thickness equation (40.61b). Hence, the zeroth order horizontal velocity is given by *f*-plane geostrophy.

Given the non-divergence condition (40.61b), the zeroth order velocity field can be written in terms of a *geostrophic streamfunction*

$$\hat{u}_0 = -\frac{\partial \hat{\psi}_0}{\partial \hat{y}} \quad \text{and} \quad \hat{v}_0 = \frac{\partial \hat{\psi}_0}{\partial \hat{x}} \quad \text{and} \quad \hat{\zeta}_0 = \hat{\nabla}^2 \hat{\psi}_0, \quad (40.63)$$

where the zeroth order streamfunction is the ratio of the zeroth order surface height to zeroth order Coriolis parameter

$$\hat{\psi}_0 = \hat{\eta}_0 / \hat{f}_0. \quad (40.64)$$

We also introduced the zeroth order vorticity, $\hat{\zeta}_0$, which plays a role in the order Ro^1 equations. The corresponding dimensionful quantities are

$$\psi = (U L) \hat{\psi}_0 \quad \text{and} \quad \zeta = (U/L) \hat{\zeta}_0. \quad (40.65)$$

We dropped the asymptotic label on the dimensional quantities to reduce clutter, and since we will only evolve the zeroth order fields. In the following, the higher order fields are referred to as *ageostrophic* fields, whereas the zeroth order fields are the geostrophic fields.

40.5.4 Order Ro^1 asymptotic equations

The zeroth order equations do not render a prognostic equation, prompting us to consider equations at order Ro^1

$$\frac{\text{D}_0 \hat{\mathbf{u}}_0}{\text{Dt}} + \hat{f}_0 \hat{\mathbf{z}} \wedge \hat{\mathbf{u}}_1 + \hat{\beta} \hat{y} \hat{\mathbf{z}} \wedge \hat{\mathbf{u}}_0 = -\hat{\nabla} \hat{\eta}_1 \quad (40.66a)$$

$$\frac{\text{D}_0(\hat{\eta}_0 - \hat{\eta}_b)}{\text{Dt}} = -\hat{\nabla} \cdot \hat{\mathbf{u}}_1. \quad (40.66b)$$

At this order, the material time derivative makes use of *only* the zeroth order geostrophic horizontal velocity

$$\frac{\text{D}_0}{\text{Dt}} = \frac{\partial}{\partial \hat{t}} + \hat{\mathbf{u}}_0 \cdot \hat{\nabla}. \quad (40.67)$$

The set of first order equations (40.66a) and (40.66b) appear to be unclosed because the evolution equation for zeroth order (geostrophic) terms is dependent on first order (ageostrophic) terms. However, the ageostrophic terms can be eliminated using two steps. First, we produce the vorticity equation from the momentum equation, which removes the ageostrophic pressure gradient $-\hat{\nabla} \hat{\eta}_1$. Second, combine the vorticity equation and continuity equation to eliminate the horizontal convergence of the ageostrophic velocity, $-\hat{\nabla} \cdot \hat{\mathbf{u}}_1$. The second step leads to the quasi-geostrophic potential vorticity equation. Although details are specific to our study of shallow water quasi-geostrophy, similar steps are frequently encountered in other balanced geophysical fluid systems.

The geostrophic vorticity equation

Taking the curl of the momentum equation (40.66a) eliminates the ageostrophic pressure gradient, $\hat{\nabla} \hat{\eta}_1$, thus producing the vorticity equation

$$\frac{\partial \hat{\zeta}_0}{\partial \hat{t}} + (\hat{\mathbf{u}}_0 \cdot \hat{\nabla}) (\hat{\zeta}_0 + \hat{\beta} \hat{y}) = -\hat{f}_0 \hat{\nabla} \cdot \hat{\mathbf{u}}_1. \quad (40.68)$$

The right hand side term arises from stretching in the presence of rotation. To this order, stretching arises just from the planetary vorticity, with stretching due to relative vorticity appearing at a higher order. Since $\hat{\beta} \hat{y}$ is time independent, we can write the vorticity equation using the geostrophic material time derivative

$$\frac{\text{D}_0 (\hat{\zeta}_0 + \hat{\beta} \hat{y})}{\text{Dt}} = -\hat{f}_0 \hat{\nabla} \cdot \hat{\mathbf{u}}_1. \quad (40.69)$$

The quasi-geostrophic potential vorticity equation

As anticipated, we need one more step to close the system since the evolution of zeroth order vorticity in equations (40.68) and (40.69) are a function of vortex stretching induced by convergence of the first order velocity. To eliminate $\hat{\nabla} \cdot \hat{\mathbf{u}}_1$, we substitute from the thickness equation (40.66b), thus leading to a prognostic equation involving just zeroth order terms

$$\frac{\partial [\hat{\zeta}_0 + \hat{\beta} \hat{y} - \hat{f}_0 (\hat{\eta}_0 - \hat{\eta}_b)]}{\partial \hat{t}} + \hat{\mathbf{u}}_0 \cdot \hat{\nabla} [\hat{\zeta}_0 + \hat{\beta} \hat{y} - \hat{f}_0 (\hat{\eta}_0 - \hat{\eta}_b)] = 0, \quad (40.70)$$

which can be written in the material form

$$\frac{\text{D}_0}{\text{Dt}} [\hat{\zeta}_0 + \hat{\beta} \hat{y} - \hat{f}_0 (\hat{\eta}_0 - \hat{\eta}_b)] = 0. \quad (40.71)$$

Finally, we introduce the geostrophic streamfunction $\hat{\psi}_0 = \hat{\eta}_0/f_0$ (equation (40.64)) to render

$$\frac{D_0}{Dt} \left[\hat{\nabla}^2 \hat{\psi}_0 + \hat{\beta} \hat{y} + \hat{f}_0 \hat{\eta}_b - \hat{f}_0^2 \hat{\psi}_0 \right] = 0. \quad (40.72)$$

Equation (40.72) is a statement of the material conservation of quasi-geostrophic potential vorticity, where material evolution is defined by the horizontal geostrophic currents (equation (40.67)). This equation is the culmination of our quest to derive a prognostic equation for the evolution of geostrophic flow. It enables us to evolve the geostrophic velocity and geostrophic free surface by probing, but not explicitly determining, the leading order ageostrophic motions. Practical use of the quasi-geostrophic method is based on time stepping the quasi-geostrophic potential vorticity, and then inverting the potential vorticity equation to diagnose the streamfunction to then determine the geostrophic velocity and free surface. That inversion requires solving an elliptic boundary value problem.

40.6 Physical elements of the theory

The asymptotic analysis of Section 40.5 worked with non-dimensional quantities, which are suitable for determining the scales required for asymptotic methods. Now that we have worked through the details, we can make use of that effort to both further the theory and to understand its physical content. For that purpose we find it essential to reintroduce physical dimensions.

40.6.1 Dimensional potential vorticity and streamfunction

To introduce physical dimensions, we invert the relations used in Section 40.3.7, in which we make use of

$$\hat{t} = T^{-1} t \quad (\hat{x}, \hat{y}) = L^{-1} (x, y) \quad \partial_{\hat{t}} = T \partial_t \quad \hat{\nabla} = L \nabla \quad (\hat{u}, \hat{v}) = U^{-1} (u, v) \quad (40.73a)$$

$$\hat{\eta} = \mathcal{H}^{-1} \Delta \eta \quad \hat{\eta}_b = \mathcal{B}^{-1} \Delta \eta_b \quad \mathcal{H} = H \text{Ro} \quad \mathcal{B} = H \text{Ro} \quad (40.73b)$$

$$\hat{f}_0 = f_0/f_0 \quad \hat{\beta} \hat{y} = \beta y / (Ro f_0) = (L/U) \beta y \quad \hat{\zeta} = (L/U) \zeta = L^2 \nabla^2 \psi. \quad (40.73c)$$

We generally drop asymptotic subscript notation on the dimensional terms to help reduce clutter.

For the potential vorticity we start from the non-dimensional quasi-geostrophic potential vorticity (40.71) and find

$$\hat{q} = \hat{\zeta}_0 + \hat{\beta} \hat{y} + \hat{f}_0 \hat{\eta}_b - \hat{f}_0 \hat{\eta}_0 \quad (40.74a)$$

$$= \frac{L}{U} (\zeta + \beta y) + \frac{\Delta \eta_b}{\mathcal{B}} - \frac{\Delta \eta}{\mathcal{H}} \quad (40.74b)$$

$$= \frac{L}{U} (\zeta + \beta y) + \frac{\Delta \eta_b}{H \text{Ro}} - \frac{\Delta \eta}{H \text{Ro}} \quad (40.74c)$$

$$= \frac{L}{U} \left[\zeta + \beta y - \frac{f_0 (\Delta \eta - \Delta \eta_b)}{H} \right] \quad (40.74d)$$

$$= \frac{L}{U} \left[\zeta + \beta y - \frac{g (\Delta \eta - \Delta \eta_b)}{f_0} \frac{1}{L_d^2} \right] \quad (40.74e)$$

$$= \frac{L}{U} [\zeta + \beta y - L_d^{-2} (\psi - \psi_b)], \quad (40.74f)$$

where $L_d = \sqrt{gH}/f$ is the shallow water deformation radius (equation (40.31)), and we introduced the geostrophic streamfunction⁴

$$\psi = (g/f_0) \Delta\eta. \quad (40.75)$$

We also wrote the contribution from topography as

$$\psi_b = (g/f_0) \Delta\eta_b, \quad (40.76)$$

which is a static field. We are thus led to the dimensionful QG potential vorticity for a single shallow water layer

$$q = f_0 (1 + \text{Ro} \hat{q}) = (\zeta + f) - L_d^{-2} (\psi - \psi_b) = f + \psi_b L_d^{-2} + (\nabla^2 - L_d^{-2}) \psi. \quad (40.77)$$

We took the liberty of adding the constant, f_0 , to the QG potential vorticity, which does not alter the dynamics but does allow us to introduce the planetary vorticity, $f = f_0 + \beta y$. The dynamically relevant portions of q appear at order Ro .

40.6.2 Contributions to the QG potential vorticity

The quasi-geostrophic potential vorticity (40.77) has three main contributions

$$q = f + \zeta - f_0 \Delta h/H, \quad (40.78)$$

where we wrote

$$L_d^{-2} (\psi - \psi_b) = f_0^2 / (g H) (g/f_0) (\Delta\eta - \Delta\eta_b) = f_0 \Delta h/H, \quad (40.79)$$

with Δh the undulations in the layer thickness due to undulations in the free surface and bottom topography. Heuristically, we can connect q to the shallow water $Q = (f + \zeta)/h$ by

$$(f + \zeta)/h \approx (f + \zeta) (1 - \Delta h/H) \approx H^{-1} (f + \zeta - f_0 \Delta h/H), \quad (40.80)$$

where we assumed βy , ζ , and Δh are order Ro whereas f_0 is order unity. We pursue this expansion more formally in Section 40.6.4. But for present purposes we can use it to summarize the various contributions to potential vorticity.

- PLANETARY VORTICITY: The contribution $f = f_0 + \beta y$ arises from planetary vorticity, with the dynamically relevant contribution for QG arising just from the βy term.
- GEOSTROPHIC RELATIVE VORTICITY: $\nabla^2 \psi = (g/f_0) \nabla^2 \eta$ is the relative vorticity of the geostrophic flow. Given the Laplacian operator emphasizes small spatial scales, the relative vorticity is most important at scales at or smaller than the deformation radius.
- EFFECTIVE BETA: The contribution from the gradient of planetary vorticity is given by $\beta y = \mathbf{x} \cdot \nabla f$. Likewise, the contribution from topography is given by

$$\psi_b L_d^{-2} = f_0 \Delta\eta_b/H = (f_0/H) (\eta_b - \bar{\eta}_b) \approx (f_0/H) \mathbf{x} \cdot \nabla \eta_b, \quad (40.81)$$

where the approximate expression made use of a Taylor series. These two contributions can be combined into an effective beta

$$\beta y + \psi_b L_d^{-2} = \mathbf{x} \cdot \nabla (f + f_0 \eta_b/H) \approx \mathbf{x} \cdot (H - \eta_b) \nabla [f/(H - \eta_b)], \quad (40.82)$$

with the final approximate expression connecting to the effective beta discussed in Section 35.6.1.

⁴ [Vallis \(2017\)](#) defines the geostrophic streamfunction in his equation (5.69) as $\psi_{\text{vallis}} = g \eta/f_0 = \psi + g H/f_0$, which differs by the constant $g H/f_0$. However, there is no difference in the dynamics since a streamfunction is unique only up to a constant.

- VERTICAL STRETCHING: As discussed in Section 31.2, shallow water fluids move as vertical extensible columns. Hence, the term $(f_0/H)\Delta h$ accounts for the contribution to potential vorticity from column stretching and squashing. It is most important at scales at or larger than those where the relative vorticity is important; i.e., at or larger than the deformation radius.
- FLOW REGIMES RELATIVE TO THE DEFORMATION RADIUS: As noted above, the term $(\nabla^2 - L_d^{-2})\psi$ signals two regimes as determined by the deformation radius. For lateral scales on the order of the deformation radius, both the relative vorticity and vortex stretching make equal contributions to the potential vorticity. For smaller scales relative vorticity is more important whereas for larger scales vortex stretching dominates.
- TWO-DIMENSIONAL TURBULENCE: In the absence of rotation, so that $f_0 = \beta = 0$, the QG potential vorticity reduces to the relative vorticity, $q = \zeta$, in which case quasi-geostrophy reduces to a non-divergent and non-rotating two-dimensional flow. This model is commonly used to study two dimensional turbulence, which can shed light on quasi-geostrophic turbulence (e.g., Section 11.3 of [Vallis \(2017\)](#)).

40.6.3 Summarizing the equations of quasi-geostrophy

The following fields are the key pieces of the single layer shallow water quasi-geostrophic theory

$$\mathbf{u} = \hat{\mathbf{z}} \wedge \nabla \psi \quad \text{and} \quad \zeta = \hat{\mathbf{z}} \cdot \nabla \wedge \mathbf{u} = \nabla^2 \psi \quad \text{and} \quad q = f + \zeta - L_d^{-2}(\psi - \psi_b) \quad (40.83a)$$

$$\psi = (g/f_0) \Delta \eta \quad \text{and} \quad \psi_b = (g/f_0) \Delta \eta_b \quad \text{and} \quad L_d = f_0^{-1} \sqrt{g H}. \quad (40.83b)$$

Evolution of the geostrophic flow

Evolution of the geostrophic flow is determined by material time changes to the QG potential vorticity following the horizontal geostrophic flow

$$\frac{Dq}{Dt} = \frac{\partial q}{\partial t} + \mathbf{u} \cdot \nabla q = \frac{\partial q}{\partial t} + \hat{\mathbf{z}} \cdot (\nabla \psi \wedge \nabla q) = \frac{\partial q}{\partial t} + J(\psi, q), \quad (40.84)$$

where the final equality introduced the Jacobian operator

$$J(\psi, q) = \partial_x \psi \partial_y q - \partial_y \psi \partial_x q. \quad (40.85)$$

For a perfect fluid, the quasi-geostrophic PV is materially constant, whereas more general flows have forcing and dissipation so that

$$\frac{Dq}{Dt} = \text{forcing} - \text{dissipation}. \quad (40.86)$$

As an algorithm, we first update the potential vorticity to a new time step, then diagnose the streamfunction, $\psi = g \eta / f_0$, by solving the elliptic boundary value problem

$$(\nabla^2 - L_d^{-2})\psi = q - f - \psi_b/L_d^2, \quad (40.87)$$

at which point we know the updated free surface and updated velocity.

Velocity and free surface equations

In the above algorithm we are not concerned with the velocity and free surface equations, but instead we are concerned with inverting the potential vorticity equation (40.87) to find the streamfunction. Even so, it can be of interest to consider these equations, which are determined by the order Ro^1 equations in Section 40.5.4

$$\partial_t \mathbf{u}_g + (\mathbf{u}_g \cdot \nabla) \mathbf{u}_g + (f_0 + \beta y) \hat{\mathbf{z}} \wedge \mathbf{u}_g + f_0 \hat{\mathbf{z}} \wedge \mathbf{u}_{ag} = -g \nabla(\eta_g + \eta_{ag}) \quad (40.88a)$$

$$\partial_t(\eta_g - \eta_b) + (\mathbf{u}_g \cdot \nabla)(\eta_g - \eta_b) = -\nabla \cdot \mathbf{u}_{ag}. \quad (40.88b)$$

In these equations we temporarily introduced subscripts to distinguish the geostrophic velocity and corresponding surface height

$$f_0 \hat{\mathbf{z}} \wedge \mathbf{u}_g = -g \nabla \eta_g \quad (40.89)$$

from the ageostrophic free surface height, η_{ag} , and ageostrophic velocity, \mathbf{u}_{ag} . Since the ageostrophic velocity generally has a nonzero horizontal convergence, but it is three dimensionally non-divergent (since the fluid layer has constant density), then we can identify a corresponding vertical component to the ageostrophic velocity that satisfies

$$\mathbf{v}_{ag} = \mathbf{u}_{ag} + \hat{\mathbf{z}} w_{ag} \quad \text{with} \quad \nabla \cdot \mathbf{v}_{ag} = \partial_x u_{ag} + \partial_y v_{ag} + \partial_z w_{ag} = 0. \quad (40.90)$$

40.6.4 Connecting quasi-geostrophic PV to shallow water PV

We here determine how potential vorticity for shallow water quasi-geostrophy relates to the full shallow water potential vorticity studied in Chapter 35. For that purpose, recall that the potential vorticity for a single layer of shallow water fluid is (Section 35.4)

$$Q = \frac{f + \zeta}{h}, \quad (40.91)$$

where $h = H + \Delta\eta - \Delta\eta_b$ (equation (40.4)) is the thickness of the layer.

To connect to the quasi-geostrophic potential vorticity, we non-dimensionalize the potential vorticity prior to performing an asymptotic expansion to order Ro^1 . For this purpose, use the scaling relations from Section 40.3.7 to write the geostrophic relative vorticity as

$$\zeta = \nabla^2 \psi = (g/f_0) \nabla^2 \Delta\eta = (g \mathcal{H})/(f_0 L^2) \hat{\nabla}^2 \hat{\eta} = (g H \text{Ro})/(f_0 L^2) \hat{\zeta}_0 = f_0 \text{Ro} (L_d/L)^2 \hat{\zeta}_0 \quad (40.92)$$

as well as

$$h = H + H \text{Ro} (\hat{\eta} - \hat{\eta}_b) \quad (40.93a)$$

$$f = f_0 (\hat{f}_0 + \text{Ro} \hat{\beta} \hat{y}). \quad (40.93b)$$

Taking the ratio and expanding to order Ro^1 leads to

$$\frac{\zeta + f}{h} = \frac{f_0}{H} \left[\frac{\widehat{f}_0 + \text{Ro} (L_d/L)^2 \widehat{\zeta}_0 + \text{Ro} \widehat{\beta} \widehat{y}}{1 + \text{Ro} (\widehat{\eta} - \widehat{\eta}_b)} \right] \quad (40.94a)$$

$$\approx \frac{f_0}{H} + \frac{f_0 \text{Ro}}{H} \left[-\widehat{f}_0 (\widehat{\eta} - \widehat{\eta}_b) + (L_d/L)^2 \widehat{\zeta}_0 + \widehat{\beta} \widehat{y} \right] \quad (40.94b)$$

$$= \frac{f_0}{H} + \frac{f_0 \text{Ro}}{H} \left[(L_d/L)^2 \widehat{\zeta}_0 + \widehat{\beta} \widehat{y} - \widehat{f}_0 (\widehat{\eta} - \widehat{\eta}_b) \right] \quad (40.94c)$$

$$= \frac{f_0}{H} + \frac{f_0 \text{Ro}}{H} \left[\frac{\zeta}{f_0 \text{Ro}} + \frac{\beta y}{f_0 \text{Ro}} - \frac{\Delta \eta}{\mathcal{H}} + \frac{\Delta \eta_b}{\mathcal{B}} \right] \quad (40.94d)$$

$$= \frac{1}{H} \left[\zeta + f_0 + \beta y - \frac{f_0 (\Delta \eta - \Delta \eta_b)}{H} \right] \quad (40.94e)$$

$$= \widehat{q} U / (H L) \quad (40.94f)$$

$$= q / H. \quad (40.94g)$$

We are thus led to the relation between the shallow water potential vorticity and the quasi-geostrophic shallow water potential vorticity

$$\frac{f + \zeta}{h} = \frac{q}{H} + \mathcal{O}(\text{Ro}^2). \quad (40.95)$$

As noted in Section 40.5.1, this connection between the potential vorticities only holds when assuming deviations in *both* the bottom topography and free surface scale according to the Rossby number, $\Delta \eta = \mathcal{H} \text{Ro}$ and $\Delta \eta_b = \mathcal{B} \text{Ro}$, thus allowing us to combine $\widehat{\eta}$ and $\widehat{\eta}_b$ in the thickness equation (40.93a).

The identity (40.95) is a consequence of the asymptotic expansion of the velocity and thickness equations. We could alternatively invert our development of quasi-geostrophy by using the quasi-geostrophic potential vorticity as the basis for deriving the governing equations. We pursue that approach to derive the layered quasi-geostrophic equations in Section 40.6.6.

40.6.5 Considering topography to be $\mathcal{O}(\text{Ro}^0)$

We here briefly consider topography to be arbitrarily large, in which case we are no longer ensured asymptotic consistency since we can no longer combine $\Delta \eta$ and $\Delta \eta_b$ into an order Rossby number fluctuation of the layer thickness. Even so, it is not uncommon to examine how an asymptotic theory performs when outside of its formal regime of validity, particularly with the advent of numerical codes to facilitate such studies. In many cases the theories continue to provide provocative, and sometimes physically relevant, information.

To see what happens, consider the quasi-geostrophic PV equation

$$(\partial_t + \mathbf{u} \cdot \nabla) [\zeta + \beta y + L_d^{-2} (\psi_b - \psi)] = 0. \quad (40.96)$$

If ψ_b is order unity whereas the other terms are order Ro , then to leading order the potential vorticity is given by the static term, $L_d^{-2} \psi_b$, so that material conservation of potential vorticity reduces to

$$\mathbf{u} \cdot \nabla \psi_b = 0 \implies \mathbf{u} \cdot \nabla \eta_b = 0. \quad (40.97)$$

This constraint means that the *f*-plane geostrophic flow is constrained to flow along lines of constant topography (isobaths), in which case the geostrophic streamfunction satisfies

$$(f/g) \mathbf{u} \cdot \nabla \eta_b = \hat{\mathbf{z}} \cdot (\nabla \eta \wedge \nabla \eta_b) \equiv J(\eta, \eta_b) = 0. \quad (40.98)$$

The order one bottom topography undulations thus provide a strong constraint on the f -geostrophic flow, making the flow align with the bottom topography and in turn aligning surface height undulations with bottom undulations. We uncovered this constraint in our analysis of topographic form stress in Section 35.9.5.

40.6.6 Two layer quasi-geostrophy

In Section 31.4 we developed the equations for an adiabatic stacked shallow water model. We here specialize those equations to a two-layer quasi-geostrophic model, with extensions to multiple layers following straightforwardly. Rather than pursue the formal asymptotic methods used previously, we here make use of our observation in Section 40.6.4 concerning the connection between shallow water and quasi-geostrophic potential vorticities.

To get started, recall from Section 35.4.6 that the shallow water potential vorticity for an arbitrary layer, labelled by the index k , is given by

$$Q_k = \frac{f + \zeta_k}{h_k}, \quad (40.99)$$

As in Section 40.6.2, we assume βy , ζ_k , and Δh_k scale as Ro , in which case we have the quasi-geostrophic potential vorticity for each layer

$$q_k = f + \zeta_k - f_0 \Delta h_k / H, \quad (40.100)$$

where ζ_k is here the geostrophic relative vorticity.

From our asymptotic analysis earlier in this section, the toughest part of that analysis concerned the derivation of the quasi-geostrophic potential vorticity equation. In the present approach, we already have the potential vorticity for each layer via equation (40.100). What we need is the velocity field to advect it. Again, we know what that velocity is: it is the f -plane geostrophic velocity for each layer. The 2-layer velocity equations are given by equations (31.68a) and (31.68b), with their geostrophic components determined by

$$f_0 \hat{z} \wedge \mathbf{u}_1 = -g \nabla(\Delta \eta_b + \Delta h_1 + \Delta h_2) \quad (40.101a)$$

$$f_0 \hat{z} \wedge \mathbf{u}_2 = -\nabla \left[g_{1/2}^r (\Delta \eta_b + \Delta h_1 + \Delta h_2) + g_{3/2}^r (\Delta \eta_b + \Delta h_2) \right]. \quad (40.101b)$$

In these equations we set the applied atmospheric pressure to a constant, and made use of the reduced gravities at the layer interfaces are

$$g_{1/2}^r = g(\rho_1 - \rho_a)/\rho_1 \approx g \quad \text{and} \quad g_{3/2}^r = g(\rho_2 - \rho_1)/\rho_1, \quad (40.102)$$

with $\rho_a \ll \rho_1$ so that the top interface reduced gravity is well approximated by the full gravity. From equations (40.101a) and (40.101b) we can identify the geostrophic streamfunctions

$$\psi_1 = (g/f_0)(\Delta \eta_b + \Delta h_1 + \Delta h_2) \quad (40.103a)$$

$$\psi_2 = (1/f_0) \left[g_{1/2}^r (\Delta \eta_b + \Delta h_1 + \Delta h_2) + g_{3/2}^r (\Delta \eta_b + \Delta h_2) \right], \quad (40.103b)$$

so that the layer geostrophic velocities are given by

$$\mathbf{u}_1 = \hat{z} \wedge \nabla \psi_1 \quad \text{and} \quad \mathbf{u}_2 = \hat{z} \wedge \nabla \psi_2. \quad (40.104)$$

We thus proceed with the usual quasi-geostrophic algorithm, whereby evolution is determined by the material time changes of the potential vorticity with advection given by the layer geostrophic flow

$$(\partial_t + \mathbf{u}_k \cdot \nabla) q_k = 0. \quad (40.105)$$

40.7 Exercises

EXERCISE 40.1: PV CONSERVATION FOR PLANETARY GEOSTROPHY

Show that the planetary geostrophic equations

$$\mathbf{f} \wedge \mathbf{u} = -g \nabla \eta \quad \text{and} \quad \frac{Dh}{Dt} = -h \nabla \cdot \mathbf{u} \quad \text{with} \quad \eta = \eta_b + h \quad (40.106)$$

are equivalent to

$$\mathbf{f} \wedge \mathbf{u} = -g \nabla \eta \quad \text{and} \quad \frac{DQ}{Dt} = 0 \quad \text{with} \quad Q = \frac{f}{h}. \quad (40.107)$$

This result shows that the shallow water PG equations may be written as an evolution equation for an approximated version of the shallow water potential vorticity, $(f + \zeta)/h \approx f/h$. This limit holds when the Rossby number is small.

EXERCISE 40.2: CONSTRAINTS ON STEADY STATE PLANETARY GEOSTROPHIC FLOW

Consider a shallow water fluid satisfying the planetary geostrophic equations developed in Section 40.4. Assume the flow is in steady state.

- (a) In what manner does potential vorticity conservation constrain the velocity field?
- (b) Consider an initially zonal geostrophic flow. In what direction (poleward or equatorward) will a fluid parcel deviate when encountering a seamount (i.e., a region of relatively shallow depth)?
- (c) Describe the path of the velocity field for the case where the ocean sea surface height undulations, $\Delta\eta$, are far smaller than undulations in the bottom topography, η_b (see Figure 31.1 for notation).
- (d) For the special case of an f -plane, show that the velocity is aligned with isolines of bottom topography.
- (e) For the special case of a flat bottom and non-zero Coriolis parameter, show that there is no meridional geostrophic velocity. That is, the flow is zonally aligned.



Continuously stratified planetary geostrophy

The planetary geostrophic equations consist of linear steady geostrophic balance (commonly also with friction) plus the full nonlinear buoyancy equation. Circulation is thus slaved to the mass field given that the flow is diagnostic whereas buoyancy is prognostic. Given that the momentum equation is linear, there is no turbulence in planetary geostrophy. Instead, the equation set allows the analyst to focus on large-scale flow of a stratified laminar fluid feeling the effects from rapid rotation in the presence of β .

We already encountered many physical properties of planetary geostrophic flow in Chapter 26, where we studied geostrophy, vorticity, thermal wind, and Taylor-Proudman. We also made use of planetary geostrophy when studying western boundary current intensification for a shallow water layer in Section 35.9. In this chapter we fill in certain of the mathematical details left out from those early discussions by deriving, through non-dimensionalization and scale analysis, the planetary geostrophic equations.

As an application of the planetary geostrophic equations, we examine a variety of vorticity budgets formed from this equation set. We start by examining the potential vorticity budget and studying how the impermeability theorem manifests for the planetary geostrophic potential vorticity. We then discuss a variety of vorticity budgets used to describe the depth integrated or depth averaged flow. Each of these two-dimensional vorticity budgets offers insights into how the large-scale flow is constrained by rapid rotation and the β -effect present on a sphere. In particular, these budgets offer insights into how forces and the curl of forces act to generate both vertical flow next to the boundaries as well as meridional flow for the full fluid column. The stratified planetary geostrophic equations form the foundation for many theories of the large-scale ocean circulation. Constraints provided by the vorticity theorems developed in this chapter provide the key reasons for why these theories are so physically compelling.

READER'S GUIDE FOR THIS CHAPTER

We here extend the shallow water discussions from Chapter 40 to more formally derive the continuously stratified planetary geostrophic equations. We make use of stratified geophysical fluid dynamics from Chapters 18 and 26, vorticity and the β -effect from Chapter 36, and potential vorticity from Chapter 38. Physical properties of stratified geostrophic mechanics were considered in Chapter 26, with an understanding of that material assumed here. We follow the non-dimensionalization used to derive the shallow water planetary geostrophic equations in Chapter 40, and further extend this material to continuously stratified quasi-geostrophy in Chapter 42.

41.1	Non-dimensionalizing the stratified Boussinesq equations	936
41.1.1	Dimensional parameters	937
41.1.2	Physical dimensions and non-dimensional parameters	938
41.1.3	Choosing the non-dimensional parameters	938
41.1.4	Relating the buoyancy scale to the Coriolis acceleration scale	939
41.1.5	Richardson number and QG/PG flow regimes	939
41.1.6	The Rossby deformation radius	940
41.1.7	Assumed values for the non-dimensional parameters	941
41.1.8	Non-dimensional Boussinesq equations	942
41.2	Summary equations for planetary geostrophy	943
41.3	Planetary geostrophic potential vorticity	944
41.3.1	Derivation	944
41.3.2	Impermeability theorem	945
41.3.3	Kinematic fluxes satisfying impermeability	946
41.4	Depth integrated vorticity budget	946
41.4.1	The β -effect, stretching, and meridional transport	947
41.4.2	Bottom kinematics and dynamics	948
41.4.3	Surface kinematics and dynamics	950
41.4.4	Summary of force curls driving depth integrated meridional flow	951
41.5	Vorticity of the depth integrated velocity	953
41.5.1	Depth integrated velocity equation	953
41.5.2	Vorticity budget	954
41.6	Vorticity equation for the depth averaged velocity	954
41.6.1	Relating the depth average velocity to boundary velocities	954
41.6.2	Formulation of the vorticity equation	955
41.6.3	Rigid lid approximation and the role of JEBAR	956
41.6.4	Further study	957

41.1 Non-dimensionalizing the stratified Boussinesq equations

In this section we non-dimensionalize the continuously stratified Boussinesq equations. As part of this process we identify a variety of non-dimensional numbers that characterize the flow. Planetary geostrophy is a rather simple asymptotic theory. Nonetheless, our work to non-dimensionalize the Boussinesq equations supports the work needed for the somewhat more complicated derivations of continuously stratified quasi-geostrophy in Chapter 42.

The starting point is the perfect fluid stratified hydrostatic Boussinesq equations (Section 24.1)

$$\frac{D\mathbf{u}}{Dt} + f \hat{\mathbf{z}} \wedge \mathbf{u} = -\nabla_z \varphi \quad (41.1a)$$

$$\frac{\partial \varphi}{\partial z} = b \quad (41.1b)$$

$$\frac{Db}{Dt} = 0 \quad (41.1c)$$

$$\nabla \cdot \mathbf{v} = 0, \quad (41.1d)$$

where $\mathbf{v} = (\mathbf{u}, w)$ is the three-dimensional velocity written using Cartesian coordinates, $b = -g(\rho - \rho_0)/\rho_0$ is the buoyancy relative to a constant reference density, ρ_0 with ρ the density. We also write $\varphi = \delta p/\rho_0$ for the dynamic pressure (dimensions of $L^2 T^{-2}$), and $\nabla_z = (\partial_x, \partial_y, 0)$ for the horizontal gradient operator. We separate a background vertical buoyancy profile from a space-time

fluctuating buoyancy

$$b = \tilde{b}(z) + b'(x, y, z, t), \quad (41.2)$$

and introduce the corresponding background squared buoyancy frequency

$$N^2 = \frac{\partial \tilde{b}}{\partial z}. \quad (41.3)$$

With this decomposition, the buoyancy equation (41.1c) takes the form

$$\frac{Db'}{Dt} + w N^2 = 0. \quad (41.4)$$

We also introduce an associated decomposition of the hydrostatic pressure

$$\varphi = \tilde{\varphi}(z) + \varphi'(x, y, z, t) \quad (41.5)$$

where $\tilde{\varphi}$ is hydrostatically balanced by \tilde{b}

$$\frac{d\tilde{\varphi}}{dz} = \tilde{b}, \quad (41.6)$$

and the fluctuating pressure, φ' , is hydrostatically balanced by b'

$$\frac{\partial \varphi'}{\partial z} = b'. \quad (41.7)$$

41.1.1 Dimensional parameters

Following the shallow water discussion in Section 40.3.1, we have the following dimensional parameters for the perfect Boussinesq fluid.

- LENGTH SCALES
 - ★ H = depth scale of a typical vertical structure in the fluid (e.g., the depth of the thermocline).
 - ★ L = horizontal/lateral length scale of motions under consideration.
 - ★ R = radius of the planet.
- VELOCITY SCALES
 - ★ U = horizontal velocity scale for fluid parcel motion.
 - ★ W = vertical velocity scale for fluid parcel motion.
- PRESSURE AND BUOYANCY SCALES: Pressure is a contact force, acting on the boundary of an arbitrary fluid region, and buoyancy arises from the gravitational force that acts to raise or lower a fluid parcel depending on its density relative to the environment. They have scales given by the following.
 - ★ Φ = scale for pressure fluctuations φ' (dimensions of pressure divided by density = length scale \times acceleration).
 - ★ B = scale of buoyancy fluctuations b' (dimensions of acceleration).
- BODY FORCES: There are two body forces acting on the fluid, one from gravity and one from Coriolis.

- ★ g = gravitational acceleration
- ★ f = Coriolis frequency

Contrary to the shallow water discussion in Section 40.3.1, we do not introduce a wave speed since it does not affect the asymptotics. We also dropped the bottom topography scale, assuming it is small for present purposes.

41.1.2 Physical dimensions and non-dimensional parameters

There are two physical dimensions in the Boussinesq system: length, L , and time, T . As for the shallow water system, there no mass since mass is determined by the density (buoyancy) and volume. The Buckingham-II theorem then says there are

$$N_{\text{dimensionless}} = 9 - 2 = 7 \quad (41.8)$$

non-dimensional parameters.

41.1.3 Choosing the non-dimensional parameters

Following the shallow water discussion in Section 40.3.4, we choose the following non-dimensional parameters.

1. VERTICAL TO HORIZONTAL ASPECT RATIO: The ratio of the vertical scale to the horizontal scale defines the aspect ratio

$$\delta_{\text{vertical/horizontal}} = \frac{\text{vertical depth scale}}{\text{horizontal length scale}} = \frac{H}{L}. \quad (41.9)$$

2. RATIO OF HORIZONTAL SCALE TO PLANETARY SCALE: The ratio of the lateral length scale to the planet radius is given by

$$\delta_{\text{horizontal/planet}} = \frac{\text{lateral length scale}}{\text{planetary length scale}} = \frac{L}{R}. \quad (41.10)$$

3. RATIO VERTICAL TO HORIZONTAL VELOCITY SCALE: The ratio of the vertical to horizontal velocity is given by

$$\frac{\text{vertical velocity scale}}{\text{horizontal velocity scale}} = \frac{W}{U}. \quad (41.11)$$

4. HYDROSTATIC NUMBER: The hydrostatic number is the ratio of the pressure gradient scale to the buoyancy scale. For the hydrostatic fluid fluctuations

$$\frac{\Phi}{H} = B, \quad (41.12)$$

where B is the scale for the buoyancy fluctuations.

5. ROSSBY NUMBER: The Rossby number is the ratio of the fluid parcel acceleration scale to the Coriolis acceleration

$$\text{Ro} = \frac{\text{parcel acceleration}}{\text{Coriolis acceleration}} = \frac{U}{f L}, \quad (41.13)$$

where we again assume time scales advectively

$$T \sim \frac{L}{U} \implies \text{Ro} = \frac{U}{f L} = \frac{1}{f T}. \quad (41.14)$$

6. GEOSTROPHIC NUMBER: The ratio of the Coriolis acceleration to the pressure gradient acceleration defines the geostrophic number

$$\text{Ge} = \frac{\text{Coriolis acceleration}}{\text{pressure gradient acceleration}}. \quad (41.15)$$

The Coriolis acceleration scales as

$$\text{Coriolis acceleration} \sim f U \quad (41.16)$$

whereas the pressure gradient acceleration from the fluctuating pressure, φ' , scales as

$$\text{pressure gradient acceleration} \sim \frac{\Phi}{L}, \quad (41.17)$$

so that

$$\text{Ge} = \frac{\text{Coriolis acceleration}}{\text{pressure gradient acceleration}} = \frac{f U}{(\Phi/L)}. \quad (41.18)$$

7. RATIO OF FLUCTUATING STRATIFICATION TO BACKGROUND STRATIFICATION: The ratio of the buoyancy frequency arising from the fluctuating buoyancy, B/H , to the background buoyancy frequency, N^2 , is given by

$$\frac{\text{fluctuating buoyancy frequency}}{\text{background buoyancy frequency}} = \frac{B/H}{N^2}. \quad (41.19)$$

41.1.4 Relating the buoyancy scale to the Coriolis acceleration scale

The fluctuating buoyancy (b') and fluctuating pressure (φ') have scales related through the hydrostatic balance. From equation (41.12) we have

$$B = \frac{\Phi}{H}. \quad (41.20)$$

Additionally, assuming geostrophic scaling, equation (41.18) means that the fluctuating pressure has a scale related to the Coriolis acceleration scale according to

$$\Phi = f U L. \quad (41.21)$$

Hence, the scale for the fluctuating buoyancy is given by

$$B = \frac{f U L}{H}. \quad (41.22)$$

41.1.5 Richardson number and QG/PG flow regimes

The *Richardson number* is the non-dimensional ratio of the squared buoyancy frequency to the squared vertical shear of the horizontal velocity

$$\text{Ri} = \frac{N^2}{|\partial_z \mathbf{u}|^2}. \quad (41.23)$$

In regions where $\text{Ri} \ll 1$, the vertical shear is stronger than the stabilizing effects from vertical stratification. In regions with small Richardson numbers, there is enough kinetic energy in the vertical shear to extract potential energy from the stratification, and this extraction process occurs

via a dynamical instability known as *Kelvin-Helmholtz instability*. In contrast, for large-scale highly stratified flow, the Richardson number is quite large, with $\text{Ri} \sim 100$ common. Large Richardson number flow regimes are where quasi-geostrophy is relevant (Chapter 42).

In our choice for dimensionless parameters, we could choose one determined by the scale for the Richardson number

$$\text{Ri} = \frac{N^2}{(U/H)^2}, \quad (41.24)$$

where we set the vertical scale equal to H , the horizontal velocity scale to U , and the squared buoyancy frequency to a scale N^2 . Alternatively, the Richardson number can be related to the Rossby and Burger numbers through

$$\text{Bu} = \left[\frac{L_d}{L} \right]^2 = \left[\frac{N H}{f L} \right]^2 = \frac{U^2 \text{Ri}}{U^2 / (\text{Ro})^2} = (\text{Ro})^2 \text{Ri}. \quad (41.25)$$

For QG flows, the horizontal length scales, L , are assumed to be on the order of the deformation radius, L_d , in which case the Burger number is close to unity. The relation (41.25) thus means that the Richardson number scales as

$$\text{Ri} \sim (\text{Ro})^{-2} \quad \text{QG flow regime.} \quad (41.26)$$

For atmospheric flows with a Rossby number order 1/10, QG flow regimes are realized with a Richardson number ~ 100 . For the ocean, the Rossby number can be even smaller, in which case QG flows are characterized by an even larger Richardson number. For planetary geostrophy, the Burger number is small, in which case PG flows are characterized by somewhat smaller Richardson numbers than QG flows.

41.1.6 The Rossby deformation radius

The combined effects of buoyancy and rotation yield the richness of continuously stratified planetary geostrophic and quasi-geostrophic motions. Hence, the buoyancy frequency and the Coriolis parameter play central role in characterizing these flow regimes. The ratio of these two frequencies, N/f , in regions of nontrivial vertical stratification is typically around 100. Hence, rotational inertial oscillations (usually just called *inertial oscillations*) have about 100 times longer period, $2\pi/f$, than buoyancy oscillations with period $2\pi/N$.

Letting the squared buoyancy frequency N^2 refer to a value typical of a particular flow regime, one can define the Rossby deformation radius

$$L_d = H N/f. \quad (41.27)$$

As defined, the deformation radius is the vertical length scale multiplied by the ratio of the buoyancy frequency to the Coriolis frequency. The ratio f/N appears frequently in rotating/stratified fluids, and is sometimes called the Prandtl ratio

$$\frac{f}{N} = \text{Prandtl ratio.} \quad (41.28)$$

With $H \approx 1$ km and $N/f \approx 100$, the Rossby deformation radius is roughly 100 km. This length scale measures the relative importance of stratification and rotation. Depending on the ratio L/L_d , we can have large or small stratification fluctuations relative to the background stratification. Furthermore, it sets the scale for unstable baroclinic waves leading to baroclinically unstable flow (see Chapter 6 of [Vallis \(2017\)](#)). For some context, recall the shallow water deformation radius is

given by equation (40.31), $L_d = \sqrt{gH}/f$, which is the ratio of the gravity wave speed to Coriolis frequency. With $N = 100 f = 10^{-2} \text{ s}^{-1}$ and $H = 10^3 \text{ m}$, the shallow water deformation radius is about an order of magnitude larger than the internal deformation radius. This means that the characteristic length scales, as set by L_d , are much larger in a shallow water fluid than in a stratified fluid.

41.1.7 Assumed values for the non-dimensional parameters

We now enumerate the assumed values for the non-dimensional parameters, again largely following the choices made for the shallow water in Section 40.3.5. These assumptions are guided by the regimes of interest to our analysis. In turn, the assumptions limit the validity of the resulting equations to those regimes.

1. SMALL VERTICAL TO HORIZONTAL ASPECT RATIO: The aspect ratio is generally small for large-scale atmospheric and oceanic fluid systems

$$\delta_{\text{vertical/horizontal}} \ll 1. \quad (41.29)$$

This assumption is part of the hydrostatic approximation (Section 23.3).

2. SMALL OR ORDER ONE RATIO OF HORIZONTAL TO PLANETARY SCALES: The ratio of the lateral length scale to the planet radius is small for quasi-geostrophic systems, and order unity for planetary geostrophy

$$\delta_{\text{horizontal/planet}} \ll 1 \quad \text{quasi-geostrophy} \quad (41.30a)$$

$$\delta_{\text{horizontal/planet}} \sim 1 \quad \text{planetary geostrophy}. \quad (41.30b)$$

3. SMALL RATIO VERTICAL TO HORIZONTAL VELOCITY SCALE: The continuity equation implies

$$\frac{W}{H} = \frac{U}{L}, \quad (41.31)$$

so that

$$W = U \frac{H}{L}. \quad (41.32)$$

As noted above, for a hydrostatic fluid the vertical to horizontal aspect ratio H/L is small, so that the vertical velocity scale is smaller than the horizontal velocity scale. Furthermore, when the fluid is close to geostrophically balanced, the vertical velocity scale is even smaller, by a factor of Ro . We will see that factor emerge in the following scale analysis.

4. UNIT HYDROSTATIC NUMBER: As already noted, the hydrostatic balance (41.1b) means that the scales for a buoyancy fluctuation and pressure fluctuation are related by (see equation (41.12))

$$\Phi = H B. \quad (41.33)$$

5. SMALL ROSSBY NUMBER: The Rossby number is assumed small

$$\text{Ro} = \frac{U}{f L} = \frac{1}{f T} \ll 1, \quad (41.34)$$

where we set the time scale according to advection, $T = L/U$.

6. UNIT GEOSTROPHIC NUMBER: The geostrophic number is assumed to be order unity

$$\text{Ge} \sim 1, \quad (41.35)$$

which means that the Coriolis acceleration and pressure gradient acceleration scale together

$$f U \sim \frac{\Phi}{L}. \quad (41.36)$$

This scaling is consistent with the momentum equation (41.1a) so long as the Rossby number is small, $\text{Ro} \ll 1$.

7. STRATIFICATION FLUCTUATIONS COMPARED TO BACKGROUND STRATIFICATION: Making use of the assumed unit geostrophic number, the ratio of the buoyancy frequency arising from the fluctuating buoyancy to the background buoyancy frequency is given by

$$\frac{B/H}{N^2} = \frac{\Phi}{H^2 N^2} = \frac{f U L}{H^2 N^2} = \frac{U}{f L} \frac{L^2 f^2}{H^2 N^2} = \text{Ro} \frac{L^2}{L_d^2}, \quad (41.37)$$

where we introduced the deformation radius (41.27) $L_d = H N/f$. It is important to keep the depth dependence of N^2 when returning to dimensional fields, particularly for the QG system in Section 42.3.7.

41.1.8 Non-dimensional Boussinesq equations

Following the shallow water approach in Section 40.3.7, we introduce non-dimensional variables according to

$$t = T \hat{t} \quad (x, y) = L (\hat{x}, \hat{y}) \quad \partial_t = \partial_{\hat{t}}/T \quad \nabla_z = \hat{\nabla}_z/L \quad \partial_z = \partial_{\hat{z}}/H \quad f = f_0 \hat{f} \quad (41.38a)$$

$$(u, v) = U (\hat{u}, \hat{v}) \quad w = W \hat{w} \quad \varphi' = f_0 U L \hat{\varphi} \quad b' = B \hat{b} = (f_0 U L / H) \hat{b}. \quad (41.38b)$$

For the second equality in the buoyancy scale, we made use of equation (41.22) to connect the buoyancy fluctuation scale to the Coriolis acceleration scale. We also make use of the following relations between scales

$$T = L/U \quad \text{advective scaling for } T \quad (41.39)$$

$$W = U H/L \quad \text{continuity scaling for } W \quad (41.40)$$

$$\text{Ro} = U/(f_0 L) = (T f_0)^{-1} \quad \text{advective scaling for } T. \quad (41.41)$$

The first relation assumes the time scale is determined by the advection time $T = L/U$, which then means that the Rossby number is the ratio of the advective frequency $1/T$ to the Coriolis frequency f_0 . Furthermore, we assume vertical velocity scales according to the continuity equation, $W = U (H/L)$. This continuity scaling for W will be seen to be an over-estimate in the following, where we find that W instead scales like $W = \text{Ro} U (H/L)$.

Non-dimensional momentum equation

Introducing the above dimensionless variables and dimensionful scales into the Boussinesq momentum equation (41.1a) renders

$$\frac{U}{T} \frac{\partial \hat{\mathbf{u}}}{\partial \hat{t}} + \frac{U^2}{L} (\hat{\mathbf{u}} \cdot \hat{\nabla}_z) \hat{\mathbf{u}} + \frac{W U}{H} \hat{\mathbf{w}} \frac{\partial \hat{\mathbf{u}}}{\partial \hat{z}} + f_0 U (\hat{\mathbf{f}} \wedge \hat{\mathbf{u}}) = -f_0 U \hat{\nabla}_z \hat{\varphi}. \quad (41.42)$$

Hence, dividing by $f_0 U$ leads to

$$\text{Ro} \left[\frac{\partial \hat{\mathbf{u}}}{\partial \hat{t}} + (\hat{\mathbf{u}} \cdot \hat{\nabla}_{\hat{z}}) \hat{\mathbf{u}} + \hat{w} \frac{\partial \hat{\mathbf{u}}}{\partial \hat{z}} \right] + (\hat{\mathbf{f}} \wedge \hat{\mathbf{u}}) = -\hat{\nabla}_z \hat{\varphi}. \quad (41.43)$$

The momentum equation is consistent with a unit geostrophy number (i.e., Coriolis acceleration balances pressure gradient acceleration) if and only if the Rossby number is small, thus eliminating the acceleration of a fluid parcel. Likewise, the non-dimensional hydrostatic balance is given by

$$\frac{\partial \hat{\varphi}}{\partial \hat{z}} = \hat{b}, \quad (41.44)$$

and the non-dimensional continuity equation is

$$\hat{\nabla} \cdot \hat{\mathbf{v}} = 0. \quad (41.45)$$

Non-dimensional buoyancy equation

The buoyancy equation (41.4) requires a bit more work to non-dimensionalize. The material time derivative takes the form

$$\frac{D b'}{D t} = \frac{B}{T} \frac{D \hat{b}}{D \hat{t}} = \frac{U}{L} \frac{f_0 U L}{H} \frac{D \hat{b}}{D \hat{t}} = \frac{f_0 U^2}{H} \frac{D \hat{b}}{D \hat{t}}, \quad (41.46)$$

where we made use of the advective scaling $T = L/U$ and continuity scaling $W = U(H/L)$. The vertical advection of background stratification is given by

$$N^2 w = N^2 W \hat{w} = N^2 U (H/L) \hat{w} = L_d^2 \frac{U f_0^2}{H L} \hat{w}, \quad (41.47)$$

where we introduced the deformation radius, $L_d = HN/f$, from equation (41.27). Bringing these two pieces together leads to

$$\text{Ro} \frac{D \hat{b}}{D \hat{t}} + \text{Bu} \hat{w} = 0, \quad (41.48)$$

where we introduced the Burger number $\text{Bu} = (L_d/L)^2$.

41.2 Summary equations for planetary geostrophy

Just like for the shallow water model in Section 40.4, the planetary geostrophic system for the stratified Boussinesq system is a rather simple asymptotic theory. For this case, we assume the horizontal scales are large compared to the deformation radius, so that

$$\text{Ro}/\text{Bu} \sim 1 \implies \text{Ro} L^2 \sim L_d^2. \quad (41.49)$$

With this scaling, and with a small Rossby number, the momentum equation (41.43) reduces to geostrophic balance. However, the continuity and buoyancy equations retain their unapproximated Boussinesq form. Hence, in dimensional form, the perfect planetary geostrophic equations for a stratified Boussinesq fluid take the form

$$\frac{D b'}{D t} + w N^2 = 0 \quad (41.50a)$$

$$f \hat{z} \wedge \mathbf{u} = -\nabla_z \varphi' \quad (41.50b)$$

$$\frac{\partial \varphi'}{\partial z} = b' \quad (41.50c)$$

$$\nabla \cdot \mathbf{v} = 0. \quad (41.50d)$$

We could just as well write the above planetary geostrophic equations in terms of the full buoyancy $b = \tilde{b}(z) + b'$, and full pressure, $p = p_0(z) + \rho_0 \varphi$. Additionally, we can include non-conservative terms such as buoyancy mixing, \dot{b} , as well as frictional accelerations and/or boundary forcing, \mathbf{F} . The following form for the frictional and diabatic planetary geostrophic flow will form the basis for many discussions in this chapter

$$f \hat{\mathbf{z}} \wedge \mathbf{u} = -\rho_0^{-1} \nabla_z p + \mathbf{F} \quad \text{frictional geostrophy} \quad (41.51a)$$

$$\frac{\partial b}{\partial t} + \mathbf{u} \cdot \nabla_z b + N^2 w = \dot{b} \quad \text{diabatic buoyancy equation} \quad (41.51b)$$

$$\frac{\partial p}{\partial z} = -\rho g \quad \text{hydrostatic} \quad (41.51c)$$

$$\nabla \cdot \mathbf{v} = \nabla_z \cdot \mathbf{u} + \frac{\partial w}{\partial z} = 0 \quad \text{incompressible flow} \quad (41.51d)$$

$$N^2 = \frac{\partial b}{\partial z} \quad \text{squared buoyancy frequency.} \quad (41.51e)$$

We will also generally assume an equation of state that is independent of pressure, so that changes in buoyancy arise only from changes in potential temperature and/or salinity

$$\dot{b} = \frac{\partial b}{\partial S} \dot{S} + \frac{\partial b}{\partial \Theta} \dot{\Theta}. \quad (41.52)$$

The buoyancy partial derivatives, $\partial b / \partial \Theta$ and $\partial b / \partial S$, will sometimes be assumed constant.

Note that the material time derivative in PG makes use of advection by the three velocity components, $\mathbf{v} = (\mathbf{u}, w)$, as seen in the buoyancy equation (41.51b), with the horizontal velocity components determined by the frictional geostrophic balance (41.50b). This situation contrasts to the case of QG, where it is only the horizontal advection by the geostrophic flow that contributes to material time evolution (Section 42.2).

41.3 Planetary geostrophic potential vorticity

In Section 38.5 we developed the potential vorticity equation for the hydrostatic Boussinesq system in the presence of horizontal friction in the momentum equation and diabatic terms in the buoyancy equation. Here we specialize that result to the case of planetary geostrophic system written in the form of equations (41.51a)-(41.51e).

41.3.1 Derivation

Derivation of the PV equation proceeds much like that for the hydrostatic Boussinesq system. The first step requires the PG vorticity equation as determined by taking the curl of the momentum equation (41.51a). The vertical component of this vorticity equation is given by (see also Section 26.5.2)

$$\beta v = f \frac{\partial w}{\partial z} + \hat{\mathbf{z}} \cdot (\nabla_z \wedge \mathbf{F}) \quad \text{with} \quad \beta = \frac{\partial f}{\partial y}. \quad (41.53)$$

Next make use of the thermal wind relation in the presence of friction

$$f \frac{\partial \mathbf{u}}{\partial z} = \hat{\mathbf{z}} \wedge \nabla_z b - \frac{\partial (\hat{\mathbf{z}} \wedge \mathbf{F})}{\partial z} \quad (41.54)$$

as well as the identities

$$N^2 \frac{Df}{Dt} = N^2 \beta v \quad (41.55a)$$

$$f \frac{DN^2}{Dt} = f \frac{\partial \dot{b}}{\partial z} - f \nabla b \cdot \frac{\partial \mathbf{v}}{\partial z} \quad (41.55b)$$

$$f \nabla b \cdot \frac{\partial \mathbf{v}}{\partial z} = f N^2 \frac{\partial w}{\partial z} - \frac{\partial (\hat{\mathbf{z}} \wedge \mathbf{F})}{\partial z} \cdot \nabla_z b, \quad (41.55c)$$

to render

$$\frac{D(f N^2)}{Dt} = N^2 \beta v + f \frac{\partial \dot{b}}{\partial z} - f \nabla b \cdot \frac{\partial \mathbf{v}}{\partial z} \quad (41.56a)$$

$$= N^2 \left[f \frac{\partial w}{\partial z} + \hat{\mathbf{z}} \cdot (\nabla_z \wedge \mathbf{F}) \right] + f \frac{\partial \dot{b}}{\partial z} - f N^2 \frac{\partial w}{\partial z} + \frac{\partial (\hat{\mathbf{z}} \wedge \mathbf{F})}{\partial z} \cdot \nabla_z b \quad (41.56b)$$

$$= f \frac{\partial \dot{b}}{\partial z} + \nabla b \cdot (\nabla \wedge \mathbf{F}) \quad (41.56c)$$

$$= \nabla \cdot (f \dot{b} \hat{\mathbf{z}} + \mathbf{F} \wedge \nabla b). \quad (41.56d)$$

We thus identify the planetary geostrophic potential vorticity

$$Q_{PG} = f N^2, \quad (41.57)$$

which is materially invariant in the absence of diabatic processes and friction

$$\frac{DQ_{PG}}{Dt} = 0 \quad \text{when } \dot{b} = 0 \text{ and } \mathbf{F} = 0. \quad (41.58)$$

We can write the general budget equation in the form of an Eulerian flux-form expression

$$\frac{\partial Q_{PG}}{\partial t} + \nabla \cdot \mathbf{J}_{PG} = 0, \quad (41.59)$$

where the potential vorticity flux is given by

$$\mathbf{J}_{PG} = \mathbf{v} Q_{PG} - \dot{b} f \hat{\mathbf{z}} + \nabla b \wedge \mathbf{F} + \nabla \wedge \mathbf{A}. \quad (41.60)$$

The vector \mathbf{A} is an arbitrary gauge function that has no impact on the potential vorticity evolution. Comparing to the hydrostatic Boussinesq case from Section 38.5.2, we see that the planetary geostrophic result follows by dropping contributions from the relative vorticity and dropping the inertial acceleration.

41.3.2 Impermeability theorem

Following the discussion in Section 39.1.2, we here verify that the potential vorticity flux vector (41.60) satisfies the impermeability theorem for buoyancy isosurfaces. We do so here for the particular case of a zero gauge function ($\mathbf{A} = 0$), in which case

$$\mathbf{v}_{PG} \cdot \nabla b = \frac{\mathbf{J}_{PG} \cdot \nabla b}{Q} = \mathbf{v} \cdot \nabla b - \dot{b} = -\frac{\partial b}{\partial t}, \quad (41.61)$$

reveals that

$$\frac{\partial b}{\partial t} + \mathbf{v}_{PG} \cdot \nabla b = 0. \quad (41.62)$$

Hence, there is zero flux of PV-substance crossing buoyancy isosurfaces, even in the presence of irreversible processes that allow matter and heat to cross those surfaces. As shown in the next subsection, we identify two more forms of the PV-substance flux vector that also satisfy impermeability, with these alternative forms differing by gauge transformations.

41.3.3 Kinematic fluxes satisfying impermeability

Following the discussion of impermeability for the Ertel PV in Section 39.1.2, we expose a purely kinematic means to derive the impermeability theorem for the PG potential vorticity. This derivation follows by computing the time tendency of the PV

$$\frac{\partial Q}{\partial t} = \frac{\partial}{\partial t} \nabla \cdot (f b \hat{z}) = \nabla \cdot \left[f \frac{\partial b}{\partial t} \hat{z} \right] \equiv -\nabla \cdot \tilde{\mathbf{J}}_{\text{PG}}, \quad (41.63)$$

where

$$\tilde{\mathbf{J}}_{\text{PG}} = -f \frac{\partial b}{\partial t} \hat{z}. \quad (41.64)$$

This form of the PV-substance flux also satisfies impermeability since

$$\tilde{\mathbf{v}}_{\text{PG}} \cdot \nabla b = \frac{\tilde{\mathbf{J}}_{\text{PG}} \cdot \nabla b}{Q} = -\frac{\partial b}{\partial t}, \quad (41.65)$$

so that

$$\frac{\partial b}{\partial t} + \tilde{\mathbf{v}}_{\text{PG}} \cdot \nabla b = 0. \quad (41.66)$$

The flux $\tilde{\mathbf{J}}_{\text{PG}}$ vanishes in the steady state, whereas the steady state form of \mathbf{J}_{PG} does not vanish. Following the discussion in Section 39.2.2, we may choose to introduce a gauge transformation to the kinematic flux $\tilde{\mathbf{J}}_{\text{PG}}$ so that it does not vanish in the steady state. Taking the small Rossby number limit of the flux (39.35) renders

$$\mathbf{J}_{\text{Q}}^{\text{marshall PG}} = -\nabla(g z + \varphi) \wedge \nabla b - f \frac{\partial b}{\partial t} \hat{z}. \quad (41.67)$$

This flux differs from $\tilde{\mathbf{J}}_{\text{PG}}$ by a curl

$$\nabla(g z + \varphi) \wedge \nabla b = \nabla \wedge [(g z + \varphi) \nabla b], \quad (41.68)$$

and it also satisfies the impermeability theorem. As discussed in Section 39.4.6, there are a variety of motivations for using one form of the PV-substance flux versus another. Some applications prefer a nonzero steady flux that also does not involve any irreversible processes, with $\mathbf{J}_{\text{Q}}^{\text{marshall PG}}$ satisfying these desires.

41.4 Depth integrated vorticity budget

In a planetary geostrophic flow, vorticity arises just from planetary rotation since relative vorticity is assumed to be negligible. With planetary vorticity a function just of latitude, a budget for the planetary geostrophic vorticity reveals how the curl of forces imparted to the fluid cause meridional motion as the fluid meets the constraints imposed by the vorticity equation. As per our discussion in Section 36.4.4, we commonly refer to force curls as a “torque” in our study of vorticity sources, with that terminology common in the study of vorticity budgets. However, one must keep in mind that more common usage in physics refers to a torque as affecting changes to angular momentum, with angular momentum generally distinct from vorticity (see Chapter 37 for a discussion of the distinction).

In this section we study the depth integrated vorticity budget for the planetary geostrophic fluid and derive implications for the meridional flow. For this purpose we focus on frictional and boundary accelerations that take the form of a vertical divergence of horizontal turbulent stresses

$$\mathbf{F} = \frac{\partial \boldsymbol{\tau}}{\partial z}. \quad (41.69)$$

The curl of this stress, as well as pressure forces, provide sources (“torques”) that generate meridional motion as revealed by the planetary geostrophic vorticity budget.

41.4.1 The β -effect, stretching, and meridional transport

In Section 26.5.2 we derived the vorticity equation for the planetary geostrophic system. We also encountered this equation when deriving the potential vorticity budget in Section 41.3.1. With friction written as a vertical divergence of horizontal turbulent stresses (equation (41.69)), the vertical component of the PG vorticity equation takes the form

$$\rho_0 \beta v = \frac{\partial}{\partial z} [\rho_0 f w + \hat{z} \cdot (\nabla_z \wedge \boldsymbol{\tau})]. \quad (41.70)$$

Vertical integration from the ocean bottom at $z = \eta_b(x, y)$ to sea surface at $z = \eta(x, y, t)$ leads to

$$\rho_0 \beta V = \underbrace{\rho_0 f [w(\eta) - w(\eta_b)]}_{\text{column stretching}} + \underbrace{\hat{z} \cdot (\nabla_z \wedge \Delta \boldsymbol{\tau})}_{\text{boundary stresses}}, \quad (41.71)$$

where

$$\Delta \boldsymbol{\tau} = \boldsymbol{\tau}(\eta) - \boldsymbol{\tau}(\eta_b) \quad (41.72)$$

is the difference in boundary stresses applied at the ocean surface and ocean bottom, and

$$V = \int_{\eta_b}^{\eta} v \, dz \quad (41.73)$$

is the depth-integrated meridional flow. For a planetary geostrophic flow, absolute vorticity is approximated by just the planetary vorticity: $\zeta_a = \zeta + f \approx f$. As revealed by the vorticity equation (41.71), vorticity sources in a planetary geostrophic fluid lead to meridional motion since that is the only way the fluid can modify its vorticity in response to the sources.

The first term on the right hand side of equation (41.71) arises from vertical stretching of the depth integrated column as measured by differences in the vertical velocity at the ocean surface and bottom. For example, vertical stretching caused by positive surface velocity, $w(\eta) > 0$, or a negative bottom velocity, $w(\eta_b) < 0$, lead to poleward motion of the fluid column. Conversely, vertical squashing leads to equatorward motion. We emphasize that the vertical stretching term arises from the difference in the boundary vertical velocity rather than the vertical velocity within the interior of the fluid.

The second term in equation (41.71) arises from differences in the vorticity imparted by surface and bottom boundary stresses. Positive vorticity imparted to the fluid through the curl of wind stresses, $\hat{z} \cdot [\nabla_z \wedge \boldsymbol{\tau}(\eta)] > 0$, or by the curl of bottom stresses, $\hat{z} \cdot [\nabla_z \wedge \boldsymbol{\tau}(\eta_b)] < 0$, lead to poleward motion of the fluid column, with opposite motion for the converse curls.

The β -effect (Section 36.7.2) is a fundamental element of the depth integrated vorticity equation (41.71). Namely, as fluid columns are stretched or squashed, they must move meridionally to maintain vorticity balance for a planetary geostrophic fluid on a rotating spherical earth. The planetary geostrophic vorticity equation restricts attention to vertical stretching through vertical motion (the w terms) and through the vorticity imparted by the curl of boundary stresses. Notably, the curl of boundary stresses also imparts vertical motion through surface and bottom Ekman layer dynamics (see Chapter 28). Hence, the right hand side of the vorticity equation (41.71) is fundamentally related to vortex stretching.

Equation (41.71) is central to mechanical descriptions of large-scale ocean circulation. For many flow regimes, the curl of the surface wind stress dominates, thus allowing us to ignore the vertical

velocity terms as well as bottom frictional stresses. Formally, we isolate the wind stress when studying a flat bottom rigid lid model, whereby $w(\eta) = w(\eta_b) = 0$. However, there can be nontrivial impacts from bottom pressure torques when flow interacts with sloping topography, with the North Atlantic and Southern Ocean providing important case studies. Other processes can be important in various flow regimes, thus prompting us to derive a full diagnostic framework to identify where these processes are important. To pursue that framework, we make use of the kinematic boundary conditions and the horizontal momentum equation to unpack the vertical velocity terms. Doing so reveals the forces and their curls that drive vertical motion at the boundaries for a planetary geostrophic flow. By doing so, we capture how these force curls contribute, through the β -effect, to meridional motion of the depth integrated flow.

41.4.2 Bottom kinematics and dynamics

The bottom kinematic boundary condition applied at $z = \eta_b(x, y)$ (Section 15.4.1) is given by

$$w = \mathbf{u} \cdot \nabla_z \eta_b \quad \text{at } z = \eta_b(x, y). \quad (41.74)$$

This relation expresses the no-normal flow condition $\hat{\mathbf{n}} \cdot \mathbf{v} = 0$ at the ocean bottom, with $\hat{\mathbf{n}} = -\nabla(z - \eta_b)/|\nabla(z - \eta_b)|$ the outward normal to the bottom. The boundary condition constrains the flow so that any horizontal motion next to a sloping bottom that is oriented either up or down the slope must have an associated vertical motion. As we see in this section, such vertical motion next to the bottom boundary arises from force curls acting to stretch or squash a fluid column. In turn, through the vorticity equation (41.71), vertical motion at the bottom leads to meridional motion of the full fluid column.

Expressions for bottom vertical velocity

The bottom kinematic boundary condition holds for all dynamical flow regimes. For the particular case of planetary geostrophy, we garner insight into the forces that drive vertical flow near the bottom by making use of the planetary geostrophic momentum equation (41.51a). Evaluating the horizontal components of this equation at the ocean bottom yields

$$\rho_0 f \hat{\mathbf{z}} \wedge \mathbf{u} = -\nabla_z p_b + \mathbf{F}_b, \quad (41.75)$$

where $p_b(x, y, t)$ is the bottom pressure and \mathbf{F}_b is the bottom friction. For the special case of a no-slip bottom, all velocity components vanish at $z = \eta_b$. In that case, we consider \mathbf{u} in equation (41.75) to be the horizontal velocity averaged within the bottom boundary layer, and w the corresponding vertical velocity leaving the boundary layer.

It is convenient to decompose the bottom horizontal velocity into its geostrophic and Ekman¹ components via

$$\rho_0 f \hat{\mathbf{z}} \wedge \mathbf{u}_g = -\nabla_z p_b \quad \text{and} \quad \rho_0 f \hat{\mathbf{z}} \wedge \mathbf{u}_e = \mathbf{F}_b, \quad (41.76)$$

so that

$$\mathbf{u}_g = \frac{\hat{\mathbf{z}} \wedge \nabla_z p_b}{\rho_0 f} \quad \text{and} \quad \mathbf{u}_e = -\frac{\hat{\mathbf{z}} \wedge \mathbf{F}_b}{\rho_0 f}. \quad (41.77)$$

The corresponding bottom vertical velocity components are determined by inserting equations (41.77) into the bottom kinematic boundary condition (41.74)

$$w_g = \frac{\hat{\mathbf{z}} \cdot (-\nabla_z \eta_b \wedge \nabla_z p_b)}{\rho_0 f} \quad \text{and} \quad w_e = \frac{\hat{\mathbf{z}} \cdot (\nabla_z \eta_b \wedge \mathbf{F}_b)}{\rho_0 f}. \quad (41.78)$$

¹Recall our discussion of Ekman mechanics in Chapter 28.

These equations reveal how the curl of inviscid pressure forces and boundary frictional forces drive a nonzero vertical motion next to the bottom, all while maintaining the bottom kinematic boundary condition (41.74). As seen by these equations, is only the projection of $\nabla_z p_b$ and \mathbf{F}_b onto the isobath direction that contributes to a nonzero vertical velocity. These along-isobath forces are needed to render a horizontal velocity that is itself misaligned with isobaths, thus satisfying the kinematic requirement for vertical motion.

To further understand the bottom pressure term, we write it as

$$w_g = \frac{\hat{\mathbf{z}} \cdot (-\nabla_z \eta_b \wedge \nabla_z p_b)}{\rho_0 f} = \frac{\hat{\mathbf{z}} \cdot [\nabla_z \wedge p_b \nabla_z \eta_b]}{\rho_0 f}. \quad (41.79)$$

The numerator is the curl of the horizontal projection of the pressure contact force along the bottom, $p_b \nabla_z \eta_b$. This term is the topographic form stress discussed in Section 20.2 for a general fluid and in Section 35.9.5 for the shallow water. We thus conclude that vertical geostrophic motion next to the bottom arises from the curl of the topographic form stress. This is an important result that will appear again within this section as well as in Sections 41.5 and 41.6.

Comments on the bottom vertical geostrophic velocity

A large part of the bottom pressure gradient driving the horizontal geostrophic flow in equation (41.77) arises from changes in bottom depth. However, that portion of the bottom pressure gradient has no impact on w_g , since it only drives horizontal flow along isobaths. We see this property by writing

$$p_b = -\rho_0 g \eta_b + p'_b \implies w_g = \frac{\hat{\mathbf{z}} \cdot (-\nabla_z \eta_b \wedge \nabla_z p'_b)}{\rho_0 f}. \quad (41.80)$$

When there is misalignment between isolines of bottom pressure and bottom topography, the geostrophic flow in a fluid column will cross isobaths. Correspondingly, with the pressure force misaligned with topographic gradients, the fluid column experiences a twisting action akin to how baroclinicity spins a fluid element if the pressure force does not act through the fluid element's center of mass (see Section 36.5).

To illustrate the above, consider the topographic bowl in Figure 41.1, with sides steep enough so that the bottom pressure gradient is dominated by the topographic slopes. Along the bottom the pressure increases moving down (increasing depth) towards the bowl center. The corresponding bottom geostrophic flow is anti-cyclonic within the bowl and largely follows isobaths. As already noted, if the geostrophic flow exactly follows isobaths, then there is no corresponding vertical component to the bottom velocity. A vertical velocity arises only in the presence of an anomalous bottom pressure gradient, $\nabla_z p'_b$, that is misaligned with the bottom slope, $\nabla \eta_b$. This bottom pressure gradient balances a geostrophic flow that deviates from isobaths thus giving rise to a nonzero w_g . Similar geometric analysis holds for the bottom friction vector, \mathbf{F}_b , and how it gives rise to a nonzero vertical Ekman velocity, w_e .

What causes misalignment between p_b and η_b ?

As we just discussed, misalignment of p_b and η_b lead to vertical geostrophic motion along the bottom. In Section 41.4.3 we will see a similar relation for vertical geostrophic motion at the ocean surface. But what causes such misalignment? The answer to this question is circular when working within the planetary geostrophic whose momentum equation is diagnostic. Even so, we can offer some insight by returning to the depth-integrated vorticity balance (41.71) and rewriting it as an

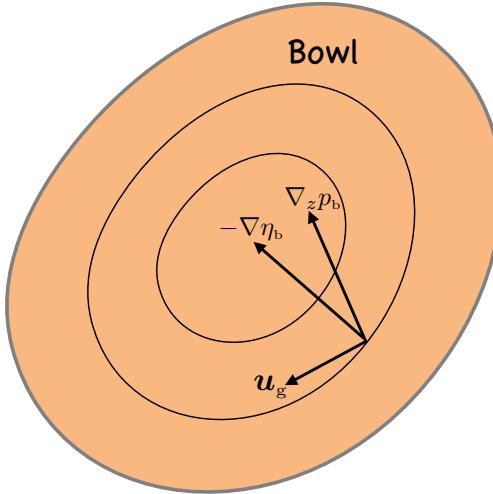


FIGURE 41.1: Depicting how bottom pressure gradients create vertical motion in planetary geostrophic flow next to a sloping bottom according to equations (41.78) and (41.80). Here we show a bowl or depression (local maximum in the depth) with $-\eta_b$ increasing inward toward the bowl center. Only those portions of $\nabla_z p_b$ and \mathbf{F}_b that are aligned parallel to the topographic slope contribute to vertical motion. We illustrate here a case where the bottom pressure gradient leads to $-\nabla \eta_b \cdot \mathbf{u}_g > 0$ so that $w_g < 0$ in the northern hemisphere and $w_g > 0$ in the southern hemisphere (see equations (41.78) and (41.80)).

expression for vertical motion

$$\rho_0 f [w(\eta_b) - w(\eta)] = -\rho_0 \beta V + \hat{\mathbf{z}} \cdot (\nabla_z \wedge \Delta \boldsymbol{\tau}). \quad (41.81)$$

Hence, vertical motion at the surface and bottom arise from meridional motion in the presence of planetary beta, plus the curl of surface and bottom stresses. The absence of planetary beta, and the absence of boundary stress curls, realizes $w(\eta_b) = w(\eta)$, with zero vertical motion the typical case.

41.4.3 Surface kinematics and dynamics

For purposes of large-scale circulation studies using the PG equations, it is generally sufficient to assume a rigid lid upper boundary condition, whereby $w(\eta) = w(0) = 0$. Even so, we find it revealing to present the results for a free surface in which there is the possibility of nonzero surface mass fluxes. The surface kinematic boundary condition for a Boussinesq fluid (Section 17.2) is given by

$$w = -\frac{Q_m}{\rho_0} + \frac{\partial \eta}{\partial t} + \mathbf{u} \cdot \nabla_z \eta \quad \text{at } z = \eta(x, y, t). \quad (41.82)$$

We retain the sea surface time tendency, $\partial_t \eta$, even though for transient solutions the time tendency is many orders of magnitude smaller than the typical vertical velocity under the planetary geostrophic regime.² Evaluating the horizontal planetary geostrophic momentum equation (26.23a) at the ocean surface renders

$$\rho_0 f \hat{\mathbf{z}} \wedge \mathbf{u} = -\nabla_z p_a + \mathbf{F}_\eta, \quad (41.83)$$

where p_a is the pressure applied to the ocean surface and \mathbf{F}_η is the corresponding horizontal friction vector. Like the bottom, we decompose the horizontal velocity into a geostrophic component and

²See Section 3.3 of [Samelson \(2011\)](#) for more details on this scaling of the planetary geostrophic system.

an Ekman component via

$$\rho_0 f \hat{z} \wedge \mathbf{u}_g = -\nabla_z p_a \quad \text{and} \quad \rho_0 f \hat{z} \wedge \mathbf{u}_e = \mathbf{F}_\eta, \quad (41.84)$$

so that

$$\mathbf{u}_g = \frac{\hat{z} \wedge \nabla_z p_a}{\rho_0 f} \quad \text{and} \quad \mathbf{u}_e = -\frac{\hat{z} \wedge \mathbf{F}_\eta}{\rho_0 f}. \quad (41.85)$$

The corresponding vertical velocity components are determined by inserting into the surface kinematic boundary condition (41.82)

$$w_{Q\eta} = -\frac{Q_m}{\rho_0} + \frac{\partial \eta}{\partial t} \quad \text{and} \quad w_g = \frac{\hat{z} \cdot (-\nabla_z \eta \wedge \nabla_z p_a)}{\rho_0 f} \quad \text{and} \quad w_e = \frac{\hat{z} \cdot (\nabla_z \eta \wedge \mathbf{F}_\eta)}{\rho_0 f}, \quad (41.86)$$

where we introduced a vertical velocity, $w_{Q\eta}$, associated with the boundary mass flux and transient sea level fluctuations. As for the bottom, the second and third of these equations reveal how the curl of inviscid pressure forces and boundary frictional forces drive a nonzero vertical motion at the ocean surface, all while maintaining the surface kinematic boundary condition (41.82). Furthermore, it is only the projection of $\nabla_z p_a$ and \mathbf{F}_η onto the direction parallel to sea surface height contours that contributes to a nonzero vertical velocity. This orientation of the surface forces is needed to render a horizontal velocity that is itself misaligned with sea surface height contours, thus satisfying the kinematics required to render vertical motion.

41.4.4 Summary of force curls driving depth integrated meridional flow

Plugging expressions (41.78), (41.80), and (41.86) into equation (41.71) renders the depth integrated planetary vorticity balance

$$\rho_0 \beta V = f (-Q_m + \rho_0 \partial_t \eta) + \hat{z} \cdot [\nabla_z \eta \wedge (\mathbf{F}_\eta - \nabla_z p_a) - \nabla_z \eta_b \wedge (\mathbf{F}_b - \nabla_z p'_b) + \nabla_z \wedge \Delta \boldsymbol{\tau}]. \quad (41.87)$$

We see how the force curls are symmetrically applied at the surface and bottom, which is part of the motivation for exposing the surface terms even though they are generally subdominant. In general, equation (41.87) shows that the depth integrated meridional flow, within the planetary geostrophic regime, is driven by the following processes.

Surface mass transport plus sea surface fluctuations

The term $f (-Q_m + \rho_0 \partial_t \eta)$ arises from mass transport across the ocean surface plus fluctuations in the sea surface height. For example, as sea surface height increases or as water leaves the ocean surface, they impart a positive surface vertical velocity, $w(\eta) > 0$, thus causing column stretching and poleward meridional depth integrated flow. In the steady state, where it is just the mass flux term that contributes, the meridional circulation is known as the Goldsborough-Stommel circulation (see [Huang and Schmitt \(1993\)](#) for a review).

Curl of turbulent boundary stresses

The term $\hat{z} \cdot (\nabla_z \wedge \Delta \boldsymbol{\tau})$ arises from the curl of the turbulent wind stress and turbulent bottom stress. The wind stress term is generally larger than the bottom turbulent stress, with many theories for ocean circulation, particularly those with a flat bottom, almost exclusively focused on the role of surface stress in forcing the planetary geostrophic vorticity budget.

Atmospheric pressure torque

The term

$$\nabla_z p_a \wedge \nabla_z \eta = \nabla_z \wedge (p_a \nabla_z \eta) = -\nabla_z \wedge (\eta \nabla_z p_a) \quad (41.88)$$

arises from differences in lines of constant atmospheric pressure and lines of constant sea surface height. Such misalignments create a torque akin to the baroclinicity detailed in Section 36.5, with these misalignments driving vertical motion and a corresponding depth integrated meridional flow.

Bottom pressure torque

The term

$$\nabla_z p_b' \wedge (-\nabla_z \eta_b) = \nabla_z \wedge (-p_b' \nabla_z \eta_b) = \nabla_z \wedge (\eta_b \nabla_z p_b') \quad (41.89)$$

arises from differences in lines of constant bottom pressure and lines of constant bottom topography. As for the atmospheric pressure torques, such misalignments create a torque that drives a depth integrated meridional flow, with this term vanishing when the bottom topography is flat. In many cases with strong flow next to sloping bottoms, this term can contribute more to the vorticity budget than the turbulent bottom stress. Indeed, in some cases it can rival contributions from the surface wind stress. We sketched out such cases for the shallow water when discussing western boundary currents in Section 35.9.5.

Surface frictional acceleration

The term

$$\hat{z} \cdot (\nabla_z \eta \wedge \mathbf{F}_\eta) = \hat{z} \cdot \left[\nabla_z \eta \wedge \frac{\partial \boldsymbol{\tau}_\eta}{\partial z} \right] \quad (41.90)$$

arises from evaluating the vertical divergence of the frictional stress at the sea surface. A finite volume boundary layer treatment of this term prompts us to integrate the stress divergence over the extent of the surface Ekman layer to render the alternative expression

$$\hat{z} \cdot \int_{-h_{e-surf}}^{\eta} \left[\nabla_z \eta \wedge \frac{\partial \boldsymbol{\tau}}{\partial z} \right] dz = \hat{z} \cdot [\nabla_z \eta \wedge \boldsymbol{\tau}(\eta)], \quad (41.91)$$

where we assumed $\boldsymbol{\tau}$ is negligible at the base of the surface Ekman layer, $z = -h_{e-surf}(x, y, t)$. The term $\hat{z} \cdot [\nabla_z \eta \wedge \boldsymbol{\tau}(\eta)]$ creates a torque from that component of the surface turbulent stress that is aligned with isolines of the sea surface height.

Bottom frictional acceleration

The term

$$\hat{z} \cdot (-\nabla_z \eta_b \wedge \mathbf{F}_b) = \hat{z} \cdot \left[-\nabla_z \eta_b \wedge \frac{\partial \boldsymbol{\tau}_b}{\partial z} \right] \quad (41.92)$$

arises from evaluating the vertical divergence of the frictional stress at the ocean bottom. As for the analogous term for the surface, we offer a finite volume Ekman boundary layer treatment to render the alternative expression

$$\hat{z} \cdot \int_{\eta_b}^{h_{e-bot}} \left[-\nabla_z \eta_b \wedge \frac{\partial \boldsymbol{\tau}}{\partial z} \right] dz = \hat{z} \cdot [-\nabla_z \eta_b \wedge \boldsymbol{\tau}(\eta_b)], \quad (41.93)$$

where we assumed $\boldsymbol{\tau}$ is negligible at the top of the bottom Ekman layer, $z = -\eta_b + h_{e-bot}$. The term $\hat{z} \cdot [-\nabla_z \eta_b \wedge \boldsymbol{\tau}(\eta_b)]$ creates a torque from that component of the bottom turbulent stress that is aligned with isobaths.

41.5 Vorticity of the depth integrated velocity

In Section 41.4 we studied the depth integrated vorticity budget for planetary geostrophic flow. We were led to see how boundary torques (i.e., the curl of boundary forces) lead to vertical motion and in turn, through the β -effect, lead to meridional motion of the depth integrated flow. In this section we present another analysis of vorticity in the planetary geostrophic regime, here focusing on vorticity of the depth integrated velocity.

41.5.1 Depth integrated velocity equation

The depth integrated horizontal velocity equation (41.51a) is given by

$$\rho_0 f \hat{z} \wedge \mathbf{U} = - \int_{\eta_b}^{\eta} \nabla_z p dz + \Delta\tau \quad (41.94)$$

where

$$\mathbf{U} = \int_{\eta_b}^{\eta} \mathbf{u} dz \quad (41.95)$$

is the depth integrated horizontal velocity, and we assumed friction in the form of the vertical divergence of a horizontal turbulent stress as in equation (41.69). For the depth integrated pressure gradient, we follow the decomposition in Section 36.10.2 by writing

$$\int_{\eta_b}^{\eta} p dz = \int_{\eta_b}^{\eta} [d(pz) - z dp] = p_a \eta - p_b \eta_b + \mathcal{P}, \quad (41.96)$$

where we used the hydrostatic balance to write $dp = -g \rho dz$, which is valid for each fluid column. We also introduced the potential energy per horizontal area of a fluid column

$$\mathcal{P} = \int_{\eta_b}^{\eta} g \rho z dz = (g \rho_0/2) (\eta^2 - \eta_b^2) + \int_{\eta_b}^{\eta} g \rho' z dz, \quad (41.97)$$

where

$$\rho' = \rho - \rho_0 \quad (41.98)$$

is the density deviation from the background reference density. These results then lead to the depth integrated horizontal pressure gradient

$$\int_{\eta_b}^{\eta} \nabla_z p dz = \nabla_z \left[\int_{\eta_b}^{\eta} p dz \right] - p_a \nabla_z \eta + p_b \nabla_z \eta_b \quad (41.99a)$$

$$= \nabla_z [p_a \eta - p_b \eta_b + \mathcal{P}] - p_a \nabla_z \eta + p_b \nabla_z \eta_b \quad (41.99b)$$

$$= \eta \nabla_z p_a - \eta_b \nabla_z p_b + \nabla_z \mathcal{P}, \quad (41.99c)$$

which in turn renders the depth integrated planetary geostrophic momentum balance

$$\rho_0 f \hat{z} \wedge \mathbf{U} = -\eta \nabla_z p_a + \eta_b \nabla_z p_b - \nabla_z \mathcal{P} + \Delta\tau. \quad (41.100)$$

The depth integrated balance is here written in terms of gradients in the surface and bottom pressures, the gradient of the potential energy per area, and the difference in turbulent stresses at the top and bottom boundaries, $\Delta\tau = \tau(\eta) - \tau(\eta_b)$.

41.5.2 Vorticity budget

Taking the curl of the depth integrated balance (41.100) annihilates the potential energy term, thus leaving

$$\rho_0 \beta V = -\rho_0 f \nabla \cdot \mathbf{U} + \hat{\mathbf{z}} \cdot \nabla \wedge [p_a \nabla \eta + \boldsymbol{\tau}(\eta) - p_b \nabla \eta_b - \boldsymbol{\tau}(\eta_b)]. \quad (41.101)$$

From Section 17.3, we know that the divergence of the depth-integrated flow for a steady Boussinesq fluid is given by

$$\rho_0 \nabla \cdot \mathbf{U} = Q_m, \quad (41.102)$$

so that

$$\rho_0 \beta V = -f Q_m + \hat{\mathbf{z}} \cdot \nabla \wedge [p_a \nabla \eta + \boldsymbol{\tau}(\eta) - p_b \nabla \eta_b - \boldsymbol{\tau}(\eta_b)]. \quad (41.103)$$

This is the vorticity equation for the depth integrated planetary geostrophic fluid. It is quite similar to the vorticity balance for a shallow water fluid as given by equation (35.126) (which considered zero atmospheric pressure). In the presence of β , meridional mass transport for the fluid column is generated by surface mass fluxes, $Q_m \neq 0$; the curl of surface form stresses and surface turbulent stresses; and the curl of topographic form stresses and bottom turbulent stresses. This result follows quite naturally when recognizing that the forces acting on a depth integrated fluid column arise from the depth integrated stresses acting in the column sides plus those acting on the top and bottom boundaries. In the absence of interior friction stresses due to horizontal strains, it is only the depth integrated pressure that acts on the column sides, and this term has zero curl. We are thus left with the curl of the form stresses and turbulent stresses on the surface and bottom boundaries, along with the contribution from mass transport.

41.6 Vorticity equation for the depth averaged velocity

In some numerical models, it is more common to have access to the depth averaged velocity

$$\bar{\mathbf{u}} = \frac{1}{D} \int_{\eta_b}^{\eta} \mathbf{u} dz \quad \text{with} \quad D = \eta - \eta_b. \quad (41.104)$$

We thus find it useful to derive the budget for the vorticity of the depth averaged velocity. In this discussion we encounter a distinct means for describing how bottom topography, in the presence of baroclinicity, generates meridional flow.

41.6.1 Relating the depth average velocity to boundary velocities

Before studying the vorticity equation, we here relate the depth averaged velocity, $\bar{\mathbf{u}}(x, y, t)$, to the surface velocity, $\mathbf{u}(x, y, z = \eta, t)$, and bottom velocity, $\mathbf{u}(x, y, z = \eta_b, t)$. This analysis exposes some general features of how the boundary flows are driven away from the depth average.

The starting point is the identity

$$\int_{\eta_b}^{\eta} \mathbf{u} dz = (\eta - \eta_b) \mathbf{u}(\eta) - \int_{\eta_b}^{\eta} \frac{\partial \mathbf{u}}{\partial z} (z - \eta_b) dz, \quad (41.105)$$

which, along with the analogous identity for the bottom flow, leads to

$$\bar{\mathbf{u}} - \mathbf{u}(\eta) = - \int_{\eta_b}^{\eta} \frac{\partial \mathbf{u}}{\partial z} \left[\frac{z - \eta_b}{\eta - \eta_b} \right] dz \quad \text{and} \quad \bar{\mathbf{u}} - \mathbf{u}(\eta_b) = \int_{\eta_b}^{\eta} \frac{\partial \mathbf{u}}{\partial z} \left[\frac{\eta - z}{\eta - \eta_b} \right] dz. \quad (41.106)$$

Note that $\mathbf{u}(\eta) - \mathbf{u}(\eta_b) = \int_{\eta_b}^{\eta} (\partial \mathbf{u} / \partial z) dz$ serves as a useful check on the manipulations leading to equation (41.106). We see that the difference between the depth averaged flow and the surface flow, $\bar{\mathbf{u}} - \mathbf{u}(\eta)$, is determined by the integral of the weighted vertical shear, with the weight linearly decreasing from unity at the surface to zero at the bottom. The minus sign in front of the integral is sensible because if the flow increases in the positive direction from the bottom to the surface, then the depth averaged flow will have a smaller magnitude than the surface flow. The converse weighting holds for computing the difference $\bar{\mathbf{u}} - \mathbf{u}(\eta_b)$.

The identities (41.106) hold for arbitrary horizontal velocity fields. Assuming the flow satisfies frictional geostrophy as per equation (41.51a) leads to the frictional thermal wind relation

$$f \mathbf{u} = \rho_0^{-1} \hat{\mathbf{z}} \wedge \nabla p - \hat{\mathbf{z}} \wedge \mathbf{F} \implies f \partial_z \mathbf{u} = -(g/\rho_0) \hat{\mathbf{z}} \wedge \nabla \rho - \hat{\mathbf{z}} \wedge \partial_z \mathbf{F}, \quad (41.107)$$

so that the velocity differences are given by

$$f [\bar{\mathbf{u}} - \mathbf{u}(\eta)] = \int_{\eta_b}^{\eta} [(g/\rho_0) \hat{\mathbf{z}} \wedge \nabla \rho - \hat{\mathbf{z}} \wedge \partial_z \mathbf{F}] \left[\frac{z - \eta_b}{\eta - \eta_b} \right] dz \quad (41.108a)$$

$$f [\bar{\mathbf{u}} - \mathbf{u}(\eta_b)] = - \int_{\eta_b}^{\eta} [(g/\rho_0) \hat{\mathbf{z}} \wedge \nabla \rho - \hat{\mathbf{z}} \wedge \partial_z \mathbf{F}] \left[\frac{\eta - z}{\eta - \eta_b} \right] dz. \quad (41.108b)$$

Hence, differences between the depth averaged flow and the boundary flows are determined by weighted integrals of the baroclinicity and vertical friction shears.

41.6.2 Formulation of the vorticity equation

To develop the vorticity equation, we start by deriving the momentum equation for the depth averaged flow. For that purpose, rearrange the depth integrated momentum budget (41.100) according to

$$\rho_0 f \hat{\mathbf{z}} \wedge \mathbf{U} = -\eta \nabla_z (p_a - p_b) - D \nabla_z p_b - \nabla_z \mathcal{P} + \Delta \boldsymbol{\tau} \quad (41.109)$$

and then divide by the thickness, $D = \eta - \eta_b$, of the ocean column to render

$$\rho_0 f \hat{\mathbf{z}} \wedge \bar{\mathbf{u}} = -(\eta/D) \nabla_z (p_a - p_b) - \nabla_z p_b + (1/D) (-\nabla_z \mathcal{P} + \Delta \boldsymbol{\tau}). \quad (41.110)$$

Taking the curl then leads to

$$\rho_0 \nabla_z \cdot (f \bar{\mathbf{u}}) = -\hat{\mathbf{z}} \cdot \nabla_z \wedge [(\eta/D) \nabla_z (p_a - p_b)] + D^{-2} \hat{\mathbf{z}} \cdot [\nabla_z \wedge (D \nabla_z \mathcal{P})] + \hat{\mathbf{z}} \cdot [\nabla_z \wedge (D^{-1} \Delta \boldsymbol{\tau})] \quad (41.111a)$$

$$= \hat{\mathbf{z}} \cdot \nabla_z \wedge [(p_a - p_b) \nabla_z (\eta/D)] - D^{-2} \hat{\mathbf{z}} \cdot [\nabla_z \wedge (\mathcal{P} \nabla_z D)] + \hat{\mathbf{z}} \cdot [\nabla_z \wedge (D^{-1} \Delta \boldsymbol{\tau})]. \quad (41.111b)$$

The vorticity budget (41.111) is a bit less tidy than that for the depth integrated budget (41.103). Nonetheless, it offers some useful insights concerning the flow. For that purpose, write the left hand side in the suggestive form

$$\nabla_z \cdot (f \bar{\mathbf{u}}) = \nabla_z \cdot [(f/D) \mathbf{U}], \quad (41.112)$$

with f/D reminiscent of the shallow water potential vorticity for the planetary geostrophic flow (Section 40.4). Motivated by this analog, we write the vorticity equation (41.111) in the form

$$\begin{aligned} \rho_0 \mathbf{U} \cdot \nabla_z (f/D) &= -\rho_0 (f/D) \nabla_z \cdot \mathbf{U} + \hat{\mathbf{z}} \cdot \nabla_z \wedge [(p_a - p_b) \nabla_z (\eta/D)] \\ &\quad - D^{-2} \hat{\mathbf{z}} \cdot [\nabla_z \wedge (\mathcal{P} \nabla_z D)] + \hat{\mathbf{z}} \cdot [\nabla_z \wedge (D^{-1} \Delta \boldsymbol{\tau})]. \end{aligned} \quad (41.113)$$

Contrary to the shallow water case, we here see that even for a perfect planetary geostrophic fluid, the depth-integrated flow does not generally follow contours of constant f/D . Even so, it is of interest to examine how the processes on the right hand side contribute to flow deviations from f/D contours. For that purpose we simplify the flow even more by making the rigid lid approximation.

41.6.3 Rigid lid approximation and the role of JEBAR

The rigid lid approximation is commonly made for studies of large-scale circulation. Indeed, it was the basis for many ocean general circulation models following the work of [Bryan \(1969\)](#). A fluid satisfying the rigid lid approximation has a vanishing horizontal divergence for the depth integrated flow

$$\text{rigid lid approximation} \implies \nabla_z \cdot \mathbf{U} = 0. \quad (41.114)$$

Furthermore, as part of the rigid lid approximation we assume the free surface undulations are much smaller than the resting ocean depth so that³

$$|\eta| \ll |\eta_b| \implies \eta/D \approx 0 \quad (41.115a)$$

$$1/(\eta - \eta_b) \approx 1/(-\eta_b) = 1/H. \quad (41.115b)$$

Hence, with the rigid lid approximation the vorticity equation (41.113) takes the simplified form

$$\rho_0 \mathbf{U} \cdot \nabla_z (f/H) = \hat{\mathbf{z}} \cdot [\nabla_z \mathcal{P} \wedge \nabla_z (1/H) + \nabla_z \wedge (H^{-1} \Delta \boldsymbol{\tau})]. \quad (41.116)$$

JEBAR drives deviations from f/H aligned flow

Equation (41.116) indicates that in the absence of boundary stresses, contours of constant f/H serve as streamlines for flow where the potential energy isolines are parallel to lines of constant H

$$\rho_0 H^2 \mathbf{U} \cdot \nabla_z (f/H) = \hat{\mathbf{z}} \cdot (\nabla_z H \wedge \nabla_z \mathcal{P}) = \hat{\mathbf{z}} \cdot (\nabla_z \wedge H \nabla_z \mathcal{P}) = -\hat{\mathbf{z}} \cdot (\nabla_z \wedge \mathcal{P} \nabla_z H). \quad (41.117)$$

The right hand side offers equivalent expressions for the Joint Effect of Baroclinicity and Relief (JEBAR). This name arises since JEBAR is nonzero only in the presence of non-flat topography (“relief”) and (as shown below) when there are horizontal density gradients (i.e., baroclinicity). Equation (41.117) says that the depth integrated flow does not generally align with f/H contours. Misalignment of the flow with f/H contours is driven by misalignments of isobaths and isolines of the depth integrated potential energy, with this misalignment referred to as JEBAR.

Let us show that contributions to JEBAR arise only from the component of the potential energy that deviates from a constant density reference state. For that purpose note that in the rigid lid approximation, the potential energy in a column, as given by equation (41.97), takes the form

$$\mathcal{P} = (1/2) g \rho_0 H^2 + g \int_{-H}^0 \rho' z dz, \quad (41.118)$$

where $\rho' = \rho - \rho_0$. With $\nabla_z H \wedge \nabla_z (\rho_0 H^2) = 0$, we are left with just the contribution from ρ' . Conversely, we see that the depth integrated horizontal flow of a steady, rigid lid, homogeneous fluid, in the absence of turbulent boundary stresses, aligns with f/H contours, which connects to the case of a single layer of shallow water fluid.

³We use the notation $\eta_b = -H$ in this subsection to correspond to the literature for rigid lid models and the JEBAR term. Note that H is the vertical depth scale, which is a constant, and it is distinct from $H(x, y) = -\eta_b(x, y)$, which is a function of horizontal position.

Relating JEBAR to pressure

JEBAR as given by the right hand side of equation (41.117) has the appearance of the curl of a form stress, and yet it is not. The reason is that \mathcal{P} is the potential energy of the fluid column rather than bottom pressure. We make this point explicit by recalling the decomposition (41.96), here specialized to the rigid lid in which

$$\mathcal{P} = H(\bar{p} - p_b) \quad \text{with} \quad \bar{p} = H^{-1} \int_{-H}^0 p dz = (-\eta_b)^{-1} \int_{\eta_b}^0 p dz. \quad (41.119)$$

Plugging into the vorticity equation (41.116) leads to

$$\rho_0 \mathbf{U} \cdot \nabla_z (f/H) = H^{-1} \hat{\mathbf{z}} \cdot [\nabla_z H \wedge \nabla_z (\bar{p} - p_b)] + \hat{\mathbf{z}} \cdot \nabla_z \wedge (H^{-1} \Delta \boldsymbol{\tau}). \quad (41.120)$$

The JEBAR term thus arises from deviations of the depth averaged pressure from the bottom pressure. We can go one more step in exposing this pressure difference through use of equation (41.106), here applied to hydrostatic pressure, in which case

$$\bar{p} - p_b = \int_{\eta_b}^0 \frac{\partial p}{\partial z} \frac{z}{\eta_b} dz = (g/H) \int_{-H}^0 z \rho dz = H^{-1} \mathcal{P} \quad (41.121)$$

so that

$$g^{-1} \nabla_z (\bar{p} - p_b) = -\rho(z = \eta_b) \nabla_z H - \int_{-H}^0 \nabla_z (\rho z/H) dz. \quad (41.122)$$

The first term arises from slopes in the bottom topography as weighted by the bottom density, whereas the second term arises from the density integral of the horizontal gradients in the depth-weighted density.

Equation (41.103) provides the budget for vorticity of the depth integrated flow, in which we find the curl of the topographic form stress leads to vortex stretching. For the vorticity of the depth averaged flow, we instead encounter the JEBAR term in equation (41.120), which is not a pure vortex stretching term. Instead, it accounts for the fact that it is the horizontal velocity flowing across isobaths, $\mathbf{u}(z = \eta_b)$, rather than depth averaged horizontal velocity, \bar{u} , that leads to vortex stretching. Hence, when studying vorticity of the depth averaged velocity, $\bar{\mathbf{u}}$, accounting for the role of vortex stretching requires us to include JEBAR.

41.6.4 Further study

Chapter 3 in [Samelson \(2011\)](#) offers more discussion of the large-scale ocean circulation based on planetary geostrophy. [Mertz and Wright \(1992\)](#) discuss the physics of how JEBAR relates to the curl of the topographic form stress as well as other mathematically equivalent forms. [Cane et al. \(1998\)](#) as well a Section 2.5 of [Drijfhout et al. \(2013\)](#) offer critical remarks on JEBAR.



Continuously stratified quasi-geostrophy

Quasi-geostrophy (QG) is the canonical *balanced model* found in geophysical fluid mechanics whereby the potential vorticity is the sole prognostic field; all other fields, such as the velocity and buoyancy, are diagnosed from potential vorticity. The process of diagnosing the allied fields requires the inversion of an elliptic operator to compute the geostrophic streamfunction from the potential vorticity. Some authors refer to this connection between potential vorticity and streamfunction as *invertibility*, with the mathematical technology required for inversion shared with many other elliptic problems in mathematical physics (see Section 3.5 for more on elliptic partial differential equations).

Quasi-geostrophy is an elegant theory of mathematical physics offering great physical and mathematical insights into the workings of large-scale geophysical fluid motions. Giving the theory full justice requires a book the size of the current one. Instead, we here aim to provide a taste of its continuously stratified realization, offering a detailed derivation that builds from Chapters 40 and 41, and sampling some of its physical and mathematical content.

READER'S GUIDE FOR THIS CHAPTER

In this chapter we extend to continuously stratified fluids the shallow water discussions of quasi-geostrophy in Chapter 40. Continuously stratified QG is not concerned with the relatively slow processes that affect stratification, but instead with the shorter time scale processes that slightly perturb that stratification. We make use of stratified geophysical fluid dynamics from Chapters 18 and 26, as well as potential vorticity from Chapter 38 and the development of continuously stratified planetary geostrophy in Chapter 41.

42.1	Open threads	960
42.2	The quasi-geostrophic PV equation	960
42.2.1	Zeroth order asymptotic equations	961
42.2.2	First order asymptotic equations	962
42.3	Reintroducing physical dimensions	963
42.3.1	Hydrostatic balance	964
42.3.2	Geostrophic balance	964
42.3.3	Material time derivative	964
42.3.4	Buoyancy equation	965
42.3.5	Vorticity equation	965
42.3.6	Velocity equation	966
42.3.7	Potential vorticity	966
42.4	Connecting quasi-geostrophic PV to Ertel PV	967

42.4.1	Non-dimensionalizing the Ertel potential vorticity	967
42.4.2	Material conservation of Ertel PV to order Ro^1	968
42.5	Mathematical expressions of the theory	969
42.5.1	The Jacobian form of geostrophic advection	969
42.5.2	Constant background buoyancy frequency	969
42.5.3	Vertical boundary conditions for ψ	970
42.5.4	Potential vorticity induction and impermeability	971
42.6	Energetics for quasi-geostrophic flow	973
42.6.1	Kinetic energy	973
42.6.2	Available potential energy	974
42.6.3	Energy conversion	974
42.6.4	Scaling APE and KE	975
42.7	Exercises	975

42.1 Open threads

- Check with Yassin concerning the boundary conditions in Section 42.5.4. What about a free upper surface?
- Finish discussion of impermeability and gauge transformation in Section 42.5.4.
- Rossby waves
- Synoptic development

42.2 The quasi-geostrophic PV equation

In deriving the quasi-geostrophic PV equation, we proceed much like for the single layer of shallow water fluid in Section 40.5. In particular, quasi-geostrophic scaling from Section 40.5.1 is relevant for both the shallow water and for the continuously stratified fluid. We employ an asymptotic expansion in the Rossby number and stop at the first nontrivial order, which is Ro^1 . For this purpose, recall the non-dimensional momentum and continuity equations from Section 41.1.8

$$\text{Ro} \left[\frac{\partial \hat{\mathbf{u}}}{\partial \hat{t}} + (\hat{\mathbf{u}} \cdot \hat{\nabla}_{\hat{z}}) \hat{\mathbf{u}} + \hat{w} \frac{\partial \hat{\mathbf{u}}}{\partial \hat{z}} \right] + \hat{\mathbf{f}} \wedge \hat{\mathbf{u}} = -\hat{\nabla}_{\hat{z}} \hat{\varphi} \quad (42.1a)$$

$$\frac{\partial \hat{\varphi}}{\partial \hat{z}} = \hat{b} \quad (42.1b)$$

$$\hat{\nabla} \cdot \hat{\mathbf{v}} = 0 \quad (42.1c)$$

$$\frac{\text{Ro}}{\text{Bu}} \frac{\text{D}\hat{b}}{\text{D}\hat{t}} + \hat{w} = 0. \quad (42.1d)$$

We expand the variables in an asymptotic series in Rossby number

$$\hat{\mathbf{u}} = \hat{\mathbf{u}}_0 + \text{Ro} \hat{\mathbf{u}}_1 + \text{Ro}^2 \hat{\mathbf{u}}_2 + \dots \quad (42.2a)$$

$$\hat{\mathbf{v}} = \hat{\mathbf{v}}_0 + \text{Ro} \hat{\mathbf{v}}_1 + \text{Ro}^2 \hat{\mathbf{v}}_2 + \dots \quad (42.2b)$$

$$\hat{w} = \hat{w}_0 + \text{Ro} \hat{w}_1 + \text{Ro}^2 \hat{w}_2 + \dots \quad (42.2c)$$

$$\hat{b} = \hat{b}_0 + \text{Ro} \hat{b}_1 + \text{Ro}^2 \hat{b}_2 + \dots \quad (42.2d)$$

$$\hat{\varphi} = \hat{\varphi}_0 + \text{Ro} \hat{\varphi}_1 + \text{Ro}^2 \hat{\varphi}_2 + \dots \quad (42.2e)$$

along with the expansion (40.55) for the Coriolis parameter

$$\hat{f} = \hat{f}_0 + \text{Ro} \hat{\beta} \hat{y}, \quad (42.3)$$

and where (equation (40.56))

$$\hat{\beta} \hat{y} = \frac{\beta y}{\text{Ro} f_0} = T \beta y. \quad (42.4)$$

As noted in Section 40.5.2, the velocity field is non-divergent at each order of Rossby number, so that

$$\nabla \cdot \hat{\mathbf{v}}_n = 0 \quad \forall n. \quad (42.5)$$

The Burger number is order unity since the horizontal length scales for the quasi-geostrophic flows of concern here are on the order of the deformation radius

$$\text{Bu} \sim 1 \implies L \sim L_d, \quad (42.6)$$

where we introduced the *internal deformation radius* from Section 41.1.6

$$L_d(z) = H \frac{N(z)}{f_0}. \quad (42.7)$$

It is important to retain the depth dependence of the Burger number through its dependence on the background stratification $N^2(z)$

$$\text{Bu}(z) = \left[\frac{L_d}{L} \right]^2 = N(z)^2 \left[\frac{H}{L f_0} \right]^2. \quad (42.8)$$

Hence, the Burger number does not commute with the vertical derivative operator.

42.2.1 Zeroth order asymptotic equations

The zeroth order asymptotic equations take the form

$$\hat{\mathbf{f}}_0 \wedge \hat{\mathbf{u}}_0 = -\hat{\nabla}_z \hat{\varphi}_0 \quad (42.9a)$$

$$\frac{\partial \hat{\varphi}_0}{\partial \hat{z}} = \hat{b}_0 \quad (42.9b)$$

$$\hat{\nabla}_z \cdot \hat{\mathbf{u}}_0 + \frac{\partial \hat{w}_0}{\partial \hat{z}} = 0 \quad (42.9c)$$

$$\hat{w}_0 = 0. \quad (42.9d)$$

The first equation represents f -plane geostrophy, which means that the horizontal velocity has zero divergence

$$\hat{\nabla}_z \cdot \hat{\mathbf{u}}_0 = 0. \quad (42.10)$$

Equation (42.9b) means the zeroth order buoyancy determines the zeroth order hydrostatic pressure. Since the horizontal velocity has zero divergence, the continuity equation (42.9c) means that the vertical velocity is depth independent

$$\frac{\partial \hat{w}_0}{\partial \hat{z}} = 0. \quad (42.11)$$

If it vanishes somewhere, such as a solid boundary, then it vanishes everywhere. This is a manifestation of the Taylor-Proudman theorem (see Section 26.5.3). Indeed, a vanishing \hat{w}_0 is required

by the zeroth-order buoyancy equation (42.9d). Hence, the non-dimensional velocity has a nonzero contribution only at order Ro^1

$$\hat{w} = \text{Ro} \hat{w}_1 + \text{Ro}^2 \hat{w}_2 + \dots, \quad (42.12)$$

thus manifesting the vertical stiffening of fluid columns found in rotating fluids. Correspondingly, the dimensionful vertical velocity has the asymptotic expansion

$$w = W\hat{w} = W\text{Ro}(\hat{w}_1 + \text{Ro} \hat{w}_2 + \dots), \quad (42.13)$$

so that to leading to order Ro^1

$$\hat{w}_1 = \frac{w}{W\text{Ro}}. \quad (42.14)$$

Since the zeroth-order horizontal velocity is non-divergent, we can introduce a geostrophic streamfunction

$$\hat{u}_0 = -\partial\hat{\psi}_0/\partial\hat{y} \quad \text{and} \quad \hat{v}_0 = \partial\hat{\psi}_0/\partial\hat{x} \quad \text{and} \quad \hat{\zeta}_0 = \hat{\nabla}^2\hat{\psi}_0, \quad (42.15)$$

where the zeroth-order streamfunction is the ratio of the zeroth order pressure to zeroth order Coriolis parameter

$$\hat{\psi}_0 = \hat{\varphi}_0/\hat{f}_0. \quad (42.16)$$

Note also that the zeroth-order system satisfies the thermal wind balance

$$\hat{f}_0 \wedge \frac{\partial \hat{u}_0}{\partial \hat{z}} = -\hat{\nabla}_z \hat{b}_0. \quad (42.17)$$

Finally, note that the zeroth order buoyancy is determined by the streamfunction through the hydrostatic balance

$$\hat{b}_0 = \frac{\partial \hat{\varphi}_0}{\partial \hat{z}} = \hat{f}_0 \frac{\partial \hat{\psi}_0}{\partial \hat{z}}. \quad (42.18)$$

42.2.2 First order asymptotic equations

The zeroth order equations do not render a prognostic equation, for which we need to consider equations at order Ro^1

$$\frac{\text{D}_0 \hat{u}_0}{\text{D}\hat{t}} + \hat{f}_0 \wedge \hat{u}_1 + \hat{\beta} \hat{y} \hat{z} \wedge \hat{u}_0 = -\hat{\nabla}_z \hat{\varphi}_1 \quad (42.19a)$$

$$\frac{\partial \hat{\varphi}_1}{\partial \hat{z}} = \hat{b}_1 \quad (42.19b)$$

$$\hat{\nabla}_z \cdot \hat{u}_1 + \frac{\partial \hat{w}_1}{\partial \hat{z}} = 0 \quad (42.19c)$$

$$\frac{1}{\text{Bu}} \frac{\text{D}_0 \hat{b}_0}{\text{D}\hat{t}} + \hat{w}_1 = 0. \quad (42.19d)$$

The first order terms are often referred to as the *ageostrophic* components, though note that all contributions higher than zeroth order constitute ageostrophic contributions.

At this order, the material time derivative makes use *only* of the zeroth order horizontal geostrophic velocity

$$\frac{\text{D}_0}{\text{D}\hat{t}} = \frac{\partial}{\partial \hat{t}} + \hat{u}_0 \cdot \hat{\nabla}. \quad (42.20)$$

To close this set of equations, we produce the vorticity equation from the momentum equation, and then combine the vorticity equation and buoyancy equation to produce the QG potential vorticity equation.

Taking the curl of the momentum equation (42.19a) eliminates the pressure gradient, $\hat{\nabla}\hat{\varphi}_1$, thus producing the vorticity equation

$$\frac{\partial \hat{\zeta}_0}{\partial \hat{t}} + (\hat{\mathbf{u}}_0 \cdot \hat{\nabla}) (\hat{\zeta}_0 + \hat{\beta} \hat{y}) = -\hat{f}_0 \hat{\nabla}_z \cdot \mathbf{u}_1. \quad (42.21)$$

We make use of the continuity equation (42.19c) to eliminate the horizontal convergence

$$\frac{\partial \hat{\zeta}_0}{\partial \hat{t}} + (\hat{\mathbf{u}}_0 \cdot \hat{\nabla}) (\hat{\zeta}_0 + \hat{\beta} \hat{y}) = \hat{f}_0 \frac{\partial \hat{w}_1}{\partial \hat{z}}. \quad (42.22)$$

The right hand side represents the contribution to vorticity evolution from stretching by the ageostrophic vertical velocity through planetary rotation. We can eliminate the ageostrophic vertical velocity through the buoyancy equation (42.19d). When doing so, it is important to keep the $Bu(z)$ depth dependence according to equation (42.8), with this depth dependence arising from the background stratification. The resulting vorticity equation is

$$\frac{\partial \hat{\zeta}_0}{\partial \hat{t}} + (\hat{\mathbf{u}}_0 \cdot \hat{\nabla}) (\hat{\zeta}_0 + \hat{\beta} \hat{y}) = -\hat{f}_0 \frac{\partial}{\partial \hat{z}} \left[\frac{1}{Bu} \frac{D_0 \hat{b}_0}{Dt} \right]. \quad (42.23)$$

We now use the identity

$$\frac{\partial}{\partial \hat{z}} \left[\frac{1}{Bu} \frac{D_0 \hat{b}_0}{Dt} \right] = \frac{\partial}{\partial \hat{z}} \left[\frac{1}{Bu} \left(\frac{\partial}{\partial \hat{t}} + \hat{\mathbf{u}}_0 \cdot \hat{\nabla} \right) \hat{b}_0 \right] \quad (42.24a)$$

$$= \frac{D_0}{Dt} \left[\frac{\partial}{\partial \hat{z}} \left(\frac{\hat{b}_0}{Bu} \right) \right] + \frac{1}{Bu} \frac{\partial \hat{\mathbf{u}}_0}{\partial \hat{z}} \cdot \hat{\nabla}_z \hat{b}_0 \quad (42.24b)$$

$$= \frac{D_0}{Dt} \left[\frac{\partial}{\partial \hat{z}} \left(\frac{\hat{b}_0}{Bu} \right) \right], \quad (42.24c)$$

where we set

$$\frac{\partial \hat{\mathbf{u}}_0}{\partial \hat{z}} \cdot \hat{\nabla}_z \hat{b}_0 = 0 \quad (42.25)$$

since the zeroth-order velocity maintains thermal wind balance (42.17). Bringing terms together then leads to the material conservation equation for quasi-geostrophic potential vorticity

$$\frac{D_0}{Dt} \left[\hat{\zeta}_0 + \hat{\beta} \hat{y} + \hat{f}_0 \frac{\partial}{\partial \hat{z}} \left(\frac{\hat{b}_0}{Bu} \right) \right] = 0. \quad (42.26)$$

42.3 Reintroducing physical dimensions

To expose the physical elements to the quasi-geostrophic theory, we find it useful, if not essential, to reintroduce physical dimensions much like we did for the shallow water in Section 40.6. For that

purpose, we write

$$\mathbf{u} = \mathbf{u}_g + \mathbf{u}_{ag} = U (\hat{\mathbf{u}}_0 + \text{Ro} \hat{\mathbf{u}}_1) \quad (42.27a)$$

$$w \equiv w_{ag} = \text{Ro} W \hat{w}_1 \quad (42.27b)$$

$$b \equiv b_g + b_{ag} = B (\hat{b}_0 + \text{Ro} \hat{b}_1) \quad (42.27c)$$

$$\varphi \equiv \varphi_g + \varphi_{ag} = f_0 U L (\hat{\varphi}_0 + \text{Ro} \hat{\varphi}_1). \quad (42.27d)$$

The details of reintroducing dimensions are straightforward but can be tedious. We provide full details to help develop experience and confidence working with the theory.

42.3.1 Hydrostatic balance

Hydrostatic balance is maintained for terms at each Rossby number order, so that we have the dimensional equations

$$\frac{\partial \varphi_g}{\partial z} = b_g \quad \text{and} \quad \frac{\partial \varphi_{ag}}{\partial z} = b_{ag}. \quad (42.28)$$

42.3.2 Geostrophic balance

The non-dimensional geostrophic balance for the zeroth order fields

$$\hat{\mathbf{f}}_0 \wedge \hat{\mathbf{u}}_0 = -\hat{\nabla}_z \hat{\varphi}_0 \quad (42.29)$$

takes on the dimensional form

$$(f_0/f_0) \hat{\mathbf{z}} \wedge \mathbf{u}_g/U = -L \nabla_z \varphi_g/(f_0 U L). \quad (42.30)$$

Cancelling factors leads to the expected form of f -plane geostrophy

$$f_0 \hat{\mathbf{z}} \wedge \mathbf{u}_g = -\nabla_z \varphi_g. \quad (42.31)$$

42.3.3 Material time derivative

For the material time derivative operator we write

$$D/Dt = \partial_t + \mathbf{u} \cdot \nabla + w \partial_z \quad (42.32a)$$

$$= (1/T) \partial_{\hat{t}} + (U/L) \hat{\mathbf{u}} \cdot \hat{\nabla} + (W/H) \hat{w} \partial_{\hat{z}} \quad (42.32b)$$

$$= (1/T) (\partial_{\hat{t}} + \hat{\mathbf{u}} \cdot \hat{\nabla} + \hat{w} \partial_{\hat{z}}) \quad (42.32c)$$

$$= (1/T) (\partial_{\hat{t}} + \hat{\mathbf{u}}_0 \cdot \hat{\nabla}_z) + (\text{Ro}/T) (\hat{\mathbf{u}}_0 \cdot \hat{\nabla}_z + \hat{w}_1 \partial_{\hat{z}}) \quad (42.32d)$$

$$= \partial_t + \mathbf{u}_g \cdot \nabla_z + \mathbf{u}_{ag} \cdot \nabla_z + w_{ag} \partial_z \quad (42.32e)$$

$$\equiv D_g/Dt + \mathbf{u}_{ag} \cdot \nabla_z + w_{ag} \partial_z, \quad (42.32f)$$

where time scales according to advection, $T = L/U$, vertical velocity scales according to continuity, $W = H U/L$, and we introduced the geostrophic material time derivative operator

$$D_g/Dt \equiv \partial_t + \mathbf{u}_g \cdot \nabla_z. \quad (42.33)$$

42.3.4 Buoyancy equation

We split buoyancy into a depth dependent static background and a deviation from the background

$$b = \tilde{b}(z) + b'(x, y, z, t), \quad (42.34)$$

with its vertical derivative

$$\partial_z b = N^2 + \partial_z b' \quad \text{with} \quad N^2 = \partial_z \tilde{b}, \quad (42.35)$$

where $N^2(z)$ is the squared buoyancy frequency for the background buoyancy field. In this manner, we can write the buoyancy equation for a perfect fluid as

$$\frac{\partial b'}{\partial t} + \mathbf{u} \cdot \nabla b' + w(N^2 + \partial_z b') = 0. \quad (42.36)$$

We non-dimensionalize this equation by making use of the following relations between the scales

$$B = f_0 U L / H \quad W = H(U/L) \quad \text{Ro} = U/(f_0 L) \quad \text{Bu} = (N H)^2 / (f_0 L)^2, \quad (42.37)$$

in which case the buoyancy equation takes the form

$$\frac{\partial b'}{\partial t} + \mathbf{u} \cdot \nabla b' + w(N^2 + \partial_z b') = \frac{B}{T} \frac{\partial \hat{b}}{\partial \hat{t}} + \frac{U B}{L} \hat{\mathbf{u}} \cdot \hat{\nabla}_z \hat{b} + W \hat{w} N^2 + \frac{W B}{H} \hat{w} \partial_{\hat{z}} \hat{b} = 0. \quad (42.38)$$

We find it useful to divide by $f_0 N^2$, so that

$$\frac{1}{f_0 N^2} \left[\frac{\partial b'}{\partial t} + \mathbf{u} \cdot \nabla b' + w(N^2 + \partial_z b') \right] = \frac{H \text{Ro}^2}{\text{Bu}} \left[\frac{\partial \hat{b}}{\partial \hat{t}} + \hat{\mathbf{u}} \cdot \hat{\nabla} \hat{b} + \hat{w} \partial_{\hat{z}} \hat{b} \right] + \text{Ro} \hat{w} = 0. \quad (42.39)$$

The vertical velocity component, \hat{w} , is nonzero only at order Ro^1 , so the term $\text{Ro} \hat{w}$ is order Ro^2 . To the same order, we drop the term, $\hat{w} \partial_{\hat{z}} \hat{b}$, that appears within the bracket and retain only the zeroth order buoyancy contribution, \hat{b}_0 . Reintroducing physical dimensions then leads to the dimensional form of the quasi-geostrophic buoyancy equation

$$\partial_t b_g + \mathbf{u}_g \cdot \nabla_z b_g + w_{ag} N^2 = D_g b_g / Dt + w_{ag} N^2 = 0. \quad (42.40)$$

This equation means that the geostrophic transport of the geostrophic buoyancy is affected by a source due to the ageostrophic vertical advection of background buoyancy

$$D_g b_g / Dt = -w_{ag} N^2 \quad (42.41)$$

42.3.5 Vorticity equation

Reintroducing dimensions to the vorticity equation (42.22) yields¹

$$\frac{\partial \hat{\zeta}_0}{\partial \hat{t}} + (\hat{\mathbf{u}}_0 \cdot \hat{\nabla}_z) (\hat{\zeta}_0 + \beta \hat{y}) - \hat{f}_0 \frac{\partial \hat{w}_1}{\partial \hat{z}} = T^2 \left[\frac{\partial \zeta}{\partial t} + \mathbf{u}_g \cdot \nabla_z (\zeta + \beta y) \right] - \frac{H}{W \text{Ro}} \frac{\partial w_{ag}}{\partial z}. \quad (42.42)$$

The identity $H/(W \text{Ro}) = f_0 T^2$ leads to the order Ro^1 vorticity equation

$$\frac{\partial \zeta_a}{\partial t} + J(\psi, \zeta_a) = f_0 \frac{\partial w_{ag}}{\partial z}. \quad (42.43)$$

¹Recall from equation (42.27b) that $W \hat{w} = \text{Ro} W \hat{w}_1 = w_{ag}$.

42.3.6 Velocity equation

The prognostic velocity equation arises from the first order asymptotic equation (42.19a)

$$\frac{D_0 \hat{\mathbf{u}}_0}{Dt} + \hat{\mathbf{f}}_0 \wedge \hat{\mathbf{u}}_1 + \hat{\beta} \hat{\mathbf{y}} \hat{\mathbf{z}} \wedge \hat{\mathbf{u}}_0 = -\hat{\nabla}_z \hat{\varphi}_1. \quad (42.44)$$

Our skills with reintroducing dimensional quantities should be sufficient to write down the dimensional velocity equation by inspection²

$$(\partial_t + \mathbf{u}_g \cdot \nabla) \mathbf{u}_g + \beta y \hat{\mathbf{z}} \wedge \mathbf{u}_g + f_0 \hat{\mathbf{z}} \wedge \mathbf{u}_{ag} = -\nabla_z \varphi_{ag}. \quad (42.45)$$

42.3.7 Potential vorticity

From equation (42.26), we identify the non-dimensional quasi-geostrophic potential vorticity

$$\hat{q} = \hat{\zeta}_0 + \hat{\beta} \hat{y} + \hat{f}_0 \frac{\partial(\hat{b}_0/Bu)}{\partial \hat{z}}. \quad (42.46)$$

Introducing dimensional quantities to the right hand side yields³

$$\hat{q} = \frac{L}{U} [\zeta_g + \beta y] + \frac{\partial}{\partial z} \left[\frac{H b_g}{B Bu} \right] \quad (42.47a)$$

$$= (1/(f_0 Ro)) (\zeta_g + \beta y) + \frac{H}{B} \frac{\partial}{\partial z} \left[\frac{b_g}{Bu} \right]. \quad (42.47b)$$

The scale for the fluctuating buoyancy is given by equation (41.22), $B = f_0 U L / H$, and the inverse Burger number is given by equation (42.8), $Bu^{-1} = [(L f_0) / (H N)]^2$, so that

$$\hat{q} = (1/(f_0 Ro)) (\zeta_g + \beta y) + \frac{H^2}{f_0 U L} \frac{L^2 f_0^2}{H^2} \frac{\partial}{\partial z} \left[\frac{b_g}{N^2} \right] \quad (42.48a)$$

$$= (1/(f_0 Ro)) (\zeta_g + \beta y) + \frac{1}{Ro} \left[\frac{\partial}{\partial z} \left(\frac{b_g}{N^2} \right) \right]. \quad (42.48b)$$

Introducing the geostrophic streamfunction,

$$\mathbf{u}_g = \hat{\mathbf{z}} \wedge \nabla_z \psi \quad \text{and} \quad \zeta = \hat{\mathbf{z}} \cdot (\nabla \wedge \mathbf{u}_g) = \nabla_z^2 \psi \quad \text{and} \quad b_g = f_0 \partial_z \psi, \quad (42.49)$$

leads to

$$q \equiv f_0 Ro \hat{q} = \zeta_g + \beta y + \frac{\partial}{\partial z} \left[\frac{f_0^2}{N^2} \frac{\partial \psi}{\partial z} \right]. \quad (42.50)$$

Just as for the shallow water case in Section 40.6.1, the potential vorticity (42.50) scales as $f_0 Ro$. We sometimes choose to add the constant f_0 to q , which has no effect on the dynamics but reveals the beta plane planetary vorticity

$$q = \underbrace{f_0 + \beta y}_{\text{planetary vorticity}} + \underbrace{\nabla_z^2 \psi}_{\text{relative vorticity}} + \underbrace{\frac{\partial}{\partial z} \left[\frac{f_0^2}{N^2} \frac{\partial \psi}{\partial z} \right]}_{\text{stretching by } f}. \quad (42.51)$$

We identify the following three contributions to the quasi-geostrophic potential vorticity.

²Recall from equation (42.27d) that $\varphi_{ag} = f_0 U L Ro \hat{\varphi}_1 = U^2 \hat{\varphi}_1$.

³Recall $\hat{f}_0 = 1$ and $\hat{\beta} \hat{y} = T \beta y = (L/U) \beta y$.

- PLANETARY VORTICITY: The planetary vorticity contribution, $f_0 + \beta y$, arises from rotation of the reference frame. The βy term is the only dynamically relevant contribution, so that we can equally well drop the f_0 contribution.
- GEOSTROPHIC RELATIVE VORTICITY: The vertical component of the geostrophic relative vorticity, $\zeta = \hat{z} \cdot (\nabla \wedge \mathbf{u}) = \nabla_z^2 \psi$, acts to bring out the smaller scale features in the streamfunction.
- VERTICAL STRETCHING: The final contribution arises from the vertical stretching in the presence of a rotating planet. Equation (42.22) helps to remind us why this term arises from vortex stretching.

42.4 Connecting quasi-geostrophic PV to Ertel PV

Following our treatment for the shallow water system in Section 40.6.4, we here determine how quasi-geostrophic potential vorticity relates to the Ertel potential vorticity from Chapter 38. For this purpose, consider the continuously stratified hydrostatic Boussinesq fluid and make use of the Ertel potential vorticity derived in Exercise 38.1

$$Q = (\boldsymbol{\omega} + \hat{z} f) \cdot \nabla b = \frac{\partial u}{\partial z} \frac{\partial b}{\partial y} - \frac{\partial v}{\partial z} \frac{\partial b}{\partial x} + (\zeta + f) \frac{\partial b}{\partial z}. \quad (42.52)$$

In a perfect fluid we have the material conservation

$$(\partial_t + \mathbf{u} \cdot \nabla_z + w \partial_z) Q = 0. \quad (42.53)$$

Our strategy in this section is to non-dimensionalize both Q and the material time operator, and then to stratify terms in equation (42.53) according to the Rossby number. We will show that material conservation of Ertel potential vorticity, when expanded asymptotically to order Ro^1 , leads to the geostrophic material conservation of quasi-geostrophic potential vorticity. The continuous stratification makes the derivation more involved than for the shallow water model. Indeed, it requires a number of manipulations commonly found with quasi-geostrophy theory, thus prompting us to expose the variety of details to offer practical experience with the formalism.

42.4.1 Non-dimensionalizing the Ertel potential vorticity

As above for the buoyancy, we are led to write the Ertel potential vorticity in the form

$$\frac{Q - f_0 N^2}{f_0 N^2} = \frac{1}{f_0 N^2} \left[\frac{\partial u}{\partial z} \frac{\partial b'}{\partial y} - \frac{\partial v}{\partial z} \frac{\partial b'}{\partial x} \right] + \frac{1}{N^2} \frac{\partial b'}{\partial z} + \frac{\beta y + \zeta}{f_0} \left[1 + \frac{1}{N^2} \frac{\partial b'}{\partial z} \right]. \quad (42.54)$$

with non-dimensionalization leading to

$$\frac{1}{f_0 N^2} \left[\frac{\partial u}{\partial z} \frac{\partial b'}{\partial y} - \frac{\partial v}{\partial z} \frac{\partial b'}{\partial x} \right] = \frac{B U}{f_0 N^2 H L} \left[\frac{\partial \hat{u}}{\partial \hat{z}} \frac{\partial \hat{b}}{\partial \hat{y}} - \frac{\partial \hat{v}}{\partial \hat{z}} \frac{\partial \hat{b}}{\partial \hat{x}} \right] = \frac{\text{Ro}^2}{\text{Bu}} \left[\frac{\partial \hat{u}}{\partial \hat{z}} \frac{\partial \hat{b}}{\partial \hat{y}} - \frac{\partial \hat{v}}{\partial \hat{z}} \frac{\partial \hat{b}}{\partial \hat{x}} \right] \quad (42.55a)$$

$$\frac{1}{N^2} \frac{\partial b'}{\partial z} = \frac{B}{H N^2} \frac{\partial \hat{b}}{\partial \hat{z}} = \frac{\text{Ro}}{\text{Bu}} \frac{\partial \hat{b}}{\partial \hat{z}} \quad (42.55b)$$

$$\frac{\beta y + \zeta}{f_0} = \text{Ro} (\hat{\beta} \hat{y} + \hat{\zeta}). \quad (42.55c)$$

The order Ro^2 terms appearing in equation (42.55a) are dropped since they do not contribute to the quasi-geostrophic potential vorticity, which involve terms only up to order Ro^1 . For the order

Ro^1 term, we only retain the zeroth order buoyancy, $\widehat{b}_0 = b_g/B$, and likewise we just retain the zeroth order vorticity, $\widehat{\zeta}_0 = (L/U) \zeta_g$. Hence, the Ertel potential vorticity is given by

$$Q = N^2 (f_0 + q_*) + \mathcal{O}(\text{Ro}^2) \quad (42.56)$$

where q_* is the order Ro^1 term

$$q_* = \text{Ro} f_0 \left[\frac{1}{\text{Bu}} \frac{\partial \widehat{b}_0}{\partial \widehat{z}} + \widehat{\beta} \widehat{y} + \widehat{\zeta}_0 \right] = \frac{f_0}{N^2} \frac{\partial b_g}{\partial z} + \beta y + \zeta_g. \quad (42.57)$$

42.4.2 Material conservation of Ertel PV to order Ro^1

The material conservation of Ertel PV now takes the form

$$(\partial_t + \mathbf{u}_g \cdot \nabla_z + w_{\text{ag}} \partial_z) (f_0 N^2 + q_* N^2) = f_0 w_{\text{ag}} \partial_z N^2 + N^2 D_g q_* / Dt = 0. \quad (42.58a)$$

We dropped the $w_{\text{ag}} \partial_z$ contribution to the advection of q_* since ageostrophic vertical advection of q_* is an order Ro^2 term. To eliminate the vertical ageostrophic velocity we make use of the buoyancy equation (42.40) so that

$$\frac{D_g q_*}{Dt} + \frac{w_{\text{ag}}}{N^2} \frac{\partial N^2}{\partial z} = \frac{D_g q_*}{Dt} - \frac{f_0}{N^4} \frac{D_g b_g}{Dt} \frac{\partial N^2}{\partial z} = 0. \quad (42.59)$$

Writing

$$\frac{\partial}{\partial z} \left[\frac{1}{N^2} \right] = -\frac{1}{N^4} \frac{\partial N^2}{\partial z} \quad (42.60)$$

leads to

$$\frac{D_g q_*}{Dt} + f_0 \left[\frac{\partial N^{-2}}{\partial z} \right] \frac{D_g b_g}{Dt} = 0. \quad (42.61)$$

Since the geostrophic material time derivative operator only involves horizontal advection, we can merge these two terms to render

$$\frac{D_g q_*}{Dt} + f_0 \left[\frac{\partial N^{-2}}{\partial z} \right] \frac{D_g b_g}{Dt} = \frac{D_g}{Dt} \left[q_* + f_0 b_g \left(\frac{\partial N^{-2}}{\partial z} \right) \right] \quad (42.62a)$$

$$= \frac{D_g}{Dt} \left[\beta y + \zeta + \frac{f_0}{N^2} \left(\frac{\partial b_g}{\partial z} \right) + f_0 b_g \left(\frac{\partial N^{-2}}{\partial z} \right) \right] \quad (42.62b)$$

$$= \frac{D_g}{Dt} \left[\beta y + \zeta + f_0 \frac{\partial}{\partial z} \left(\frac{b_g}{N^2} \right) \right] \quad (42.62c)$$

$$= \frac{D_g q}{Dt} \quad (42.62d)$$

$$= 0. \quad (42.62e)$$

In the penultimate step we introduced the quasi-geostrophic potential vorticity given by equation (42.51)

$$q = \beta y + \zeta + f_0 \frac{\partial}{\partial z} \left[\frac{b_g}{N^2} \right] = q_* + f_0 b_g \frac{\partial N^{-2}}{\partial z}. \quad (42.63)$$

We have thus established how material conservation of Ertel potential vorticity, when expanded asymptotically to order Ro^1 , leads to the geostrophic material conservation of quasi-geostrophic potential vorticity.

42.5 Mathematical expressions of the theory

In this section we sample various mathematical expressions of quasi-geostrophic theory as well as boundary conditions at the top and bottom of the domain.

42.5.1 The Jacobian form of geostropic advection

The geostrophic velocity, as a horizontally non-divergent field, can be written in terms of the quasi-geostrophic streamfunction

$$\mathbf{u}_g = \hat{\mathbf{z}} \wedge \nabla \psi. \quad (42.64)$$

We can thus write the following equivalent forms for the material time derivative of quasi-geostrophic PV

$$\frac{Dq}{Dt} = \frac{\partial q}{\partial t} + \mathbf{u}_g \cdot \nabla q \quad (42.65a)$$

$$= \frac{\partial q}{\partial t} + (\hat{\mathbf{z}} \wedge \nabla \psi) \cdot \nabla q \quad (42.65b)$$

$$= \frac{\partial q}{\partial t} + (\nabla \psi \wedge \nabla q) \cdot \hat{\mathbf{z}} \quad (42.65c)$$

$$= \frac{\partial q}{\partial t} + J(\psi, q). \quad (42.65d)$$

The final equality introduced the Jacobian operator $J(\psi, q) = (\nabla \psi \wedge \nabla q) \cdot \hat{\mathbf{z}}$, which is a notation commonly used in the geophysical fluids literature.⁴

For a perfect fluid, in which $Dq/Dt = 0$, a steady state (zero Eulerian time derivative) is realized when

$$\mathbf{u}_g \cdot \nabla q = (\nabla \psi \wedge \nabla q) \cdot \hat{\mathbf{z}} = J(\psi, q) = 0. \quad (42.66)$$

Hence, the velocity is parallel to surfaces of constant q . We are ensured that these equalities hold if the streamfunction is a function only of the potential vorticity

$$\psi = \psi(q) \implies (\nabla \psi \wedge \nabla q) \cdot \hat{\mathbf{z}} = J(\psi, q) = 0. \quad (42.67)$$

As the steady state is of physical interest, this functional relation between streamfunction and potential vorticity commonly arises in applications.

42.5.2 Constant background buoyancy frequency

Consider the quasi-geostrophic potential vorticity for the special case of a constant background buoyancy frequency, $N^2 = \text{constant}$, in which the relative potential vorticity (42.50) takes on the form

$$q - \beta y = \nabla_z^2 \psi + \frac{\partial}{\partial z} \left[\frac{f_0^2}{N^2} \frac{\partial \psi}{\partial z} \right] \quad (42.68a)$$

$$= \frac{\partial^2 \psi}{\partial x^2} + \frac{\partial^2 \psi}{\partial y^2} + \frac{f_0^2}{N^2} \frac{\partial^2 \psi}{\partial z^2} \quad (42.68b)$$

$$= [\partial_{xx} + \partial_{yy} + \partial_{zz}] \psi. \quad (42.68c)$$

⁴Recall that we also encountered the Jacobian form for non-divergent two-dimensional advection in Section 34.3.3 when studying the non-divergent barotropic flows.

For the final equality we introduced the vertical coordinate

$$\tilde{z} = (N/f) z. \quad (42.69)$$

Since $|N/f| \gg 1$ the stably stratified flows considered in QG, \tilde{z} is a *stretched* vertical coordinate so that the Laplacian operator acting on ψ is anisotropic. The linear operator acting on ψ remains elliptic even in the more general case of a depth dependent stratification, thus warranting the use of elliptical solvers when performing the inversion numerically.

42.5.3 Vertical boundary conditions for ψ

We need boundary conditions on the geostrophic streamfunction, ψ , to invert the elliptic quasi-geostrophic PV equation (42.51). For lateral boundaries, one may choose periodicity, whereby the boundaries are absent. Alternatively, we may choose to set the normal component of the flow to zero for the inviscid case, in which case ψ is a constant along material boundaries as discussed in Section 17.4.2. The top and bottom boundaries require extra considerations as discussed here.

Buoyancy equation at the boundaries

To develop the vertical boundary conditions on the streamfunction, consider the quasi-geostrophic buoyancy equation (42.40)

$$D_g b_g / Dt + w_{ag} N^2 = 0. \quad (42.70)$$

Inserting the geostrophic streamfunction, $b_g = f_0 \partial_z \psi$, leads to

$$f_0 (\partial_t + \mathbf{u}_g \cdot \nabla_z) \partial_z \psi + w_{ag} N^2 = 0. \quad (42.71)$$

Rigid top and bottom boundaries

At a rigid lid surface boundary or flat bottom boundary, $w_{ag} = 0$, in which case the boundary condition reduces to geostrophic advection of boundary buoyancy.

$$(\partial_t + \mathbf{u}_g \cdot \nabla_z) \partial_z \psi = 0 \quad \text{at } z = 0, \bar{\eta_b} \quad \text{if } w_{ag} = 0 \text{ at top \& bottom boundaries.} \quad (42.72)$$

That is, the geostrophic buoyancy is materially invariant at the boundaries.

Sloping bottom boundary condition

With a nonzero slope in the bottom topography, $\nabla \eta_b \neq 0$, the bottom kinematic boundary condition (Section 15.4.1) says that velocity is constrained so that

$$\mathbf{v} \cdot \hat{\mathbf{n}} = 0 \implies w = \mathbf{u} \cdot \nabla \eta_b, \quad (42.73)$$

where $\hat{\mathbf{n}} = -\nabla(z - \eta_b)/|\nabla(z - \eta_b)|$ is the boundary's outward normal. Expanding this kinematic boundary condition leads to

$$w_{ag} = (\mathbf{u}_g + \mathbf{u}_{ag}) \cdot \nabla \eta_b, \quad (42.74)$$

which means there is vertical motion so long as the horizontal motion is not aligned with isobaths. To be clear on the implications of this boundary condition, it is useful to examine the asymptotics

by non-dimensionalizing⁵

$$\mathbf{u}_g = U \hat{\mathbf{u}}_0 \quad (42.75a)$$

$$w_{ag} = W \text{Ro} \hat{w}_1 \quad (42.75b)$$

$$\mathbf{u}_{ag} = U \text{Ro} \hat{\mathbf{u}}_1 \quad (42.75c)$$

$$\nabla \eta_b = (\mathcal{B}/L) \hat{\nabla} \hat{\eta}_b = (H \text{Ro}/L) \hat{\nabla} \hat{\eta}_b, \quad (42.75d)$$

which brings the kinematic boundary condition (42.74) to

$$f_0 H \text{Ro}^2 (\hat{\mathbf{u}}_0 + \text{Ro} \hat{\mathbf{u}}_1) \cdot \hat{\nabla} \hat{\eta}_b = W \text{Ro} \hat{w}_1 \implies (\hat{\mathbf{u}}_0 + \text{Ro} \hat{\mathbf{u}}_1) \cdot \hat{\nabla} \hat{\eta}_b = \hat{w}_1. \quad (42.76)$$

Asymptotic consistency implies that

$$\hat{\mathbf{u}}_0 \cdot \hat{\nabla} \hat{\eta}_b = \hat{w}_1 \implies \mathbf{u}_g \cdot \nabla \eta_b = w_{ag}. \quad (42.77)$$

Hence, for quasi-geostrophic flow, any projection of the horizontal geostrophic velocity in a direction not aligned with isobaths leads to an ageostrophic vertical velocity component.

Use of the bottom kinematic boundary condition in the buoyancy equation (42.71) leads to the bottom boundary evolution of buoyancy

$$f_0 \partial_t (\partial_z \psi) + \mathbf{u}_g \cdot \nabla_z [f_0 \partial_z \psi + N^2 \eta_b] = 0. \quad \text{at } z = \bar{\eta}_b. \quad (42.78)$$

The boundary condition is evaluated at $z = \bar{\eta}_b$ since the more precise boundary location, $z = \eta_b(x, y)$, is higher order in Rossby number.

42.5.4 Potential vorticity induction and impermeability

The potential vorticity equation is the sole prognostic equation required to evolve the quasi-geostrophic flow. Consequently, one often discards the the quasi-geostrophic velocity and buoyancy equations. Even so, we found it useful to use the buoyancy equation in Section 42.3.4 as part of connecting quasi-geostrophic and Ertel potential vorticities. Likewise, there are occasions when it is useful to examine the velocity equation, with a similar discussion provided in Section 40.6.3. In this section we directly connect the velocity and buoyancy equations and clearly reveals their connection to the potential vorticity equation.

Combining the velocity and buoyancy equations

For this purpose, consider the quasi-geostrophic velocity and buoyancy equations derived in Section 42.3

$$(\partial_t + \mathbf{u}_g \cdot \nabla) \mathbf{u}_g + \hat{\mathbf{z}} \wedge (\beta y \mathbf{u}_g + f_0 \mathbf{u}_{ag}) = -\nabla_z \varphi_{ag} + \mathbf{F} \quad (42.79a)$$

$$(\partial_t + \mathbf{u}_g \cdot \nabla_z) b_g + w_{ag} N^2 = \dot{b} \quad (42.79b)$$

$$f_0 \hat{\mathbf{z}} \wedge \mathbf{u}_g + \nabla_z \varphi_g = 0 \quad (42.79c)$$

$$\nabla \cdot \mathbf{u}_g = \partial_x u_g + \partial_y v_g = 0 \quad (42.79d)$$

$$\nabla \cdot \mathbf{v}_{ag} = \nabla \cdot \mathbf{u}_{ag} + \partial_z w_{ag} = 0, \quad (42.79e)$$

⁵The bottom topography slope is non-dimensionalized according to the shallow water discussion in Section 40.3.4.

where we added a non-conservative force per mass, \mathbf{F} (e.g., friction, wind stress), and irreversible buoyancy source, \dot{b} (e.g., diffusion, boundary fluxes). Taking $-\hat{\mathbf{z}} \wedge$ on the velocity equation and multiplying the buoyancy equation by f_0/N^2 leads to

$$(\partial_t + \mathbf{u}_g \cdot \nabla)(-\hat{\mathbf{z}} \wedge \mathbf{u}_g) + \beta y \mathbf{u}_g + f_0 \mathbf{u}_{ag} = \hat{\mathbf{z}} \wedge \nabla_z \varphi_{ag} - \hat{\mathbf{z}} \wedge \mathbf{F} \quad (42.80a)$$

$$(\partial_t + \mathbf{u}_g \cdot \nabla_z)(f_0 b_g/N^2) + f_0 w_{ag} = f_0 \dot{b}/N^2. \quad (42.80b)$$

Introduce the following vector fields

$$\mathbf{D} \equiv -\hat{\mathbf{z}} \wedge \mathbf{u}_g + (f_0/N^2) b_g \hat{\mathbf{z}} = \nabla_z \psi + (f_0/N)^2 \partial_z \psi \hat{\mathbf{z}} \quad (42.81a)$$

$$\mathbf{R} \equiv -\hat{\mathbf{z}} \wedge \mathbf{F} + (f_0/N^2) \dot{b} \hat{\mathbf{z}}, \quad (42.81b)$$

with \mathbf{D} built from both the velocity and buoyancy fields and \mathbf{R} built from the corresponding non-conservative tendencies. These vectors combine the velocity and buoyancy equations into a single vector equation

$$(\partial_t + \mathbf{u}_g \cdot \nabla) \mathbf{D} + \beta y \mathbf{u}_g + f_0 \mathbf{v}_{ag} = \hat{\mathbf{z}} \wedge \nabla_z \varphi_{ag} + \mathbf{R}. \quad (42.82)$$

Potential vorticity induction

The divergence of \mathbf{D} yields the relative quasi-geostrophic potential vorticity

$$\nabla \cdot \mathbf{D} = q - \beta y. \quad (42.83)$$

In analogy to Gauss's law of electromagnetism, we refer to \mathbf{D} as the quasi-geostrophic potential vorticity *induction vector*.⁶ Additionally, the potential vorticity equation can be written (see Exercise 42.5)

$$(\partial_t + \mathbf{u}_g \cdot \nabla) q = \nabla \cdot \mathbf{R}, \quad (42.84)$$

so that the quasi-geostrophic potential vorticity flux vector

$$\mathbf{J}_q = \mathbf{u}_g q - \mathbf{R} \quad (42.85)$$

allows us to write the potential vorticity equation in the Eulerian flux-form

$$\partial_t q = -\nabla \cdot \mathbf{J}_q. \quad (42.86)$$

Kinematic PV flux and impermeability

Taking the Eulerian time derivative of the potential vorticity induction equation (42.83) renders

$$\partial_t q = \partial_t (\nabla \cdot \mathbf{D}) = \nabla \cdot (\partial_t \mathbf{D}) \equiv -\nabla \cdot \mathbf{J}_q^{\text{kin}}, \quad (42.87)$$

where we introduced the kinematic form of the potential vorticity flux

$$\mathbf{J}_q^{\text{kin}} \equiv -\partial_t \mathbf{D} = \mathbf{J}_q + \nabla \wedge \mathbf{A}, \quad (42.88)$$

with \mathbf{A} a gauge function.

which is analogous to the kinematic Ertel PV flux discussed in Section 39.1.2.

⁶This connection between potential vorticity dynamics and electromagnetism was pointed out by [Schneider et al. \(2003\)](#) and further examined by [Maddison and Marshall \(2013\)](#).

42.6 Energetics for quasi-geostrophic flow

Consider a quasi-geostrophic fluid configured with flat upper (rigid lid) and lower boundaries, and assume the lateral boundaries are either periodic or constant streamfunction on solid boundaries. These restrictive assumptions allow us to more readily study energetics within the fluid, sans the impacts from boundary effects. For this purpose we make use of the quasi-geostrophic vorticity and buoyancy equations

$$\frac{\partial \zeta}{\partial t} + \mathbf{u} \cdot \nabla \zeta = f_0 \frac{\partial w}{\partial z} \quad (42.89a)$$

$$\frac{\partial b}{\partial t} + \mathbf{u} \cdot \nabla b = -w N^2, \quad (42.89b)$$

where all labels are dropped from the variables to reduce clutter, and where

$$\frac{D}{Dt} = \frac{\partial}{\partial t} + \mathbf{u} \cdot \nabla \quad b = f_0 \frac{\partial \psi}{\partial z} \quad u = -\frac{\partial \psi}{\partial y} \quad v = \frac{\partial \psi}{\partial x} \quad \zeta = \nabla_z^2 \psi. \quad (42.90)$$

42.6.1 Kinetic energy

The kinetic energy per mass for the total fluid domain is given by the integral

$$\mathcal{K} = \frac{1}{2} \int \mathbf{u} \cdot \mathbf{u} dV = \frac{1}{2} \int \nabla_z \psi \cdot \nabla_z \psi dV, \quad (42.91)$$

and its time derivative is

$$\frac{d\mathcal{K}}{dt} = \int \nabla_z \psi \cdot \nabla_z (\partial \psi / \partial t) dV. \quad (42.92)$$

For this result, we noted that the fluid domain has a constant volume to thus allow the time derivative to move inside the integral without introducing boundary terms. Manipulation renders

$$\frac{d\mathcal{K}}{dt} = \int \nabla_z \psi \cdot \nabla_z (\partial \psi / \partial t) dV \quad (42.93a)$$

$$= \int [\nabla_z \cdot [\psi \nabla_z (\partial \psi / \partial t)] - \psi \partial (\nabla_z^2 \psi) / \partial t] dV \quad (42.93b)$$

$$= - \int \psi (\partial \zeta / \partial t) dV, \quad (42.93c)$$

where we dropped the lateral boundary term and introduced relative vorticity. Use of the vorticity equation (42.89a) yields

$$\frac{d\mathcal{K}}{dt} = - \int \psi (\partial \zeta / \partial t) dV = \int \psi [\mathbf{u} \cdot \nabla_z \zeta - f_0 \partial_z w] dV. \quad (42.94)$$

The first term vanishes since

$$\int \psi (\mathbf{u} \cdot \nabla_z \zeta) dV = \int \psi \nabla_z \cdot (\mathbf{u} \zeta) dV = \int [\nabla_z \cdot (\psi \mathbf{u} \zeta) - \nabla_z \psi \cdot \mathbf{u} \zeta] dV = 0, \quad (42.95)$$

where the boundary term vanishes and $\mathbf{u} \cdot \nabla_z \psi = 0$ since ψ is the streamfunction for the horizontal geostrophic flow. We are thus left with the expression for the kinetic energy evolution

$$\frac{d\mathcal{K}}{dt} = - \int \psi f_0 \frac{\partial w}{\partial z} dV. \quad (42.96)$$

Since the top and bottom are assumed flat, the vertical velocity vanishes on these boundaries (rigid lid top and flat bottom), in which case

$$\frac{d\mathcal{K}}{dt} = - \int \psi f_0 \frac{\partial w}{\partial z} dV = - \int f_0 \left[\frac{\partial(w\psi)}{\partial z} - w \frac{\partial\psi}{\partial z} \right] dV = \int f_0 w \frac{\partial\psi}{\partial z} dV. \quad (42.97)$$

Making use of $b = f_0 \partial\psi/\partial z$ leads to

$$\frac{d\mathcal{K}}{dt} = \int w b dV. \quad (42.98)$$

Kinetic energy thus increases when vertical motion is positively correlated with anomalous buoyancy. For example, upward motion ($w > 0$) of a positive buoyancy anomaly (relatively light water has $b > 0$) increases kinetic energy, as does downward motion of a negative buoyancy anomaly. This behavior is also reflected in the full fluid system discussed in Section 22.3.

42.6.2 Available potential energy

Available potential energy was introduced in Section 24.8 within the context of the Boussinesq fluid. An approximate form of the APE is given by equation (24.159), which we here write in the form

$$\mathcal{A}_{\text{bouss}} \approx \frac{1}{2} \int \left[\frac{b}{N} \right]^2 dV. \quad (42.99)$$

With $b = f_0 \partial\psi/\partial z$, the quasi-geostrophic APE is given by

$$\mathcal{A} = \frac{1}{2} \int \left[\frac{f_0}{N} \frac{\partial\psi}{\partial z} \right]^2 dV. \quad (42.100)$$

Taking a time derivative leads to

$$\frac{d\mathcal{A}}{dt} = \int \left[\frac{f_0}{N} \right]^2 \frac{\partial\psi}{\partial z} \frac{\partial}{\partial t} \frac{\partial\psi}{\partial z} dV = \int \frac{f_0}{N^2} \frac{\partial\psi}{\partial z} [-w N^2 - \nabla_z \cdot (\mathbf{u} b)] dV, \quad (42.101)$$

where we used the buoyancy equation (42.89b) for the second equality. The second term vanishes since

$$\int \frac{f_0}{N^2} \frac{\partial\psi}{\partial z} [\nabla_z \cdot (\mathbf{u} b)] dV = \int \left[\frac{f_0^2}{N^2} \frac{\partial\psi}{\partial z} \right] \mathbf{u} \cdot \nabla_z \left[\frac{\partial\psi}{\partial z} \right] dV = \frac{1}{2} \int \nabla_z \cdot \left[\mathbf{u} \left(\frac{f_0}{N} \frac{\partial\psi}{\partial z} \right)^2 \right] dV = 0. \quad (42.102)$$

Consequently, the quasi-geostrophic APE has a time derivative given by

$$\frac{d\mathcal{A}}{dt} = - \int w f_0 \frac{\partial\psi}{\partial z} dV = - \int w b dV. \quad (42.103)$$

42.6.3 Energy conversion

Notice how the evolution of kinetic energy involves the relative vorticity equation, whereas evolution of the APE involves the buoyancy equation. However, their sum remains constant in time since as the kinetic energy increases through buoyancy work, the available potential energy decreases

$$\frac{d(\mathcal{K} + \mathcal{A})}{dt} = 0. \quad (42.104)$$

The buoyancy work conversion term is given by

$$\text{buoyancy work} = \int w f_0 \frac{\partial \psi}{\partial z} dV = \int w b dV, \quad (42.105)$$

which has the same form as that encountered for the conversion between potential and kinetic energy in the unapproximated equations studied in Section 22.3.

42.6.4 Scaling APE and KE

The scale for the kinetic energy is given by

$$\mathcal{K} = \frac{1}{2} \int (\nabla_z \psi \cdot \nabla_z \psi) dV \sim L^{-2} \Psi^2 V \quad (42.106)$$

and the scale for the APE is

$$\mathcal{A} = \frac{1}{2} \int \left[\frac{f_0}{N} \frac{\partial \psi}{\partial z} \right]^2 dV \sim H^{-2} (f_0/N)^2 \Psi^2 V = L_d^{-2} \Psi^2 V, \quad (42.107)$$

where we wrote Ψ for the streamfunction scale, V for the domain volume, and $L_d = H(N/f_0)$ is the deformation radius (see equation (42.7)). Taking the ratio yields

$$\frac{\mathcal{K}}{\mathcal{A}} \sim \left[\frac{L_d}{L} \right]^2 = \left[\frac{H}{L} \right]^2 \left[\frac{N}{f_0} \right]^2 = \text{Bu}. \quad (42.108)$$

Hence, the Burger number is the ratio of the quasi-geostrophic kinetic energy scale to the quasi-geostrophic available potential energy scale. A large Burger number means that the horizontal scales of the flow are smaller than the deformation radius, in which case the QG dynamics is dominated by its kinetic energy. In contrast, for scales larger than the deformation radius (not much larger, as then the flow would not satisfy QG scaling), the Burger number is less than unity, in which case the QG dynamics is dominated by available potential energy.

42.7 Exercises

EXERCISE 42.1: A VARIETY OF POTENTIAL VORTICITIES

Give the mathematical expressions for potential vorticity in the following fluid models. Define all terms in the respective expressions. Give the physical dimensions and/or SI units for the potential vorticity. Hint: the answers can be found somewhere in this book.

- (a) Ertel PV for compressible fluid in a rotating reference frame
- (b) Ertel PV for an incompressible (Boussinesq) fluid in a rotating reference frame
- (c) Single shallow water layer on a beta plane
- (d) Continuously stratified planetary geostrophic flow
- (e) Continuously stratified quasi-geostrophic flow on a beta plane

EXERCISE 42.2: QUASI-GEOSTROPHIC PV EVOLUTION WITH VERTICAL FRICTION

The first part of this problem involves elements of the asymptotic method used for deriving the QG equations, only now with the advent of a non-zero friction. Use is made to incorporate the non-dimensionalization detailed in Section 28.1, which provides a detailed discussion of the Ekman number and Ekman layers. The second part of the question makes use of the thermal wind balance to connect vertical viscous momentum transfer to horizontal buoyancy transfer.

- (a) Derive the material evolution equation for quasi-geostrophic PV in a continuously stratified Boussinesq fluid in the presence of friction, \mathbf{F} . Assume the Ekman number is on the order of the Rossby number, so that the zeroth order asymptotic solution satisfies the usual inviscid f -plane geostrophic balance. Friction only appears in the first order equations.
- (b) Assume friction arises just from vertical shears in the horizontal velocity, so that

$$\mathbf{F} = \frac{\partial}{\partial z} \left[\nu \frac{\partial \mathbf{u}}{\partial z} \right], \quad (42.109)$$

where $\nu = \nu(z)$ is a vertical eddy viscosity that is a function of depth (dimensions of squared length per time). Also assume an approximate form of quasi-geostrophic PV in which we drop relative vorticity (i.e., quasi-geostrophic PV is dominated by planetary vorticity and stretching). Determine the form for the vertical eddy viscosity so that the approximate form of quasi-geostrophic PV is laterally diffused via

$$\frac{Dq^{\text{approx}}}{Dt} = A \nabla_z^2 q^{\text{approx}}, \quad (42.110)$$

where A is a constant eddy diffusivity for the potential vorticity.

Hint: to leading order, the friction operator is a function just of the geostrophic velocity.

EXERCISE 42.3: QUASI-GEOSTROPHIC ω -EQUATION

As we discovered in this chapter, for quasi-geostrophy the vertical component to the velocity is non-zero only at first order in Rossby number, whereas the zeroth order flow is horizontal and geostrophic. To time step the horizontal geostrophic flow it is not necessary to explicitly compute the ageostrophic vertical velocity. However, there are cases where it is of interest. In this exercise we derive the ω -equation for quasi-geostrophic flow, thus providing a diagnostic expression for the ageostrophic vertical velocity. The name for this equation originates from the atmospheric community where ω is the common symbol for transport across pressure surfaces. Here, we make use of the Boussinesq system so that the vertical velocity component is across depth surfaces.

An outline for the derivation of the traditional form for the ω -equation is given in Section 5.4 of [Vallis \(2017\)](#) for the anelastic version of quasi-geostrophy and for the f -plane. Here we work with the Boussinesq system and consider a β -plane. Nonetheless, the solution is nearly the same as in [Vallis \(2017\)](#). Hence, your job throughout this exercise is to fully explain the derivation and show each of the relevant steps.

- (a) Cross-multiply the dimensional buoyancy and vorticity equations from Section 42.3 to eliminate the time derivative, thus deriving a diagnostic equation for the ageostrophic vertical velocity that is valid to order Ro^1 .
- (b) The diagnostic equation you should have derived takes the form

$$\mathcal{L}w = \sigma, \quad (42.111)$$

where

$$\mathcal{L} = N^2 \nabla_z^2 + f_0^2 \frac{\partial^2}{\partial z^2} \quad (42.112)$$

is a linear partial differential operator and

$$\sigma = f_0 \partial_z [J(\psi, \zeta + \beta y)] - \nabla_z^2 J(\psi, b) \quad (42.113)$$

is a source term. The source is a function of the geostrophic flow and the buoyancy field. For vertically stable stratification, $N^2 > 0$, characterize the differential operator, \mathcal{L} , according to the elliptic, hyperbolic, or parabolic classes discussed in Section 3.4. What about when $N^2 < 0$? What new physical phenomena do you expect when $N^2 < 0$?

EXERCISE 42.4: HOSKINS' FORM OF THE QUASI-GEOSTROPHIC ω -EQUATION

We here rederive the ω -equation from Exercise 42.3 using methods introduced by [Hoskins et al. \(1978\)](#). It is not necessary to have solved Exercise 42.3 to solve the present exercise.

Hoskins' approach reveals an insightful form for the source function contributing to vertical motion. As in Exercise 42.3, we work with the adiabatic and hydrostatic Boussinesq system (see Section 41.1)

$$\frac{Du}{Dt} - fv = -\frac{\partial \varphi}{\partial x} \quad (42.114a)$$

$$\frac{Dv}{Dt} + fu = -\frac{\partial \varphi}{\partial y} \quad (42.114b)$$

$$\frac{\partial \varphi}{\partial z} = b \quad (42.114c)$$

$$\frac{Db}{Dt} = 0 \quad (42.114d)$$

$$\nabla \cdot \mathbf{v} = \nabla_z \cdot \mathbf{u} + \frac{\partial w}{\partial z} = 0 \quad (42.114e)$$

$$\frac{D}{Dt} = \frac{\partial}{\partial t} + \mathbf{v} \cdot \nabla. \quad (42.114f)$$

In Exercise 42.3, we derived the ω -equation making use of the buoyancy equation and vorticity equation. [Hoskins et al. \(1978\)](#) worked with the momentum equation rather than the vorticity equation. For this purpose, rather than consider an asymptotic expansion, Hoskins exactly decomposed the horizontal velocity into its geostrophic and ageostrophic components

$$\mathbf{u} = \mathbf{u}_g + \mathbf{u}_{ag}, \quad (42.115)$$

with the geostrophic velocity balancing the horizontal gradient of the full pressure field

$$f \mathbf{u}_g = \hat{\mathbf{z}} \wedge \nabla \varphi. \quad (42.116)$$

This definition for \mathbf{u}_g is distinct from that arising from an asymptotic expansion, whereby the geostrophic velocity is the zeroth order term balancing the zeroth order pressure gradient (see Section 42.2.1 or Exercise 42.3). We are generally able to access the full hydrostatic pressure field through knowledge of the buoyancy field, in which case there is no need to make an asymptotic expansion of pressure.

Hoskins' definition for the geostrophic velocity brings the horizontal momentum equations into the rather elegant form

$$\frac{D\mathbf{u}}{Dt} + f \hat{\mathbf{z}} \wedge \mathbf{u}_{ag} = 0, \quad (42.117)$$

with the pressure gradient annihilated since it exactly balances the geostrophic velocity. In this manner, the material evolution of horizontal velocity is determined solely by the ageostrophic Coriolis acceleration. Again, there has been no approximation made thus far. Rather, we have only introduced a strategic decomposition of the velocity field as per Hoskins.

At this point we make the quasi-geostrophic approximation by setting the momentum equation equal to

$$\frac{\partial \mathbf{u}_g}{\partial t} + (\mathbf{u}_g \cdot \nabla_z) \mathbf{u}_g + f \hat{z} \wedge \mathbf{u}_{ag} = 0 \quad (42.118)$$

and the buoyancy equation equal to

$$\frac{\partial b}{\partial t} + \mathbf{u}_g \cdot \nabla_z b + N^2 w = 0, \quad (42.119)$$

where $N^2(z)$ is a prescribed static background stratification. That is, both the horizontal geostrophic velocity and the buoyancy are advected just by the geostrophic velocity. The buoyancy equation is the same as derived to order Ro^1 using asymptotic methods (Section 42.2.2). However, Hoskins' momentum equation (42.118) has no pressure gradient on the right hand side, whereas an asymptotic approach has contributions from higher order pressure gradients (Section 42.2.2). Hoskins' approach dispenses with such pressure terms by defining the geostrophic velocity using the full pressure field.

- (a) Show that the evolution of horizontal buoyancy gradients by the horizontal geostrophic currents can be written

$$\left[\frac{\partial}{\partial t} + \mathbf{u}_g \cdot \nabla_z \right] |\nabla_z b|^2 = \mathbf{Q} \cdot \nabla_z b. \quad (42.120)$$

Hence, horizontal buoyancy gradients grow in magnitude in regions where the horizontal buoyancy gradient projects positively onto the \mathbf{Q} -vector. Write the expression for the vector \mathbf{Q} . Hint: A general version of this result was derived in Exercise 13.2.

- (b) Show that the quasi-geostrophic ω -equation on an f -plane can be written

$$N^2 \nabla_z^2 w + f_0^2 \frac{\partial^2 w}{\partial z^2} = 2 \nabla_z \cdot \mathbf{Q}. \quad (42.121)$$

We see that the source for vertical motion is the divergence of the \mathbf{Q} -vector. This formula offers useful insight into the origin of vertical motion, with [Hoskins et al. \(1978\)](#) offering examples. For this part of the exercise, you are to fully explain the derivation of equation (42.121) and show each of the relevant steps.

- (c) The equation (42.121) for the vertical velocity takes the form

$$\mathcal{L}w = \sigma, \quad (42.122)$$

where

$$\mathcal{L} = N^2 \nabla_z^2 + f_0^2 \frac{\partial^2}{\partial z^2} \quad (42.123)$$

is a linear partial differential operator and

$$\sigma = 2 \nabla_z \cdot \mathbf{Q} \quad (42.124)$$

is a source term. The source term is a function of the geostrophic flow and the buoyancy field. For vertically stable stratification, $N^2 > 0$, characterize the differential operator \mathcal{L} according to the elliptic, hyperbolic, or parabolic classes discussed in Section 3.4. What about when $N^2 < 0$? What new physical phenomena do you expect when $N^2 < 0$?

EXERCISE 42.5: QGPV FLUX-FORM EQUATION WITH NON-CONSERVATIVE PROCESSES

Derive the quasi-geostrophic potential vorticity equation (42.84). Show all the relevant steps. Hint: the key step requires showing that

$$\nabla \cdot [(\mathbf{u}_g \cdot) \mathbf{D}] = (\mathbf{u}_g \cdot) \nabla \cdot \mathbf{D}. \quad (42.125)$$

To do so, it is useful to express \mathbf{u}_g and b_g in terms of the geostrophic streamfunction.



Part VIII

Elements of generalized vertical coordinates

Generalized vertical coordinates (GVCs) offer a mathematical framework for describing stratified fluid mechanics. Hence, they appear in many guises throughout geophysical fluid mechanics theory and numerical modeling. GVCs were introduced by [Starr \(1945\)](#) for atmospheric modeling and [Bleck \(1978\)](#) for ocean modeling. There is a growing use of GVC-based numerical ocean (e.g., [Griffies et al. \(2020\)](#)) and atmospheric models. This usage prompts the need to master their use for analysis, model formulation, and theory, thus providing motivation for this part of the book.

GVCs are both time dependent and non-orthogonal coordinates, which represents both their strength and their complexity. Both features require extra mathematical care beyond that needed with the more familiar static and orthogonal coordinates considered in Chapter 8. The key reason to favor the use of non-orthogonal GVCs, rather than a locally orthogonal coordinate set, is that gravity plays a dominant role in orienting geophysical fluid motion. Hence, it is strategically useful to decompose the equations of motion so that lateral motions are perpendicular to gravity, just like with familiar Cartesian geopotential vertical coordinates. That is, we orient horizontal motions the same regardless whether we use Cartesian coordinates or GVCs, since doing so removes the vertical pressure force from the horizontal equations of motion.

If we were to locally rotate the components of the velocity vector to be parallel and perpendicular to the surface of constant generalized vertical coordinate, as per a locally orthogonal coordinate description, then that representation would introduce a portion of the vertical pressure gradient into the equations for lateral motion. Having the vertical pressure gradient appear in each of the three components to the equations of motion makes it very difficult to isolate the hydrostatic pressure force. In turn, it makes it difficult to describe nearly all of the basic features of geophysical flows, such as the geostrophic and hydrostatic balances. We thus consider locally orthogonal generalized vertical coordinates to be of minimal use for the dynamical equations of geophysical fluid mechanics. Even so, they are useful for orienting the tracer diffusion operator within the ocean interior, as discussed in Section 50.3.

By studying the chapters in this part of the book, we develop the mathematics of generalized vertical coordinates, and then build up the kinematics and dynamics of stratified fluid mechanics using these coordinates and then encounter some example applications. Here is a brief summary of these chapters.

- MATHEMATICAL FOUNDATIONS: Chapter 43 initiates our study of the mathematics of generalized vertical coordinates. As noted already, their time dependence and their non-orthogonality present some complexity in both concept and detail. Even so, with some experience, generalized vertical coordinates can become a versatile member of our theoretical and numerical toolkit.
- Chapters 44 and 45 describes elements of fluid kinematics and dynamics using generalized vertical coordinates.
- Chapter 46 formulates the hydrostatic Boussinesq equations using buoyancy as the vertical coordinate. This chapter specializes elements from the previous chapters, and in so doing provides a mathematical and physical basis for *isopycnal* models of the ocean and isentropic models of the atmosphere.
- Chapter 47 provides further insights into eddy and mean flow decomposition, this time making use of thickness weighted averaging (TWA) for the shallow water model. We position this chapter in this part of this book since it makes use of the layered perspective of generalized vertical coordinates, particularly isopycnal coordinates. Indeed, the adiabatic stacked shallow water model is a discrete realization of the perfect fluid isopycnal equations. The TWA

formalism of this chapter offers a technically less difficult rendition of the formalism when applied to the continuously stratified isopycnal coordinates.

Mathematical foundations

In this chapter we establish the mathematical foundations of generalized vertical coordinates (GVCs) used to describe rotating and stratified fluid mechanics.

READER'S GUIDE TO THIS CHAPTER

We make use of the general tensor analysis detailed in Chapter 7. We mostly consider just the spatial tensors in this chapter, consistent with the Newtonian perspective whereby time is universal. However, since the vertical coordinate is a function of time, we follow the space-time perspective from Section 7.9 for transforming the partial derivative operator. Chapters directly relying on the material from this chapter include the fluid kinematics discussed in Chapter 44, the general vertical coordinate dynamics discussed in Chapter 45, and the tracer equation diffusion and stirring operators discussed in Chapter 50,

43.1	Introducing GVCs	986
43.1.1	Relating Cartesian and generalized vertical coordinates	986
43.2	Spatial basis vectors	988
43.2.1	A few more points on the transformation matrix	988
43.2.2	Expressions for the basis vectors	988
43.3	Basis one-forms	989
43.3.1	A few more points on the inverse transformation matrix	989
43.3.2	GVC basis one-forms	989
43.3.3	Verifying the bi-orthogonality relation	990
43.4	Triple product identities	990
43.5	Position vector	991
43.6	Transforming components of a vector	992
43.7	Velocity	993
43.7.1	Contravariant components to the velocity vector	993
43.7.2	Covariant components to the velocity vector	994
43.7.3	Introducing the material time derivative	994
43.7.4	Equivalence to the Cartesian velocity representation	994
43.8	Metric tensor	994
43.8.1	Jacobian of transformation	995
43.8.2	Covariant and contravariant representations	995
43.9	Volume element and the Levi-Civita tensor	996
43.10	Vector cross product of basis vectors	996
43.11	Partial derivative operators	996
43.11.1	Analytical derivation	996
43.11.2	Geometrical derivation	997

43.12 Material time derivative	998
43.13 Divergence of a vector and the divergence theorem	998
43.14 The diffusion operator	999
43.14.1 Continuous expression	999
43.14.2 Layer thickness weighted diffusion operator	1000
43.15 Vorticity	1000
43.15.1 The components	1001
43.15.2 Transforming from Cartesian coordinates	1001
43.16 Velocity circulation	1001

43.1 Introducing GVCs

In this chapter we present the mathematics of generalized vertical coordinates (GVC), with Figure 43.1 offering a schematic of how these coordinates monotonically partition the vertical direction. Such coordinates are of particular use for stratified fluid mechanics, where it is often convenient to make use of a vertical coordinate distinct from, but uniquely related to, the geopotential vertical coordinate, z . For example, in hydrostatic compressible fluids, such as the large-scale atmosphere, pressure is a convenient choice since it absorbs the appearance of density in many formula such as mass continuity as discussed in Section 44.10.2 and the geostrophic balance given by equation (26.1a). Hence, a natural expression of the compressible hydrostatic equations of motion make use of pressure rather than geopotential for the vertical coordinate. Two other common vertical coordinates include the isopycnal coordinate and the terrain following coordinate.

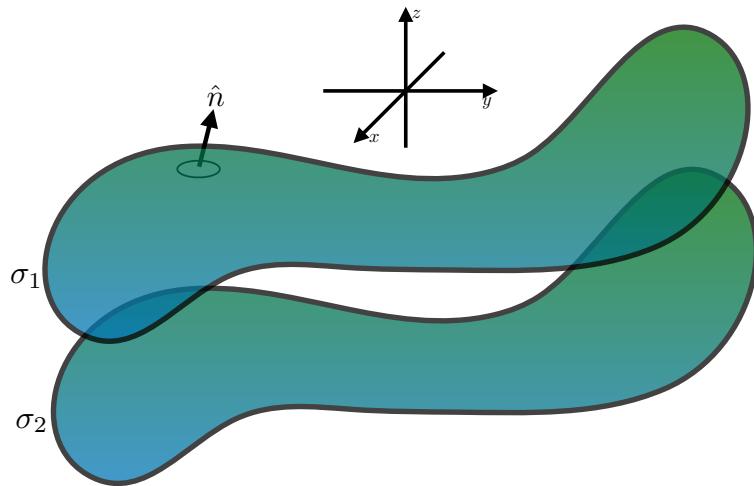


FIGURE 43.1: This stylized schematic illustrates the geometry of two surfaces of constant generalized vertical coordinate $\sigma(x, y, z, t) = \sigma_1$ and $\sigma(x, y, z, t) = \sigma_2$, here showing patches on two such surfaces. The surfaces are generally undulating in space and time yet are assumed to monotonically layer the fluid. Monotonic layering means that the surface normal, \hat{n} , always has a non-zero projection onto the vertical: $\hat{n} \cdot \hat{z} \neq 0$. That is, the surfaces never become vertical nor do they overturn. It also means that there is a 1-to-1 invertible relation between σ and geopotential, so that specifying $(x, y, \sigma(x, y, z, t))$ is sufficient to yield a unique z .

43.1.1 Relating Cartesian and generalized vertical coordinates

We make use of the symbol σ for a generalized vertical coordinate. This coordinate is *not* orthogonal to the horizontal spatial coordinates x, y . This is a central property of GVCs that

influences nearly all aspects of their calculus. To help develop the mathematics for transforming between Cartesian coordinates and GVCs, it is important to distinguish the two coordinate systems. For that purpose we write the time coordinate and spatial Cartesian coordinates according to

$$\xi^\alpha = (\xi^0, \xi^a) = (\xi^0, \xi^1, \xi^2, \xi^3) = (t, x, y, z) \text{ with } \alpha = 0, 1, 2, 3, \text{ and } a = 1, 2, 3. \quad (43.1)$$

As defined, the tensor label a runs over the spatial coordinates 1, 2, 3 whereas α also includes the time coordinate with $\alpha = 0$. The corresponding generalized vertical coordinates are denoted with an overbar

$$\xi^{\bar{\alpha}} = (\xi^{\bar{0}}, \xi^{\bar{1}}, \xi^{\bar{2}}, \xi^{\bar{3}}) = (\bar{t}, \bar{x}, \bar{y}, \sigma). \quad (43.2)$$

The 1-to-1 coordinate transformation between Cartesian and GVC coordinates is written

$$\xi^{\bar{0}} = \xi^0 \iff \bar{t} = t \quad (43.3a)$$

$$\xi^{\bar{1}} = \xi^1 \iff \bar{x} = x \quad (43.3b)$$

$$\xi^{\bar{2}} = \xi^2 \iff \bar{y} = y \quad (43.3c)$$

$$\xi^{\bar{3}} = \sigma(x, y, z, t), \quad (43.3d)$$

with the final relation expressing the generalized vertical coordinate as a function of space and time. We ordered the appearance of independent variables in σ with time in the last position, which is the conventional ordering in this book for functions of space and time even though the zeroth coordinate is time.

The coordinate transformation is invertible so that we can write

$$\xi^0 = \xi^{\bar{0}} \quad (43.4a)$$

$$\xi^1 = \xi^{\bar{1}} \quad (43.4b)$$

$$\xi^2 = \xi^{\bar{2}} \quad (43.4c)$$

$$\xi^3 = \xi^3(\bar{x}, \bar{y}, \sigma, \bar{t}) = \xi^3(x, y, \sigma, t). \quad (43.4d)$$

The relation

$$\xi^3 = \xi^3(\xi^{\bar{a}}) = \xi^3(x, t, \sigma, t) \quad (43.5)$$

provides the vertical position of a given σ surface. Since $\xi^3 = z$ one commonly writes

$$z = z(x, y, \sigma, t). \quad (43.6)$$

However, this expression is prone to confusion since the meaning of z is overloaded.¹ Namely, one meaning ascribes z to a particular value of the vertical position; i.e., $z = -100$ m. The other meaning, as on the right hand side, is for z as the vertical coordinate function of a particular σ surface, with this value a function of space, time, and σ . Learning to distinguish when z refers to a particular vertical position or as a coordinate function takes some practice, and those who routinely work with generalized vertical coordinates typically find no problem with the overloaded meanings. Indeed, after reading this chapter we should be able to hold the two meanings in our mind without confusion. As a reminder, we commonly write the vertical coordinate function as

$$z = \eta(x, y, \sigma, t), \quad (43.7)$$

¹We use the term *overloaded* as in computer science where a symbol has more than one meaning or usage.

where η is used throughout this book to represent the vertical position of a surface, such as the ocean free surface, solid-earth topography, or a generalized vertical coordinate surface. Hence, for example, the vertical position of a pressure surface of chosen value p is given by the functional relation

$$\xi^3 = z = \eta(\bar{x}, \bar{y}, p, \bar{t}) = \eta(x, y, p, t). \quad (43.8)$$

We make use of the η nomenclature where it seems useful but gradually sprinkle in more use of the $z(x, y, \sigma, t)$ notation since it is proves to be natural for many of the formulations.

43.2 Spatial basis vectors

Making use of the tensor formalism from Chapter 7, consider the transformation of the Cartesian basis vectors into their corresponding GVC representation. This transformation takes is given by

$$\vec{e}_{\bar{a}} = \Lambda_{\bar{a}}^a \vec{e}_a, \quad (43.9)$$

where the transformation matrix is

$$\Lambda_{\bar{a}}^a = \begin{bmatrix} \partial x / \partial \bar{x} & \partial x / \partial \bar{y} & \partial x / \partial \sigma \\ \partial y / \partial \bar{x} & \partial y / \partial \bar{y} & \partial y / \partial \sigma \\ \partial z / \partial \bar{x} & \partial z / \partial \bar{y} & \partial z / \partial \sigma \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ \partial \eta / \partial \bar{x} & \partial \eta / \partial \bar{y} & \partial \eta / \partial \sigma \end{bmatrix}, \quad (43.10)$$

where the second equality used our preferred notation $z = \eta(x, y, \sigma, t)$ for the vertical position of a σ surface. The diagonal unit values for the space-space components arise since a horizontal position in Cartesian and GVCs is the same and the horizontal directions are orthogonal. Likewise, the time coordinate does not change when changing \bar{x} , \bar{y} , or σ . Additionally, $\partial x / \partial \sigma = \partial y / \partial \sigma = 0$ since the horizontal position remains unchanged when moving across a GVC surface. In contrast, a non-zero value for $\partial \eta / \partial \bar{x}$ and $\partial \eta / \partial \bar{y}$ arise since we generally change vertical position when moving horizontally along a sloped σ surface. Finally, the element $\partial \eta / \partial \sigma$ is nonzero due to vertical stratification of the fluid when represented using general vertical coordinates.

43.2.1 A few more points on the transformation matrix

To further detail how to produce elements of the transformation matrix (43.10), it is crucial to ensure that the proper variables are held fixed when performing the partial derivatives. For example, consider the top row where we compute

$$\Lambda_{\bar{a}}^1 = [[\partial x / \partial \bar{x}]_{\bar{y}, \sigma} \quad [\partial x / \partial \bar{y}]_{\bar{x}, \sigma} \quad [\partial x / \partial \sigma]_{\bar{x}, \bar{y}}] \quad (43.11)$$

Since $x = \bar{x}$, all elements vanish except for the first. Namely, $[\partial x / \partial \bar{y}]_{\bar{x}, \sigma} = 0$ since x cannot change when \bar{x} is fixed. The same idea leads to the results for y derivatives.

43.2.2 Expressions for the basis vectors

Use of the transformation matrix (43.10) renders the spatial components of the GVC basis vectors

$$\vec{e}_{\bar{1}} = \hat{\mathbf{x}} + \hat{\mathbf{z}} (\partial \eta / \partial \bar{x}) \quad (43.12a)$$

$$\vec{e}_{\bar{2}} = \hat{\mathbf{y}} + \hat{\mathbf{z}} (\partial \eta / \partial \bar{y}) \quad (43.12b)$$

$$\vec{e}_{\bar{3}} = \hat{\mathbf{z}} (\partial \eta / \partial \sigma). \quad (43.12c)$$

The basis vectors $\vec{e}_{\bar{1}}$ and $\vec{e}_{\bar{2}}$ have a vertical component due to sloping GVC surfaces. These basis vectors lie within the tangent plane of the GVC surface. The basis vector $\vec{e}_{\bar{3}}$ is purely vertical and has a non-unit magnitude due to the inverse vertical stratification, $\partial \eta / \partial \sigma$. The left panel of Figure 43.2 illustrates the basis vectors.

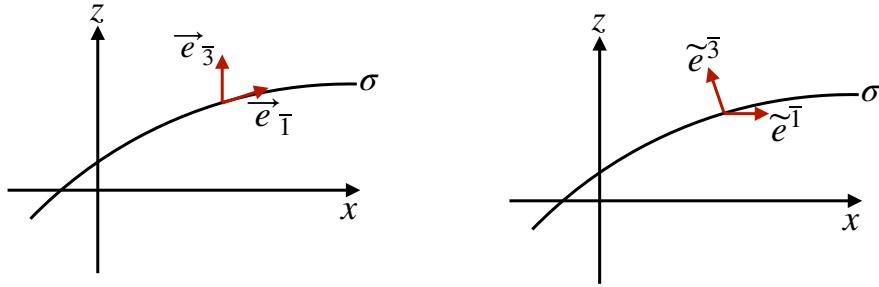


FIGURE 43.2: Illustrating the basis vectors (left panel) and basis one-forms (right panel) for generalized vertical coordinates. The \vec{e}_3 basis vector is vertical whereas \vec{e}_1 and \vec{e}_2 lie within the tangent plane to the σ surface. As a complement, the basis one-form \tilde{e}^3 is normal to the σ surface whereas the basis one-forms \tilde{e}^1 and \tilde{e}^2 are horizontal.

43.3 Basis one-forms

The basis one-forms are obtained by transforming from Cartesian into GVCs through use of the inverse transformation

$$\tilde{e}^{\bar{a}} = \Lambda_{\bar{a}}^{\bar{a}} \tilde{e}^a, \quad (43.13)$$

where the inverse transformation matrix takes the form

$$\Lambda^{\bar{a}}_a = \begin{bmatrix} \partial \bar{x}/\partial x & \partial \bar{x}/\partial y & \partial \bar{x}/\partial z \\ \partial \bar{y}/\partial x & \partial \bar{y}/\partial y & \partial \bar{y}/\partial z \\ \partial \sigma/\partial x & \partial \sigma/\partial y & \partial \sigma/\partial z \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ \partial \sigma/\partial x & \partial \sigma/\partial y & \partial \sigma/\partial z \end{bmatrix}. \quad (43.14)$$

As for the transformation matrix (43.10), the unit diagonal values arise since a horizontal position in Cartesian and GVCs is the same and the horizontal directions are orthogonal. Likewise, $\partial \bar{x}/\partial z = \partial \bar{y}/\partial z = 0$ since the horizontal position on a GVC surface remains unchanged when moving across a depth surface. The nonzero values for $\partial \sigma/\partial x$, $\partial \sigma/\partial y$, and $\partial \sigma/\partial z$, arise in the presence of horizontal and vertical stratification of the generalized vertical coordinate.

43.3.1 A few more points on the inverse transformation matrix

When computing elements of the inverse transformation matrix (43.14), it is crucial to ensure that the proper variables are held fixed. For example, consider the top row where we compute

$$\Lambda^{\bar{1}}_a = [[\partial \bar{x}/\partial x]_{y,z} \quad [\partial \bar{x}/\partial y]_{x,z} \quad [\partial \bar{x}/\partial z]_{x,y}]. \quad (43.15)$$

Just as for the transformation matrix (43.11), since $x = \bar{x}$, all but the first element vanish in equation (43.15). Namely, $[\partial \bar{x}/\partial y]_{x,z} = 0$ since the \bar{x} cannot change when x is fixed. The same idea holds for the \bar{y} row.

43.3.2 GVC basis one-forms

Use of the inverse transformation matrix (43.14) renders the spatial components of the GVC basis one-forms

$$\tilde{e}^{\bar{1}} = \hat{x} \quad (43.16a)$$

$$\tilde{e}^{\bar{2}} = \hat{y} \quad (43.16b)$$

$$\tilde{e}^{\bar{3}} = \tilde{e}^a \partial_a \sigma = \hat{x} (\partial \sigma / \partial x) + \hat{y} (\partial \sigma / \partial y) + \hat{z} (\partial \sigma / \partial z) = \nabla \sigma. \quad (43.16c)$$

The left panel of Figure 43.2 illustrates the basis one-forms.

43.3.3 Verifying the bi-orthogonality relation

The basis one-forms satisfy the bi-orthogonality relation (7.25) with the basis vectors

$$\vec{e}^{\bar{a}} \cdot \vec{e}_{\bar{b}} = \delta_{\bar{b}}^{\bar{a}}. \quad (43.17)$$

This identity is trivial to verify for all $\bar{a} = 1, 2, 3$.

43.4 Triple product identities

We find various occasions to make use of a suite of triple product identities that hold for GVCs. For this purpose we write σ as a composite function as in Section 7.9.3

$$\sigma = \sigma(x, y, z, t) = \sigma[x, y, z(\bar{t}, \bar{x}, \bar{y}, \sigma), t], \quad (43.18)$$

with $\eta(\bar{x}, \bar{y}, \sigma, \bar{t})$ written as $z(\bar{x}, \bar{y}, \sigma, \bar{t})$ as it here eases the manipulations. Use of the chain rule leads to the space-time differential increment

$$d\sigma = dt \left[\frac{\partial \sigma}{\partial t} \right]_{x,y,z} + dx \left[\frac{\partial \sigma}{\partial x} \right]_{t,y,z} + dy \left[\frac{\partial \sigma}{\partial y} \right]_{t,x,z} + dz \left[\frac{\partial \sigma}{\partial z} \right]_{t,x,y}. \quad (43.19)$$

Likewise, writing $z = z[\bar{t}, \bar{x}, \bar{y}, \sigma]$ leads to the space-time differential increment dz

$$dz = d\bar{t} \left[\frac{\partial z}{\partial \bar{t}} \right]_{\bar{x}, \bar{y}, \sigma} + d\bar{x} \left[\frac{\partial z}{\partial \bar{x}} \right]_{\bar{t}, \bar{y}, \sigma} + d\bar{y} \left[\frac{\partial z}{\partial \bar{y}} \right]_{\bar{t}, \bar{x}, \sigma} + d\sigma \left[\frac{\partial z}{\partial \sigma} \right]_{\bar{t}, \bar{x}, \bar{y}}. \quad (43.20)$$

We note the identities

$$\left[\frac{\partial \sigma}{\partial z} \right]_{t,x,y} \left[\frac{\partial z}{\partial \sigma} \right]_{\bar{t}, \bar{x}, \bar{y}} = 1 \quad d\bar{t} = dt \quad d\bar{x} = dx \quad d\bar{y} = dy, \quad (43.21)$$

which follow since $t = \bar{t}$, $x = \bar{x}$, and $y = \bar{y}$. Substituting equation (43.20) into equation (43.19) and making use of the identities (43.21) yields

$$0 = dt \left[\left[\frac{\partial \sigma}{\partial t} \right]_{x,y,z} + \left[\frac{\partial \sigma}{\partial z} \right]_{t,x,y} \left[\frac{\partial z}{\partial \bar{t}} \right]_{\bar{x}, \bar{y}, \sigma} \right] \\ + dx \left[\left[\frac{\partial \sigma}{\partial x} \right]_{t,y,z} + \left[\frac{\partial \sigma}{\partial z} \right]_{t,x,y} \left[\frac{\partial z}{\partial \bar{x}} \right]_{\bar{t}, \bar{y}, \sigma} \right] + dy \left[\left[\frac{\partial \sigma}{\partial y} \right]_{t,x,z} + \left[\frac{\partial \sigma}{\partial z} \right]_{t,x,y} \left[\frac{\partial z}{\partial \bar{y}} \right]_{\bar{t}, \bar{x}, \sigma} \right]. \quad (43.22)$$

For this equation to hold with general increments dt , dx , and dy requires that each bracketed term vanish, which in turn leads to the following set of triple product identities²

$$\left[\frac{\partial \sigma}{\partial z} \right]_{t,x,y} \left[\frac{\partial z}{\partial \bar{t}} \right]_{\bar{x}, \bar{y}, \sigma} = - \left[\frac{\partial \sigma}{\partial t} \right]_{x,y,z} \quad (43.23a)$$

$$\left[\frac{\partial \sigma}{\partial z} \right]_{t,x,y} \left[\frac{\partial z}{\partial \bar{x}} \right]_{\bar{t}, \bar{y}, \sigma} = - \left[\frac{\partial \sigma}{\partial x} \right]_{t,y,z} \quad (43.23b)$$

$$\left[\frac{\partial \sigma}{\partial z} \right]_{t,x,y} \left[\frac{\partial z}{\partial \bar{y}} \right]_{\bar{t}, \bar{x}, \sigma} = - \left[\frac{\partial \sigma}{\partial y} \right]_{t,x,z} \quad (43.23c)$$

²These identities are directly analogous to the Maxwell relations from thermodynamics, with an introduction in Section 21.6 and full details in the book by [Callen \(1985\)](#).

If the vertical stratification, $\partial\sigma/\partial z$, is non-zero, the triple product identities are equivalent to

$$\left[\frac{\partial z}{\partial \bar{t}} \right]_{\bar{x}, \bar{y}, \sigma} = - \frac{[\partial\sigma/\partial t]_{x,y,z}}{[\partial\sigma/\partial z]_{t,x,y}} = - \left[\frac{\partial\sigma}{\partial t} \right]_{x,y,z} \left[\frac{\partial z}{\partial\sigma} \right]_{t,x,y} \quad (43.24a)$$

$$\left[\frac{\partial z}{\partial \bar{x}} \right]_{\bar{t}, \bar{y}, \sigma} = - \frac{[\partial\sigma/\partial x]_{t,y,z}}{[\partial\sigma/\partial z]_{t,x,y}} = - \left[\frac{\partial\sigma}{\partial x} \right]_{t,y,z} \left[\frac{\partial z}{\partial\sigma} \right]_{t,x,y} \quad (43.24b)$$

$$\left[\frac{\partial z}{\partial \bar{y}} \right]_{\bar{t}, \bar{x}, \sigma} = - \frac{[\partial\sigma/\partial y]_{t,x,z}}{[\partial\sigma/\partial z]_{t,x,y}} = - \left[\frac{\partial\sigma}{\partial y} \right]_{t,x,z} \left[\frac{\partial z}{\partial\sigma} \right]_{t,x,y}. \quad (43.24c)$$

Since $t = \bar{t}$, $x = \bar{x}$, and $y = \bar{y}$ we can write these identities in the more succinct form

$$\left[\frac{\partial z}{\partial \bar{t}} \right]_\sigma = \left[\frac{\partial \eta}{\partial \bar{t}} \right]_\sigma = - \frac{[\partial\sigma/\partial t]_z}{[\partial\sigma/\partial z]} \quad (43.25a)$$

$$\left[\frac{\partial z}{\partial \bar{x}} \right]_\sigma = \left[\frac{\partial \eta}{\partial \bar{x}} \right]_\sigma = - \frac{[\partial\sigma/\partial x]_z}{[\partial\sigma/\partial z]} \quad (43.25b)$$

$$\left[\frac{\partial z}{\partial \bar{y}} \right]_\sigma = \left[\frac{\partial \eta}{\partial \bar{y}} \right]_\sigma = - \frac{[\partial\sigma/\partial y]_z}{[\partial\sigma/\partial z]}, \quad (43.25c)$$

where we reintroduced the notation $\eta(x, y, \sigma, t) = z(x, y, \sigma, t)$. These identities are quite useful for manipulating equations involving GVCs. In particular, equations (43.25b) and (43.25c) provide alternate expressions for the slope of σ isosurfaces relative to the horizontal plane (see Section 43.11).

43.5 Position vector

We are familiar with locating a point in space using Cartesian coordinates as in Figure 1.1. What about specifying the position using GVCs? We can do so by making use of the basis vectors (43.12a)-(43.12c) so that the position of an arbitrary point in space is given by

$$\mathcal{P} = \xi^{\bar{a}} \vec{e}_{\bar{a}} \quad (43.26a)$$

$$= \bar{x} [\hat{x} + (\partial\eta/\partial\bar{x}) \hat{z}] + \bar{y} [\hat{y} + (\partial\eta/\partial\bar{y}) \hat{z}] + \sigma (\partial\eta/\partial\sigma) \hat{z} \quad (43.26b)$$

$$= \hat{x} \bar{x} + \hat{y} \bar{y} + \hat{z} [\bar{x} (\partial\eta/\partial\bar{x}) + \bar{y} (\partial\eta/\partial\bar{y}) + \sigma (\partial\eta/\partial\sigma)] \quad (43.26c)$$

$$= \hat{x} \bar{x} + \hat{y} \bar{y} + \hat{z} \xi^{\bar{a}} \partial_{\bar{a}} \eta. \quad (43.26d)$$

We identify the following properties as a means to help understand these expressions, with Figure 43.3 offering a schematic.

- The expression (43.26b) has horizontal positions \bar{x} and \bar{y} multiplying the basis vectors $\vec{e}_{\bar{x}}$ and $\vec{e}_{\bar{y}}$, with these vectors oriented parallel to a surface of constant GVC as in Figure 43.3. Likewise, the third term, $\sigma (\partial\eta/\partial\sigma) \hat{z}$, positions the point vertically according to the value of the GVC and its inverse stratification.
- Consider the case of $\bar{y} = 0$ so that

$$\mathcal{P} = \bar{x} \hat{x} + \hat{z} [\bar{x} (\partial\eta/\partial\bar{x}) + \sigma (\partial\eta/\partial\sigma)] \quad (43.27a)$$

$$= \bar{x} \hat{x} + \hat{z} (\partial\eta/\partial\sigma) [\bar{x} (\partial\sigma/\partial z)_x (\partial\eta/\partial\bar{x})_\sigma + \sigma] \quad (43.27b)$$

$$= \bar{x} \hat{x} + \hat{z} (\partial\eta/\partial\sigma) [-\bar{x} (\partial\sigma/\partial x)_z + \sigma], \quad (43.27c)$$

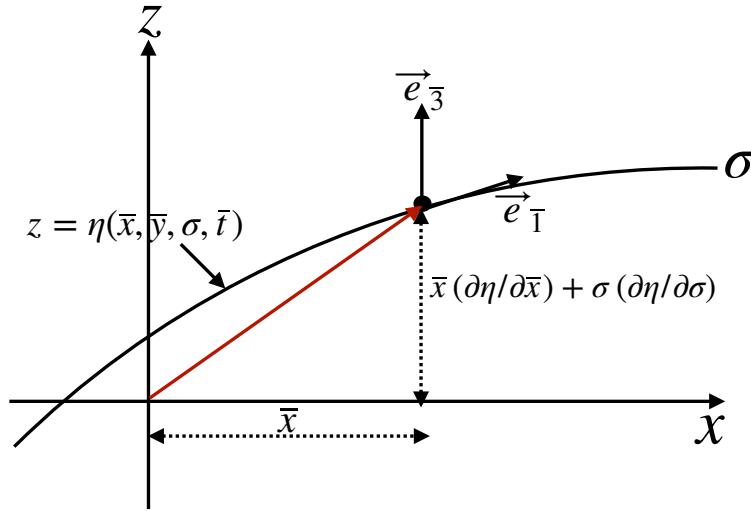


FIGURE 43.3: The position of a point in space as represented using GVCs following equation (43.27a). For this example, $\bar{y} = 0$ so that the horizontal position is determined by the coordinate $\bar{x} = x$, whereas the vertical position is determined by $\bar{x} (\partial z / \partial \bar{x}) + \sigma (\partial z / \partial \sigma) = \bar{x} (\partial \eta / \partial \bar{x}) + \sigma (\partial \eta / \partial \sigma)$.

where we used the triple product identity (43.25b) for the final equality. Consequently, a horizontal position vector is realized using GVC coordinates with $\sigma = \bar{x} (\partial \sigma / \partial x)$. That is, a horizontal position vector crosses surfaces of constant GVC when the GVC surface has a nonzero horizontal slope.

- The projection of the position vector onto the basis one-forms leads to

$$\mathcal{P} \cdot \tilde{e}^{\bar{b}} = \xi^{\bar{a}} \tilde{e}_{\bar{a}} \cdot \tilde{e}^{\bar{b}} = \xi^{\bar{b}}. \quad (43.28)$$

This result follows from the orthogonality relation (7.25). So the projection of the position vector onto a basis one-form picks out the corresponding coordinate value.

- Equation (43.7) provides the spatial dependence for the vertical position of the surface of constant GVC

$$z = z(\xi^{\bar{a}}) = \eta(\xi^{\bar{a}}). \quad (43.29)$$

At any particular time instance we can perform a Taylor series about a reference geopotential $z_0 = \eta_0$, so that

$$\eta(\xi^{\bar{a}}) \approx \eta_0 + \xi^{\bar{a}} \partial_{\bar{a}} \eta. \quad (43.30)$$

We can thus write the position (43.26d) in the form

$$\mathcal{P} = \hat{x} \bar{x} + \hat{y} \bar{y} + \hat{z} [\eta - \eta_0]. \quad (43.31)$$

Taking the reference geopotential as $\eta_0 = 0$ recovers the Cartesian expression. Since the position vector is a geometric object, it is reassuring that the GVC representation is the same as the Cartesian representation; it is merely a reorganization of the basis vectors and corresponding coordinates.

43.6 Transforming components of a vector

Consider a vector field \vec{F} with Cartesian representation

$$\vec{F} = \mathbf{F} = F^a \vec{e}_a = F^x \hat{x} + F^y \hat{y} + F^z \hat{z}. \quad (43.32)$$

The corresponding GVC components are related through the transformation matrix

$$F^{\bar{a}} = \Lambda_{\bar{a}}^{\bar{a}} F^a. \quad (43.33)$$

Making use of the transformation matrix (43.14) yields the relations between GVC components and Cartesian components

$$F^{\bar{1}} = F^1 \quad \text{and} \quad F^{\bar{2}} = F^2 \quad \text{and} \quad F^{\bar{3}} = \nabla \sigma \cdot \mathbf{F}, \quad (43.34)$$

where we wrote

$$\nabla \sigma \cdot \mathbf{F} = (\partial \sigma / \partial x) F^1 + (\partial \sigma / \partial y) F^2 + (\partial \sigma / \partial z) F^3. \quad (43.35)$$

The vector field thus can be represented in GVC coordinates as

$$\vec{F} = F^{\bar{a}} \vec{e}_{\bar{a}} = F^1 \vec{e}_{\bar{1}} + F^2 \vec{e}_{\bar{2}} + (\nabla \sigma \cdot \mathbf{F}) \vec{e}_{\bar{3}}. \quad (43.36)$$

Similarly, the covariant components transform as $F_{\bar{a}} = \Lambda_{\bar{a}}^a F_a$, where use of the inverse transformation matrix (43.14) renders

$$F_{\bar{1}} = F_1 + (\partial z / \partial \bar{x}) F_3 = F_1 + (\partial \eta / \partial \bar{x}) F_3 \quad (43.37a)$$

$$F_{\bar{2}} = F_2 + (\partial z / \partial \bar{y}) F_3 = F_2 + (\partial \eta / \partial \bar{y}) F_3 \quad (43.37b)$$

$$F_{\bar{3}} = (\partial z / \partial \sigma) F_3 = (\partial \eta / \partial \sigma) F_3, \quad (43.37c)$$

and the expression for the vector field

$$\vec{F} = F_{\bar{a}} \vec{e}^{\bar{a}} = [F_1 + (\partial \eta / \partial \bar{x}) F_3] \vec{e}^{\bar{1}} + [F_2 + (\partial \eta / \partial \bar{y}) F_3] \vec{e}^{\bar{2}} + (\partial \eta / \partial \sigma) F_3 \vec{e}^{\bar{3}}. \quad (43.38)$$

Recall also that for Cartesian coordinates the contravariant and covariant components to a vector are identical: $F^a = F_a$.

43.7 Velocity

As an example of the results from Section 43.6, we here represent the velocity vector, considering both covariant and contravariant representations. As for the position vector detailed in Section 43.5, we are assured that both representations lead to the same velocity vector since the velocity is an objective geometric object (i.e., an arrow with a magnitude). In Section 43.7.4 we verify that the transformation formalism indeed respects this equivalence, with the GVC representations equivalent to the Cartesian representation

$$\vec{v} = u \hat{x} + v \hat{y} + w \hat{z}. \quad (43.39)$$

43.7.1 Contravariant components to the velocity vector

Following Section 43.6, we have the contravariant velocity components

$$v^{\bar{1}} = u \quad \text{and} \quad v^{\bar{2}} = v \quad \text{and} \quad v^{\bar{3}} = \mathbf{v} \cdot \nabla \sigma. \quad (43.40)$$

Use of the basis vectors (43.12a)-(43.12c) then leads to

$$\vec{v} = v^{\bar{a}} \vec{e}_{\bar{a}} \quad (43.41a)$$

$$= u \vec{e}_{\bar{1}} + v \vec{e}_{\bar{2}} + (\mathbf{v} \cdot \nabla \sigma) \vec{e}_{\bar{3}} \quad (43.41b)$$

$$= u [\hat{x} + (\partial \eta / \partial \bar{x}) \hat{z}] + v [\hat{y} + (\partial \eta / \partial \bar{y}) \hat{z}] + (\mathbf{v} \cdot \nabla \sigma) (\partial \eta / \partial \sigma) \hat{z}. \quad (43.41c)$$

43.7.2 Covariant components to the velocity vector

The covariant velocity components are given by

$$v_{\bar{1}} = u + (\partial\eta/\partial\bar{x}) w \quad \text{and} \quad v_{\bar{2}} = v + (\partial\eta/\partial\bar{y}) w \quad \text{and} \quad v_{\bar{3}} = (\partial\eta/\partial\sigma) w. \quad (43.42)$$

The one-form basis (43.16a)–(43.16c) thus leads to the velocity vector

$$\vec{v} = v_{\bar{a}} \tilde{e}^{\bar{a}} = [u + (\partial\eta/\partial\bar{x}) w] \hat{x} + [v + (\partial\eta/\partial\bar{y}) w] \hat{y} + w (\partial\eta/\partial\sigma) \nabla\sigma. \quad (43.43)$$

43.7.3 Introducing the material time derivative

The material evolution for the generalized vertical coordinate can be written

$$\frac{D\sigma}{Dt} = \frac{\partial\sigma}{\partial t} + \mathbf{v} \cdot \nabla\sigma = \dot{\sigma}, \quad (43.44)$$

with $\dot{\sigma}$ symbolizing any process contributing to motion across σ isosurfaces (as fully explained in Section 44.4). Using the expression (43.44) in the velocity vector expression (43.41c) leads to

$$\vec{v} = u [\hat{x} + (\partial\eta/\partial\bar{x}) \hat{z}] + v [\hat{y} + (\partial\eta/\partial\bar{y}) \hat{z}] + (\mathbf{v} \cdot \nabla\sigma) (\partial\eta/\partial\sigma) \hat{z}. \quad (43.45a)$$

$$= u [\hat{x} + (\partial\eta/\partial\bar{x}) \hat{z}] + v [\hat{y} + (\partial\eta/\partial\bar{y}) \hat{z}] + (\dot{\sigma} - \partial\sigma/\partial t) (\partial\eta/\partial\sigma) \hat{z} \quad (43.45b)$$

$$= u \hat{x} + v \hat{y} + [\partial\eta/\partial\bar{t} + \mathbf{u} \cdot \nabla_\sigma z + (\partial\eta/\partial\sigma)\dot{\sigma}] \hat{z}, \quad (43.45c)$$

where the final equality made use of the triple product (43.24a): $(\partial\sigma/\partial t) (\partial\eta/\partial\sigma) = -\partial\eta/\partial\bar{t}$. In the steady state and in the absence of material changes to σ , the three dimensional flow lies within a surface of constant σ , whereby $\mathbf{v} \cdot \nabla\sigma = 0$ and

$$\vec{v} = u [\hat{x} + (\partial\eta/\partial\bar{x}) \hat{z}] + v [\hat{y} + (\partial\eta/\partial\bar{y}) \hat{z}] \quad \text{if } \partial_t\sigma = 0 \text{ and } \dot{\sigma} = 0. \quad (43.46)$$

However, in general there are transient fluctuations and material changes to σ so that $\mathbf{v} \cdot \nabla\sigma \neq 0$.

43.7.4 Equivalence to the Cartesian velocity representation

Use of the triple product identities (43.25b)–(43.25c) allows us to manipulate both expressions (43.41c) and (43.43) to recover the Cartesian expression

$$\vec{v} = u \hat{x} + v \hat{y} + w \hat{z}. \quad (43.47)$$

Another way to see this identity is to note that in equation (43.45c), the vertical component is an expression for the material time derivative of the vertical position

$$w = \frac{Dz}{Dt} = \frac{\partial\eta}{\partial\bar{t}} + \mathbf{u} \cdot \nabla_\sigma \eta + \frac{\partial\eta}{\partial\sigma} \dot{\sigma}. \quad (43.48)$$

We derive this identity in Section 44.5 where we discuss further kinematic results using GVCs.

43.8 Metric tensor

Recall from Section 7.1 that we make use of a metric tensor to measure the distance between two points in space. The GVC representation of the metric tensor is given by

$$g_{\bar{a}\bar{b}} = \vec{e}_{\bar{a}} \cdot \vec{e}_{\bar{b}} = \begin{bmatrix} 1 + (\partial\eta/\partial\bar{x})^2 & (\partial\eta/\partial\bar{x})(\partial\eta/\partial\bar{y}) & (\partial\eta/\partial\bar{x})(\partial\eta/\partial\sigma) \\ (\partial\eta/\partial\bar{x})(\partial\eta/\partial\bar{y}) & 1 + (\partial\eta/\partial\bar{y})^2 & (\partial\eta/\partial\bar{y})(\partial\eta/\partial\sigma) \\ (\partial\eta/\partial\bar{x})(\partial\eta/\partial\sigma) & (\partial\eta/\partial\bar{y})(\partial\eta/\partial\sigma) & (\partial\eta/\partial\sigma)^2 \end{bmatrix}, \quad (43.49)$$

and the GVC representation of the inverse metric tensor is given by the somewhat simpler form

$$g^{\bar{a}\bar{b}} = \tilde{e}^{\bar{a}} \cdot \tilde{e}^{\bar{b}} = \begin{bmatrix} 1 & 0 & \partial\sigma/\partial x \\ 0 & 1 & \partial\sigma/\partial y \\ \partial\sigma/\partial x & \partial\sigma/\partial y & |\nabla\sigma|^2 \end{bmatrix}. \quad (43.50)$$

Proof that $g^{\bar{a}\bar{b}} g_{\bar{b}\bar{c}} = \delta_{\bar{c}}^{\bar{a}}$ requires use of the triple product identities (43.25b) and (43.25c).

43.8.1 Jacobian of transformation

The determinant of the GVC representation of the metric tensor (43.49) is

$$\det(g_{\bar{a}\bar{b}}) = (\partial z/\partial\sigma)^2 = (\partial\eta/\partial\sigma)^2 \quad (43.51)$$

so that the Jacobian of transformation (Section 7.5) is the specific thickness

$$\frac{\partial(x, y, z)}{\partial(\bar{x}, \bar{y}, \sigma)} = \frac{\partial z}{\partial\sigma} = \frac{\partial\eta}{\partial\sigma}. \quad (43.52)$$

The coordinate transformation from Cartesian to generalized vertical is invertible only so long as the Jacobian remains nonzero and single-signed, meaning the fluid retains a monotonic vertical stratification of GVC surfaces. The invertible relation between z and σ means that each point in the vertical can be uniquely specified by either of the two vertical coordinates, z or σ . For example, the Jacobian for pressure as the generalized vertical coordinate in a hydrostatic fluid is given by³

$$\frac{\partial z}{\partial\sigma} = \frac{\partial z}{\partial p} = -\frac{1}{\rho g}, \quad (43.53)$$

which is indeed single-signed since the mass density is always positive.

43.8.2 Covariant and contravariant representations

The metric tensor allows us to convert between the covariant and contravariant representations of a vector via the identity (Section 7.2.3)

$$F_{\bar{a}} = g_{\bar{a}\bar{b}} F^{\bar{b}}. \quad (43.54)$$

We use triple product identities (43.25b)-(43.25c) to verify that this relation agrees with the transformation matrix approach detailed in Section 43.6. For example,

$$F_{\bar{1}} = g_{\bar{1}\bar{b}} F^{\bar{b}} \quad (43.55a)$$

$$= [1 + (\partial\eta/\partial\bar{x})^2] F^{\bar{1}} + (\partial\eta/\partial\bar{x})(\partial\eta/\partial\bar{y}) F^{\bar{2}} + (\partial\eta/\partial\bar{x})(\partial\eta/\partial\sigma) F^{\sigma} \quad (43.55b)$$

$$= [1 + (\partial\eta/\partial\bar{x})^2] F^1 + (\partial\eta/\partial\bar{x})(\partial\eta/\partial\bar{y}) F^2 + (\partial\eta/\partial\bar{x})(\partial\eta/\partial\sigma) \nabla\sigma \cdot \mathbf{F} \quad (43.55c)$$

$$= F^1 + (\partial\eta/\partial\bar{x}) F^3 \quad (43.55d)$$

$$= F_1 + (\partial\eta/\partial\bar{x}) F_3, \quad (43.55e)$$

where the final equality holds since $F^1 = F_1$ and $F^3 = F_3$ for Cartesian tensor components.

³We derive the hydrostatic balance in Section 23.3.

43.9 Volume element and the Levi-Civita tensor

The square root of the determinant of the metric tensor (43.49) is

$$\sqrt{\det(g_{\bar{a}\bar{b}})} = \partial z / \partial \sigma = \partial \eta / \partial \sigma \quad (43.56)$$

so that the volume element (Section 7.5) is

$$dV = (\partial z / \partial \sigma) dx dy d\sigma. \quad (43.57)$$

The covariant Levi-Civita tensor (Section 7.7) has the GVC representations

$$\varepsilon_{\bar{a}\bar{b}\bar{c}} = (\partial z / \partial \sigma) \epsilon_{\bar{a}\bar{b}\bar{c}} \quad \varepsilon^{\bar{a}\bar{b}\bar{c}} = (\partial z / \partial \sigma)^{-1} \epsilon^{\bar{a}\bar{b}\bar{c}} \quad (43.58)$$

where ϵ is the permutation symbol introduced in Section 1.4.1 with its components independent of coordinate representation.

43.10 Vector cross product of basis vectors

We now verify the relation (7.79) for the cross product of two basis vectors using GVCs

$$\vec{e}_{\bar{a}} \wedge \vec{e}_{\bar{b}} = \varepsilon_{\bar{a}\bar{b}\bar{c}} \tilde{e}^{\bar{c}} \implies \vec{e}_{\bar{a}} \wedge \vec{e}_{\bar{b}} = (\partial z / \partial \sigma) \epsilon_{\bar{a}\bar{b}\bar{c}} \tilde{e}^{\bar{c}}. \quad (43.59)$$

Making use of the basis vectors from Section 43.2 and the basis one-forms from Section 43.3 renders

$$\vec{e}_{\bar{x}} \wedge \vec{e}_{\bar{y}} = \hat{z} - \hat{x} (\partial \eta / \partial \bar{x}) - \hat{y} (\partial \eta / \partial \bar{y}) = (\partial z / \partial \sigma) \nabla \sigma = \varepsilon_{\bar{x}\bar{y}\sigma} \tilde{e}^{\sigma} \quad (43.60a)$$

$$\vec{e}_{\bar{y}} \wedge \vec{e}_{\sigma} = \hat{x} (\partial z / \partial \sigma) = \varepsilon_{\bar{y}\sigma\bar{x}} \tilde{e}^{\bar{x}} \quad (43.60b)$$

$$\vec{e}_{\sigma} \wedge \vec{e}_{\bar{x}} = \hat{y} (\partial z / \partial \sigma) = \varepsilon_{\sigma\bar{x}\bar{y}} \tilde{e}^{\bar{y}}. \quad (43.60c)$$

43.11 Partial derivative operators

We here consider the partial derivative operators and their transformation between coordinate systems. These identities are used throughout GVC calculus. Given the importance of these expressions, we offer two derivations. Notably, the geometric derivation in Section 43.11.2 requires minimal use of the previous tensor formalism.

43.11.1 Analytical derivation

The partial derivative operators in GVCs are computed via $\partial_{\bar{a}} = \Lambda_{\bar{a}}^a \partial_a$. Including also the time component leads to the relations

$$\partial_{\bar{t}} = \partial_t + (\partial z / \partial \bar{t}) \partial_z = \partial_t + (\partial \eta / \partial \bar{t}) \partial_z \quad (43.61a)$$

$$\partial_{\bar{x}} = \partial_x + (\partial z / \partial \bar{x}) \partial_z = \partial_x + (\partial \eta / \partial \bar{x}) \partial_z \quad (43.61b)$$

$$\partial_{\bar{y}} = \partial_y + (\partial z / \partial \bar{y}) \partial_z = \partial_y + (\partial \eta / \partial \bar{y}) \partial_z \quad (43.61c)$$

$$\partial_{\sigma} = (\partial z / \partial \sigma) \partial_z = (\partial \eta / \partial \sigma) \partial_z. \quad (43.61d)$$

We can make use of the triple product identities (43.25b) and (43.25c) to express the slope of a constant GVC surface in the equivalent manners

$$\mathbf{S} = \nabla_\sigma \eta = \nabla_\sigma z = -(\partial\sigma/\partial z)^{-1} \nabla_z \sigma \quad (43.62)$$

where we introduced the standard shorthand notation

$$\nabla_\sigma = \hat{\mathbf{x}} \partial/\partial \bar{x} + \hat{\mathbf{y}} \partial/\partial \bar{y} \quad \text{and} \quad \nabla_z = \hat{\mathbf{x}} \partial/\partial x + \hat{\mathbf{y}} \partial/\partial y. \quad (43.63)$$

It is common to transform between the horizontal gradient operators, in which case we write

$$\nabla_\sigma = \nabla_z + (\nabla_\sigma z) \partial_z \equiv \nabla_z + \mathbf{S} \partial_z. \quad (43.64)$$

We emphasize that ∇_σ is merely a shorthand for the two partial derivative operators and that it only has components in the horizontal directions.

43.11.2 Geometrical derivation

We provide a geometric derivation for the lateral derivative operator that complements the previous analytical derivation. This operator is computed by taking the difference of a function along surfaces of constant generalized vertical coordinate, but with the lateral distance computed in the horizontal direction as show in Figure 43.4. This feature of the horizontal derivative operator is a key aspect of the GVCs' non-orthogonality.

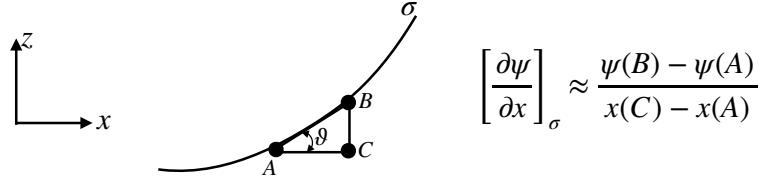


FIGURE 43.4: A surface of constant generalized vertical coordinate, σ , along with a local tangent plane with a slope $\tan \vartheta$ with respect to the horizontal plane. This figure illustrates the identities (43.66a)-(43.66d), with these identities relating a lateral derivative taken along the GVC surface to horizontal and vertical derivatives taken along orthogonal Cartesian axes.

Consider the geometry shown in Figure 43.4, which shows a generalized vertical coordinate surface (constant σ surface) along with a sample tangent plane with a slope

$$S^x = \frac{\text{rise}}{\text{run}} = \tan \vartheta = \frac{z(B) - z(C)}{x(C) - x(A)} \approx \left[\frac{\partial z}{\partial x} \right]_{\sigma} = -\frac{(\partial \sigma/\partial x)_z}{(\partial \sigma/\partial z)} \quad (43.65)$$

relative to the horizontal. We readily verify the following identities based on finite difference operations for an arbitrary function

$$\left[\frac{\partial \psi}{\partial x} \right]_{\sigma} \approx \frac{\psi(B) - \psi(A)}{x(C) - x(A)} \quad (43.66a)$$

$$= \frac{\psi(C) - \psi(A)}{x(C) - x(A)} + \frac{\psi(B) - \psi(C)}{x(C) - x(A)} \quad (43.66b)$$

$$= \frac{\psi(C) - \psi(A)}{x(C) - x(A)} + \left[\frac{z(B) - z(C)}{x(C) - x(A)} \right] \frac{\psi(B) - \psi(C)}{z(B) - z(C)} \quad (43.66c)$$

$$= \left[\frac{\partial \psi}{\partial x} \right]_z + S^x \left[\frac{\partial \psi}{\partial z} \right]_x. \quad (43.66d)$$

Taking the continuum limit then leads to the relations between horizontal derivatives computed on constant σ surfaces to those computed on constant z surfaces

$$\left[\frac{\partial}{\partial x} \right]_{\sigma} = \left[\frac{\partial}{\partial x} \right]_z + \left[\frac{\partial z}{\partial x} \right]_{\sigma} \frac{\partial}{\partial z} = \left[\frac{\partial}{\partial x} \right]_z + \left[\frac{\partial \eta}{\partial x} \right]_{\sigma} \frac{\partial}{\partial z} \quad (43.67a)$$

$$\left[\frac{\partial}{\partial y} \right]_{\sigma} = \left[\frac{\partial}{\partial y} \right]_z + \left[\frac{\partial z}{\partial y} \right]_{\sigma} \frac{\partial}{\partial z} = \left[\frac{\partial}{\partial y} \right]_z + \left[\frac{\partial \eta}{\partial y} \right]_{\sigma} \frac{\partial}{\partial z}, \quad (43.67b)$$

which can be written in the shorthand vector notation

$$\nabla_{\sigma} = \hat{x} \left[\frac{\partial}{\partial x} \right]_{\sigma} + \hat{y} \left[\frac{\partial}{\partial y} \right]_{\sigma} = \nabla_z + (\nabla_{\sigma} z) \partial_z = \nabla_z + (\nabla_{\sigma} \eta) \partial_z. \quad (43.68)$$

43.12 Material time derivative

Making use of the relations for the partial derivative operators in Section 43.11 allows us to write the material time derivative in the following equivalent forms

$$\frac{D}{Dt} = \left[\frac{\partial}{\partial t} \right]_z + \mathbf{u} \cdot \nabla_z + w \frac{\partial}{\partial z} \quad (43.69a)$$

$$= \left[\frac{\partial}{\partial t} \right]_{\sigma} - (\partial \eta / \partial \bar{t}) \partial_z + \mathbf{u} \cdot [\nabla_{\sigma} - (\nabla_{\sigma} \eta) \partial_z] + w \partial / \partial z \quad (43.69b)$$

$$= \left[\frac{\partial}{\partial t} \right]_{\sigma} + \mathbf{u} \cdot \nabla_{\sigma} + [w - \mathbf{u} \cdot \nabla_{\sigma} \eta - \partial \eta / \partial \bar{t}] (\partial \sigma / \partial z) \partial / \partial \sigma \quad (43.69c)$$

$$= \left[\frac{\partial}{\partial t} \right]_{\sigma} + \mathbf{u} \cdot \nabla_{\sigma} + \frac{D\sigma}{Dt} \frac{\partial}{\partial \sigma} \quad (43.69d)$$

$$= \left[\frac{\partial}{\partial t} \right]_{\sigma} + \mathbf{u} \cdot \nabla_{\sigma} + \frac{\partial z}{\partial \sigma} \frac{D\sigma}{Dt} \frac{\partial}{\partial z}. \quad (43.69e)$$

The equality (43.69d) made use of the identity (43.48), which is itself derived in Section 44.5 where we discuss further kinematic results using GVCs. Besides differences in the spatial operators, it is important to note that the time derivative operators are computed on constant geopotential and constant GVC surfaces, respectively. However, the horizontal velocity component is the *same* for both forms of the material time derivative

$$(u, v) = \frac{D(x, y)}{Dt}. \quad (43.70)$$

43.13 Divergence of a vector and the divergence theorem

Making use of the general expression (7.15) for the covariant divergence of a vector renders the GVC expression

$$\nabla_{\bar{a}} F^{\bar{a}} = [\det(g_{\bar{a}\bar{b}})]^{-1/2} \partial_{\bar{a}} \left[[\det(g_{\bar{a}\bar{b}})]^{1/2} F^{\bar{a}} \right] = (\partial z / \partial \sigma)^{-1} \partial_{\bar{a}} [(\partial z / \partial \sigma) F^{\bar{a}}]. \quad (43.71)$$

Recall that the GVC vector components, $F^{\bar{a}}$, are related to the Cartesian components in equation (43.34), and the GVC components of the partial derivative operator, $\partial_{\bar{a}}$, are related to the Cartesian operator in equation (43.63).

When making use of the divergence theorem (Section 7.18), we require the product of the volume element and the covariant divergence. For GVCs this product takes on the form

$$(\nabla_{\bar{a}} F^{\bar{a}}) dV = \partial_{\bar{a}} [(\partial z / \partial \sigma) F^{\bar{a}}] d\bar{x} d\bar{y} d\sigma, \quad (43.72)$$

which reduces to a boundary integral when integrating over a volume.

43.14 The diffusion operator

As an explicit example of the covariant divergence operator (43.71), we here consider the diffusion operator discussed in Chapter 48. The derivation here recovers much of what we just discussed in Section 43.13, yet we make use of a bit less tensor formalism though at the cost of more algebra.

43.14.1 Continuous expression

The diffusion operator is the convergence of the diffusive flux

$$\mathcal{R} = -\nabla \cdot \mathbf{J}, \quad (43.73)$$

where \mathbf{J} is a vector field. Let us convert the pieces of this operator from Cartesian coordinates into GVC coordinates, making use of the transformation of partial derivative operators given in Section 43.11. Also, we make use of the shorthand $z(x, y, \sigma, t)$ rather than $\eta(x, y, \sigma, t)$

$$-\mathcal{R} = \nabla \cdot \mathbf{J} \quad (43.74a)$$

$$= \nabla_z \cdot \mathbf{J}^h + \partial_z J^z \quad (43.74b)$$

$$= (\nabla_{\sigma} - \nabla_{\sigma} z \partial_z) \cdot \mathbf{J}^h + (\sigma_z) \partial_{\sigma} J^z \quad (43.74c)$$

$$= \sigma_z \left[z_{\sigma} \nabla_{\sigma} \cdot \mathbf{J}^h + (\hat{z} \partial_{\sigma} - \nabla_{\sigma} z \partial_{\sigma}) \cdot \mathbf{J} \right] \quad (43.74d)$$

$$= \sigma_z \left[\nabla_{\sigma} \cdot (z_{\sigma} \mathbf{J}^h) - \mathbf{J}^h \cdot \nabla_{\sigma} (z_{\sigma}) + \partial_{\sigma} J^z - \partial_{\sigma} (\nabla_{\sigma} z \cdot \mathbf{J}) + \mathbf{J} \cdot \partial_{\sigma} (\nabla_{\sigma} z) \right] \quad (43.74e)$$

$$= \sigma_z \left[\nabla_{\sigma} \cdot (z_{\sigma} \mathbf{J}^h) + \partial_{\sigma} J^z - \partial_{\sigma} (\nabla_{\sigma} z \cdot \mathbf{J}^h) \right] \quad (43.74f)$$

$$= \sigma_z \left(\nabla_{\sigma} \cdot (\partial_{\sigma} z \mathbf{J}^h) + \partial_{\sigma} [(\hat{z} - \nabla_{\sigma} z) \cdot \mathbf{J}] \right) \quad (43.74g)$$

$$= \sigma_z \left[\nabla_{\sigma} \cdot (z_{\sigma} \mathbf{J}^h) + \partial_{\sigma} (z_{\sigma} \nabla_{\sigma} \cdot \mathbf{J}) \right], \quad (43.74h)$$

where we used

$$z_{\sigma} \nabla_{\sigma} = \hat{z} - \nabla_{\sigma} z \quad (43.75)$$

to reach the final equality, and made use of the shorthand

$$z_{\sigma} = \partial z / \partial \sigma \quad \text{and} \quad \sigma_z = \partial \sigma / \partial z = (z_{\sigma})^{-1}. \quad (43.76)$$

The coordinate transformations in Section 43.6 for vector components reveal that the expression (43.74h) is identical to equation (43.71) derived using formal tensor methods. Likewise, multiplying by the volume element

$$dV = dx dy dz = dx dy z_{\sigma} d\sigma, \quad (43.77)$$

leads to

$$-\mathcal{R} dV = \left[\nabla_{\sigma} \cdot (z_{\sigma} \mathbf{J}^h) + \partial_{\sigma} (z_{\sigma} \nabla_{\sigma} \cdot \mathbf{J}) \right] dx dy d\sigma, \quad (43.78)$$

which is identical to the expression (43.72).

43.14.2 Layer thickness weighted diffusion operator

Consider a prescribed increment, $\delta\sigma$, separating two σ isosurfaces. This increment commutes with the horizontal operator ∇_σ , acting within the layer. We can thus formally consider the following layer-integrated or thickness weighted form of the diffusion operator

$$-\mathcal{R} \delta V = \left[\nabla_\sigma \cdot (\delta\sigma z_\sigma \mathbf{J}^h) + \delta\sigma \partial_\sigma (z_\sigma \nabla\sigma \cdot \mathbf{J}) \right] \delta x \delta y \quad (43.79a)$$

$$= \frac{1}{\delta z} \left[\nabla_\sigma \cdot (\delta\sigma z_\sigma \mathbf{J}^h) + \delta\sigma \partial_\sigma (z_\sigma \nabla\sigma \cdot \mathbf{J}) \right] \delta x \delta y \delta z \quad (43.79b)$$

$$= \frac{1}{h} \left[\nabla_\sigma \cdot (h \mathbf{J}^h) + \Delta_\sigma (z_\sigma \nabla\sigma \cdot \mathbf{J}) \right] \delta x \delta y h, \quad (43.79c)$$

where we introduced the infinitesimal layer thickness

$$h = z_\sigma \delta\sigma \quad (43.80)$$

and the non-dimensional differential operator

$$\Delta_\sigma \equiv \delta\sigma \frac{\partial}{\partial\sigma}. \quad (43.81)$$

Cancelling the volume element on both sides leads to the diffusion operator

$$\mathcal{R} = -\frac{1}{h} \left[\nabla_\sigma \cdot (h \mathbf{J}^h) + \Delta_\sigma (z_\sigma \nabla\sigma \cdot \mathbf{J}) \right]. \quad (43.82)$$

This form is commonly found in the numerical modeling literature when considering generalized vertical coordinate models.

We make the following comments concerning the diffusion operator in equation (43.82).

- Our introduction of the layer thickness $h = z_\sigma \delta\sigma$ is treated a bit more formally in Sections 44.10 and 44.11 by considering a vertical integral over a coordinate layer. Even so, the resulting diffusion operator is the same as that derived here.
- The thickness weighted flux, $h \mathbf{J}^h$, is oriented within the horizontal plane. However, its contribution to the diffusion operator is computed by taking its divergence using the operator ∇_σ rather than the horizontal operator ∇_z . This distinction is fundamental to how operators, such as advection and diffusion, appear using generalized vertical coordinates.
- The flux $z_\sigma \nabla\sigma \cdot \mathbf{J}$ is commonly referred to as the dia-surface subgrid scale flux.
- For the special case of a diffusive flux with zero component parallel to $\nabla\sigma$, the diffusion operator reduces to

$$\mathcal{R} = -\frac{1}{h} \left[\nabla_\sigma \cdot (h \mathbf{J}^h) \right] \quad \text{if } \nabla\sigma \cdot \mathbf{J} = 0. \quad (43.83)$$

The neutral diffusion operator of Section 50.3.3 is an example of such an operator, with σ in that case given by the locally referenced potential density.

43.15 Vorticity

As detailed in Chapter 36, vorticity is the curl of the velocity

$$\vec{\omega} = \text{curl}(\vec{v}), \quad (43.84)$$

where the curl has components (Section 7.17)

$$\text{curl}(\vec{v}) = \tilde{e}_a \varepsilon^{abc} \partial_b v_c = \tilde{e}_{\bar{a}} \varepsilon^{\bar{a}\bar{b}\bar{c}} \partial_{\bar{b}} v_{\bar{c}}. \quad (43.85)$$

43.15.1 The components

We identify the contravariant components of the vorticity via

$$\omega^{\bar{a}} = \varepsilon^{\bar{a}\bar{b}\bar{c}} \partial_{\bar{b}} v_{\bar{c}} = (\partial z / \partial \sigma)^{-1} \epsilon^{\bar{a}\bar{b}\bar{c}} \partial_{\bar{b}} v_{\bar{c}} \quad (43.86)$$

where we made use of equation (43.58) to introduce the permutation symbol. Expanding the components leads to

$$\omega^{\bar{1}} = (\partial \sigma / \partial z) (\partial_{\bar{2}} v_{\bar{3}} - \partial_{\bar{3}} v_{\bar{2}}) \quad (43.87a)$$

$$\omega^{\bar{2}} = (\partial \sigma / \partial z) (\partial_{\bar{3}} v_{\bar{1}} - \partial_{\bar{1}} v_{\bar{3}}) \quad (43.87b)$$

$$\omega^{\bar{3}} = \omega^\sigma = (\partial \sigma / \partial z) (\partial_{\bar{1}} v_{\bar{2}} - \partial_{\bar{2}} v_{\bar{1}}). \quad (43.87c)$$

43.15.2 Transforming from Cartesian coordinates

The above approach works solely with the GVC coordinates. An alternative approach connects the GVC vorticity components and the Cartesian vorticity components. For that purpose we use the transformation matrix via

$$\omega^{\bar{a}} = \Lambda_a^{\bar{a}} \omega^a, \quad (43.88)$$

where ω^a are the Cartesian components

$$\boldsymbol{\omega} = \hat{x} \left(\frac{\partial w}{\partial y} - \frac{\partial v}{\partial z} \right) + \hat{y} \left(\frac{\partial u}{\partial z} - \frac{\partial w}{\partial x} \right) + \hat{z} \left(\frac{\partial v}{\partial x} - \frac{\partial u}{\partial y} \right). \quad (43.89)$$

Making use of the transformation matrix $\Lambda_a^{\bar{a}}$ from equation (43.14) yields (as in Section 43.6)

$$\omega^{\bar{x}} = \omega^x = \frac{\partial w}{\partial y} - \frac{\partial v}{\partial z} \quad \text{and} \quad \omega^{\bar{y}} = \omega^y = \frac{\partial u}{\partial z} - \frac{\partial w}{\partial x} \quad \text{and} \quad \omega^\sigma = \boldsymbol{\omega} \cdot \nabla \sigma. \quad (43.90)$$

Note that for isopycnal coordinates in a Boussinesq fluid, ω^σ equals to the potential vorticity when the vorticity is the absolute vorticity (Section 46.2). That is, the potential vorticity is the isopycnal component of the absolute vorticity.

43.16 Velocity circulation

The velocity circulation (Section 33.3) is given by the closed oriented path integral of the velocity projected into the direction of the path

$$\mathcal{C} \equiv \oint_{\partial\mathcal{S}} \mathbf{v} \cdot d\mathbf{x} \quad (43.91)$$

where $d\mathbf{x}$ is the vector line element along the path and $\partial\mathcal{S}$ is the closed path defining the boundary to a two-dimensional surface \mathcal{S} . Stokes' Theorem from Section 2.6 leads to the identity

$$\mathcal{C} = \oint_{\partial\mathcal{S}} \mathbf{v} \cdot d\mathbf{x} = \int_{\mathcal{S}} (\nabla \wedge \mathbf{v}) \cdot \hat{\mathbf{n}} d\mathcal{S} = \int_{\mathcal{S}} \boldsymbol{\omega} \cdot \hat{\mathbf{n}} d\mathcal{S}, \quad (43.92)$$

where $\hat{\mathbf{n}}$ is the outward normal vector orienting the area element $d\mathcal{S}$ according to the right-hand rule applied to the bounding circuit. These results are all written in a generally covariant manner (Section 6.1) so that they hold for an arbitrary coordinate representation.

As a particular case, consider the circulation around a closed path on a constant σ surface, in which

$$\hat{\mathbf{n}} = \frac{\nabla\sigma}{|\nabla\sigma|} \quad (43.93)$$

is the outward normal and

$$\boldsymbol{\omega} \cdot \hat{\mathbf{n}} = \frac{\omega^\sigma}{|\nabla\sigma|} \quad (43.94)$$

where $\omega^\sigma = \boldsymbol{\omega} \cdot \nabla\sigma$ (equation (43.90)). So long as the vertical stratification remains non-zero ($\partial\sigma/\partial z \neq 0$) we can write the area factor in the form

$$\frac{d\mathcal{S}}{|\nabla\sigma|} = \frac{d\mathcal{S}}{\sqrt{(\partial\sigma/\partial x)^2 + (\partial\sigma/\partial y)^2 + (\partial\sigma/\partial z)^2}} \quad (43.95a)$$

$$= \frac{d\mathcal{S}}{|\partial\sigma/\partial z| \sqrt{[(\partial\sigma/\partial x)/(\partial\sigma/\partial z)]^2 + [(\partial\sigma/\partial y)/(\partial\sigma/\partial z)]^2 + 1}} \quad (43.95b)$$

$$= \frac{d\mathcal{S}}{|\partial\sigma/\partial z| \sqrt{1 + \tan^2 \vartheta}} \quad (43.95c)$$

$$= \left| \frac{\partial z}{\partial \sigma} \right| |\cos \vartheta| d\mathcal{S} \quad (43.95d)$$

$$= \left| \frac{\partial z}{\partial \sigma} \right| dA. \quad (43.95e)$$

The equality (43.95c) introduces the angle, ϑ , between the boundary surface and the horizontal plane as in Figure 43.4. The squared slope of this surface given by

$$\tan^2 \vartheta = \frac{\nabla_z \sigma \cdot \nabla_z \sigma}{(\partial\sigma/\partial z)^2} = \nabla_\sigma z \cdot \nabla_\sigma z. \quad (43.96)$$

The equality (43.95d) made use of a trigonometric identity, and the equality (43.95e) introduced the horizontal projection of the area,

$$dA = |\cos \vartheta| d\mathcal{S}. \quad (43.97)$$

Bringing these results together leads to the expression for circulation around a closed loop on a constant σ surface

$$\mathcal{C}_{\sigma-\text{surface}} = \int_{\mathcal{S}} (\boldsymbol{\omega} \cdot \nabla\sigma) |\partial z/\partial \sigma| dA. \quad (43.98)$$



Kinematics

In providing a mechanistic description of budgets within the ocean or atmosphere, it is often useful to measure the material or momentum transfer through a surface. This transport is termed the *dia-surface transport*. Our discussion in this chapter unifies ideas developed for kinematic boundary conditions in Section 15.4 with transport across an arbitrary surface in the fluid interior. We do so by making use of the generalized vertical coordinates (GVCs) first introduced in Chapter 43. We make use of the dia-surface transport formulation to express the material time derivative operator using GVCs. This form for the material time derivative allows us to decompose the vertical velocity into motion relative to a moving GVC surface. In turn, we are afforded a means to reinterpret the velocity vector and corresponding particle trajectories. GVC kinematics also provides a means to express the subduction of fluid into the ocean interior beneath the mixed layer base. We close the chapter with derivations of the GVC version of mass continuity and the tracer equation. We also introduce the layer integrated version of the continuity and tracer equations, with the layer integrated equations appropriate for discrete numerical fluid models.

READER'S GUIDE TO THIS CHAPTER

We introduced mathematical properties of generalized vertical coordinates (GVCs) in Chapter 43, including the calculus using these non-orthogonal coordinates. It is essential to have a working knowledge of that material to understand the present chapter. Later in Chapter 45 we detail the dynamical equations using GVCs, with material in that chapter relying on the kinematics presented here. Following the treatment in Chapter 43, we here use the symbol σ to denote a generalized vertical coordinate.

44.1	Loose threads	1004
44.2	Example generalized vertical coordinates	1004
44.2.1	Ocean free surface	1005
44.2.2	Solid earth boundary	1005
44.2.3	Ocean mixed layer base	1005
44.2.4	Interior generalized vertical coordinate surfaces	1005
44.3	Specific thickness	1006
44.4	The dia-surface transport	1007
44.4.1	Flow normal to the GVC surface	1007
44.4.2	Accounting for movement of the surface	1008
44.4.3	We only care about divergent surface motion	1009
44.4.4	Cross GVC transport in terms of GVC material evolution	1009
44.4.5	Defining the dia-surface transport	1009
44.4.6	Expressions for the dia-surface velocity component	1010

44.4.7 An alternative definition of dia-surface velocity component	1011
44.4.8 Area integrated dia-surface transport for incompressible flows	1011
44.5 Material time derivative	1013
44.6 Vertical velocity and dia-surface velocity	1014
44.6.1 Decomposing the vertical velocity	1014
44.6.2 Another form of the vertical velocity decomposition	1014
44.7 The velocity vector and fluid particle trajectories	1016
44.8 Subduction across the mixed layer base	1017
44.9 Mass continuity	1018
44.9.1 Cartesian coordinates	1018
44.9.2 Generalized vertical coordinates	1019
44.10 Layer integrated mass continuity	1020
44.10.1 Compressible fluids	1021
44.10.2 Mass continuity using pressure coordinates	1022
44.10.3 Incompressible flow	1023
44.10.4 Rescaled geopotential coordinates	1023
44.11 Layer integrated tracer equation	1024

44.1 Loose threads

- Develop the notion of overturning streamfunctions as per Sections 32.2 to 32.4 of the MOM5 elements guide. Work through the issues for both Boussinesq and non-Boussinesq.
- Develop the notion of depth integrated transport as per Sections 32.2 to 32.4 of the MOM5 elements guide. Work through the issues for both Boussinesq and non-Boussinesq.

44.2 Example generalized vertical coordinates

We here consider some generalized vertical coordinates that will prove of use for our discussion in this chapter, with Figure 44.1 illustrating the examples.

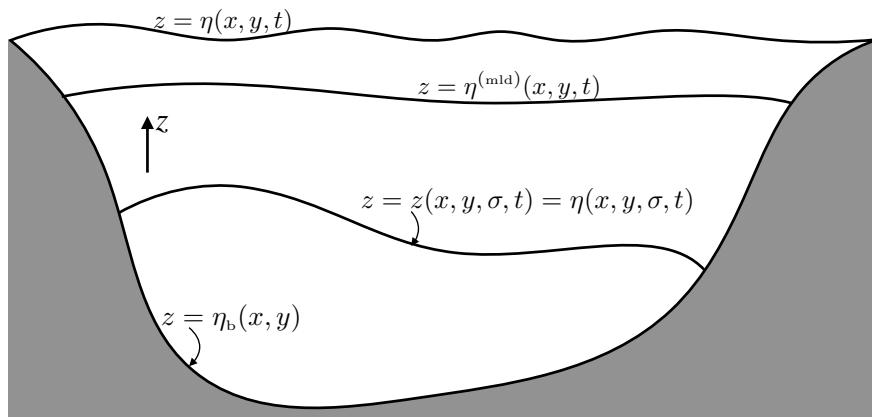


FIGURE 44.1: Example generalized vertical coordinates, $\sigma(x, y, z, t)$. The ocean free surface can be represented mathematically by $\sigma(x, y, z, t) = z - \eta(x, y, t) = 0$; the ocean mixed layer base by $\sigma(x, y, z, t) = z - \eta^{(\text{mld})}(x, y, t) = 0$; the position of an interior GVC coordinate σ with depth $z = z(x, y, \sigma, t) = \eta(x, y, \sigma)$; and the solid earth bottom $\sigma(x, y, z) = z - \eta_b(x, y) = 0$.

44.2.1 Ocean free surface

The first surface is the ocean free surface, whose kinematic boundary conditions were derived in Section 15.4.3. Here, water and tracer penetrate this surface through precipitation, evaporation, river runoff (when applied as an upper ocean boundary condition), and sea ice melt. Momentum exchange arises from stresses between the ocean and atmosphere or ice. The ocean free surface can be represented mathematically by the identity

$$\sigma(x, y, z, t) = z - \eta(x, y, t) = 0 \quad \text{ocean free surface.} \quad (44.1)$$

This identity holds so long as we assume the surface height η is smooth and contains no overturns at the scales of interest. That is, we assume breaking surface waves are filtered from the description.

44.2.2 Solid earth boundary

We may describe the solid Earth lower boundary mathematically by using the time independent expression

$$\sigma(x, y, z) = z + H(x, y) = z - \eta_b(x, y) = 0 \quad \text{ocean bottom,} \quad (44.2)$$

where we introduce the two common symbols used for the bottom, $\eta_b = -H$. We generally prefer η_b since H is used elsewhere in this book for vertical scale. As detailed in Section 15.4.1, we typically assume that there is no fluid mass transport through the solid Earth. However, in the case of geothermal heating, we may consider an exchange of heat between the ocean and the solid Earth. Momentum exchange through the action of stresses occur between the solid Earth and ocean fluid.

44.2.3 Ocean mixed layer base

Let

$$\sigma = z - \eta^{(\text{mld})}(x, y, t) = 0 \quad (44.3)$$

represent the vertical position of the ocean mixed layer base. The corresponding normal vector is given by

$$\hat{\mathbf{n}}^{(\text{mld})} = \frac{\nabla(z - \eta^{(\text{mld})})}{|\nabla(z - \eta^{(\text{mld})})|} = \frac{\nabla(\hat{z} - \nabla\eta^{(\text{mld})})}{\sqrt{1 + |\nabla\eta^{(\text{mld})}|^2}}. \quad (44.4)$$

This example is relevant for the study of ocean ventilation, whereby we are interested in measuring the transport of fluid that enters the ocean interior across the mixed layer base (see Section 44.8).

44.2.4 Interior generalized vertical coordinate surfaces

Within the ocean interior, transport across surfaces of constant generalized vertical coordinate $\sigma = \sigma(x, y, z, t)$ constitutes the dia-surface transport affecting budgets of mass, tracer, and momentum within layers bounded by two generalized vertical coordinate surfaces. A canonical example is provided by isopycnal layers formed by surfaces of constant potential density (or equivalently constant buoyancy surfaces) as used in isopycnal ocean models as well as theoretical descriptions of adiabatic ocean dynamics. The vertical position of this surface is written in one of two equivalent manners

$$z = z(x, y, \sigma, t) = \eta(x, y, \sigma, t). \quad (44.5)$$

The first expression exposes the functional dependence of the vertical position of the σ surface at a horizontal position and time. In Section 43.1.1 we discussed the potential for confusion between writing z as a particular vertical position versus a function, thus motivating $z = \eta(x, y, \sigma, t)$.

However, by now we should have sufficient experience with generalized vertical coordinates so that we can well distinguish when z refers to a particular vertical position versus $z(x, y, \sigma, t)$ as a coordinate function. For this reason we only infrequently use $z = \eta(x, y, \sigma, t)$ in this chapter.

44.3 Specific thickness

As mentioned in Section 43.8.1, a surface of constant generalized vertical coordinate can be successfully used to partition the vertical so long as the transformation between the generalized vertical coordinate and the geopotential coordinate is invertible. The Jacobian of transformation is given by

$$\frac{\partial z}{\partial \sigma} = z_\sigma, \quad (44.6)$$

which must be single signed for suitable generalized vertical coordinates. This constraint means that we do not allow the surfaces to overturn, which is the same assumption made about the ocean surface, $z = \eta(x, y, t)$, and solid earth bottom, $z = \eta_b(x, y)$. This restriction places a limitation on the ability of certain GVC models (e.g., isopycnal models) to describe non-hydrostatic processes, such as the overturning common in Kelvin-Helmholtz billows and gravitational convection. Note that for both the solid earth bottom and ocean free surface

$$\frac{\partial z}{\partial \sigma} = 1 \quad \text{ocean free surface and fluid/solid interface.} \quad (44.7)$$

Furthermore, this relation also holds, trivially, for geopotential coordinates in which $\sigma = z$.

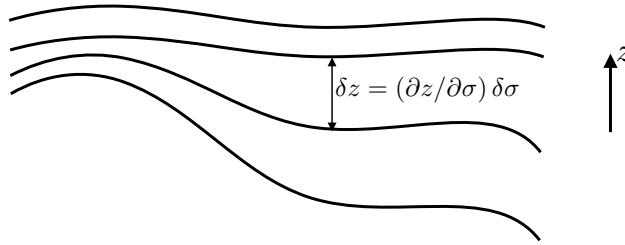


FIGURE 44.2: Illustrating the thickness between surfaces of constant generalized vertical coordinates, $\delta z = (\partial z / \partial \sigma) \delta \sigma$. In regions with larger magnitude for the specific thickness, $\partial z / \partial \sigma$, or equivalently smaller vertical stratification of the σ surfaces, $\partial \sigma / \partial z$, the layer thicknesses are further apart. The converse holds where $\partial z / \partial \sigma$ is small (equivalently $\partial \sigma / \partial z$ is large).

We refer to the Jacobian z_σ as the *specific thickness* and sometimes find it useful to write it as

$$h = z_\sigma = \frac{\partial z}{\partial \sigma}. \quad (44.8)$$

This name is motivated by noting that the vertical thickness of an infinitesimal layer of coordinate thickness $\delta \sigma$ is given by

$$\delta z = \frac{\partial z}{\partial \sigma} \delta \sigma = h \delta \sigma, \quad (44.9)$$

with Figure 44.2 providing an example with finitely thick layers. For example, if $\sigma = b(x, y, z, t)$ (buoyancy or potential density as in isopycnal models), then the thickness of a buoyancy layer is given by

$$\delta z = \frac{\partial z}{\partial \sigma} \delta b = N^{-2} \delta b, \quad (44.10)$$

with

$$N^2 = \frac{\partial b}{\partial z} \quad (44.11)$$

the squared buoyancy frequency (Section 25.5.4) in a Boussinesq fluid (Chapter 24). For a hydrostatic fluid using pressure as the vertical coordinate, the thickness of a pressure layer is

$$\delta z = \frac{\partial z}{\partial p} \delta p = -\frac{1}{\rho g} \delta p \quad (44.12)$$

where we used the hydrostatic relation (Section 23.3)

$$\frac{\partial p}{\partial z} = -\rho g \quad (44.13)$$

with g the constant acceleration due to effective gravity. Note that we assume the layer thickness is positive, $\delta z > 0$. For this purpose, with hydrostatic pressure we might choose to consider negative pressure increments, $\delta p < 0$, as this corresponds to vertically upward movement in a fluid column.

44.4 The dia-surface transport

In this section we develop the concept of dia-surface transport and derive its expression in terms of the material time derivative of the GVC surface.

44.4.1 Flow normal to the GVC surface

At an arbitrary point on a surface of constant generalized vertical coordinate (see Figure 44.3), the rate at which fluid moves in the direction normal to the surface is given by

$$\text{RATE OF FLUID FLOW IN DIRECTION } \hat{\mathbf{n}} = \mathbf{v} \cdot \hat{\mathbf{n}}, \quad (44.14)$$

where

$$\hat{\mathbf{n}} = \frac{\nabla \sigma}{|\nabla \sigma|}, \quad (44.15)$$

is the surface unit normal. Two examples are useful to ground this expression in common experience. For the ocean free surface, $\sigma = z - \eta(x, y, t) = 0$, the unit normal takes the form

$$\hat{\mathbf{n}} = \frac{\nabla(z - \eta)}{|\nabla(z - \eta)|} = \frac{\hat{\mathbf{z}} - \nabla \eta}{\sqrt{1 + |\nabla \eta|^2}}, \quad (44.16)$$

whereas at the solid Earth bottom, $\sigma = z - \eta_b(x, y) = 0$,

$$\hat{\mathbf{n}} = -\frac{\nabla(z - \eta_b)}{|\nabla(z - \eta_b)|} = -\frac{\hat{\mathbf{z}} - \nabla \eta_b}{\sqrt{1 + |\nabla \eta_b|^2}}. \quad (44.17)$$

Introducing the material time derivative

$$\frac{D\sigma}{Dt} = \frac{\partial \sigma}{\partial t} + \mathbf{v} \cdot \nabla \sigma \quad (44.18)$$

to equation (44.14) leads to the identity

$$\mathbf{v} \cdot \hat{\mathbf{n}} = \frac{1}{|\nabla \sigma|} \left[\frac{D\sigma}{Dt} - \frac{\partial \sigma}{\partial t} \right]. \quad (44.19)$$

Hence, the component to the velocity of a fluid particle that is normal to a GVC surface is proportional to the difference between the material time derivative of the surface coordinate and its partial time derivative.

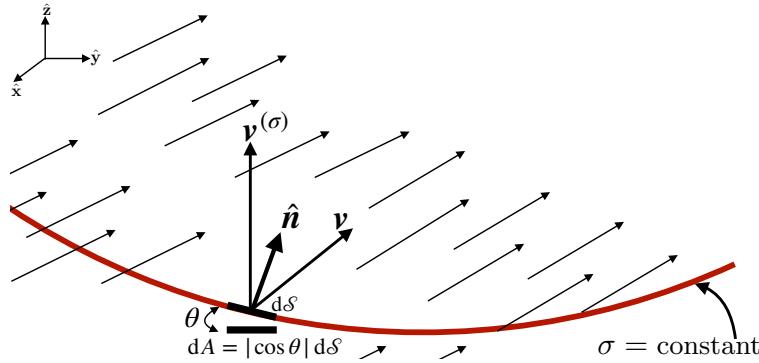


FIGURE 44.3: A surface of constant generalized vertical coordinate, $\sigma = \text{constant}$, within a fluid. The normal direction, $\hat{\mathbf{n}} = \nabla\sigma/|\nabla\sigma|$, points in the direction of increasing σ . We show an example velocity vector for a fluid particle, \mathbf{v} , at a point on the surface as well as the velocity, $\mathbf{v}^{(\sigma)}$, of a point that lives on the surface. Note that kinematics is only concerned with the normal component to the surface velocity, $\mathbf{v}^{(\sigma)} \cdot \hat{\mathbf{n}}$, as per equation (44.25). We require dynamical information to obtain information about the tangential component of $\mathbf{v}^{(\sigma)}$, but such information is not required for this chapter. Following equation (44.30), the horizontal projection of the surface area element is given by $dA = |\cos\vartheta| d\delta$, where ϑ is the angle between the surface and the horizontal and $dA = dx dy$.

44.4.2 Accounting for movement of the surface

A generalized vertical coordinate surface is generally moving. So to diagnose the net transport of fluid penetrating the surface requires us to subtract the velocity of the surface, $\mathbf{v}^{(\sigma)}$, from the velocity of a fluid particle. We are thus led to

$$\text{RATE THAT FLUID CROSSES A MOVING GVC SURFACE} = \hat{\mathbf{n}} \cdot (\mathbf{v} - \mathbf{v}^{(\sigma)}). \quad (44.20)$$

We next develop a kinematic property of the surface velocity, or more precisely the normal component to that velocity. For that purpose, consider an infinitesimal increment in both space and time under which σ undergoes an infinitesimal change

$$\delta\sigma = \delta\mathbf{x} \cdot \nabla\sigma + \delta t \partial_t\sigma. \quad (44.21)$$

Now restrict attention to a point fixed on a constant σ surface, in which

$$\delta\sigma = \delta\mathbf{x}^{(\sigma)} \cdot \nabla\sigma + \delta t \partial_t\sigma = 0, \quad (44.22)$$

where $\delta\mathbf{x}^{(\sigma)}$ is a differential increment following the moving surface. We define the velocity of that point as

$$\mathbf{v}^{(\sigma)} = \frac{\delta\mathbf{x}^{(\sigma)}}{\delta t}, \quad (44.23)$$

in which case equation (44.22) implies that at each point within the fluid,

$$\frac{\partial\sigma}{\partial t} + \mathbf{v}^{(\sigma)} \cdot \nabla\sigma = 0. \quad (44.24)$$

We can likewise write this equation as one for the normal component of the surface velocity

$$\mathbf{v}^{(\sigma)} \cdot \hat{\mathbf{n}} = -\frac{1}{|\nabla\sigma|} \frac{\partial\sigma}{\partial t}. \quad (44.25)$$

Hence, we reach the sensible result that the normal component to the velocity of the surface vanishes when the surface is static.

44.4.3 We only care about divergent surface motion

For the kinematics of fluid motion relative to a surface of constant generalized vertical coordinates, we are only concerned with the normal component to the surface velocity, $\mathbf{v}^{(\sigma)} \cdot \hat{\mathbf{n}}$. That is, we are only concerned with divergent motion of the surface, defined as motion parallel to the surface normal direction, $\hat{\mathbf{n}}$. We have no concern for rotational or tangential motion, which is motion perpendicular to $\hat{\mathbf{n}}$. Even so, some authors, by fiat, choose to set to zero the tangential component of the surface motion. In fact, specification of the tangential surface velocity component is generally not available without extra information about the surface motion, nor is its specification necessary for developing kinematic properties of fluid motion relative to arbitrary generalized vertical coordinate surfaces. Hence, we make no statement about tangential motion of the surface.

44.4.4 Cross GVC transport in terms of GVC material evolution

Using expression (44.25) in equation (44.20) leads to the net flux of fluid crossing the GVC surface

$$\hat{\mathbf{n}} \cdot (\mathbf{v} - \mathbf{v}^{(\sigma)}) = \frac{1}{|\nabla\sigma|} \frac{D\sigma}{Dt}. \quad (44.26)$$

The material time derivative of the GVC surface thus vanishes if no fluid crosses the surface. Notably, this result holds for motion of the fluid as defined by the barycentric velocity, \mathbf{v} , of Section 16.1.2. For multi-component fluids, $\dot{\sigma} = 0$ can still, in principle, be associated with trace matter exchange across the surface via diffusion so long as the net matter crossing the surface is zero. But this situation is rather fine tuned and thus unlikely. Additionally, matter diffusion also occurs with heat diffusion, in which case $\dot{\sigma} = 0$ only occurs in the absence of both matter and heat diffusion, which then means that σ is a material surface.

44.4.5 Defining the dia-surface transport

The area normalizing the volume flux in equation (44.26) is the area $d\delta$ of an infinitesimal patch on the surface of constant generalized vertical coordinate with outward unit normal $\hat{\mathbf{n}}$. We now follow the trigonometry discussed in Section 43.16 to introduce the horizontal projection of this area, dA , which is more convenient to work with for many purposes. So long as the vertical stratification remains non-zero ($\partial\sigma/\partial z \neq 0$) we can write the area factor in the form

$$\frac{d\delta}{|\nabla\sigma|} = \frac{d\delta}{\sqrt{(\partial\sigma/\partial x)^2 + (\partial\sigma/\partial y)^2 + (\partial\sigma/\partial z)^2}} \quad (44.27a)$$

$$= \frac{d\delta}{|\partial\sigma/\partial z| \sqrt{[(\partial\sigma/\partial x)/(\partial\sigma/\partial z)]^2 + [(\partial\sigma/\partial y)/(\partial\sigma/\partial z)]^2 + 1}} \quad (44.27b)$$

$$= \frac{d\delta}{|\partial\sigma/\partial z| \sqrt{1 + \tan^2 \vartheta}} \quad (44.27c)$$

$$= \left| \frac{\partial z}{\partial \sigma} \right| |\cos \vartheta| d\delta \quad (44.27d)$$

$$= \left| \frac{\partial z}{\partial \sigma} \right| dA. \quad (44.27e)$$

The equality (44.27c) introduced the angle, ϑ , between the boundary surface and the horizontal plane. The squared slope of this surface given by (see Section 43.11)

$$\tan^2 \vartheta = \frac{\nabla_z \sigma \cdot \nabla_z \sigma}{(\partial\sigma/\partial z)^2} = \nabla_\sigma z \cdot \nabla_\sigma z. \quad (44.28)$$

The equality (44.27d) made use of a trigonometric identity so that

$$|\cos \vartheta|^{-1} = |z_\sigma \nabla \sigma|. \quad (44.29)$$

Furthermore, the equality (44.27e) introduced the horizontal projection of the area,

$$dA = |\cos \vartheta| d\mathcal{S}. \quad (44.30)$$

We now introduce the *dia-surface velocity component* for the GVC coordinate

$$w^{(\dot{\sigma})} = \frac{\partial z}{\partial \sigma} \frac{D\sigma}{Dt} = z_\sigma \dot{\sigma}, \quad (44.31)$$

which measures the volume of fluid passing through the surface, per unit horizontal area, per unit time

$$w^{(\dot{\sigma})} \equiv \hat{\mathbf{n}} \cdot (\mathbf{v} - \mathbf{v}^{(\sigma)}) \frac{d\mathcal{S}}{dA} \quad (44.32)$$

$$= \frac{(\text{VOLUME/TIME}) \text{ FLUID THROUGH SURFACE}}{\text{HORIZONTAL AREA OF SURFACE}}, \quad (44.33)$$

so that

$$w^{(\dot{\sigma})} dA \equiv \hat{\mathbf{n}} \cdot (\mathbf{v} - \mathbf{v}^{(\sigma)}) d\mathcal{S}. \quad (44.34)$$

The velocity component $w^{(\dot{\sigma})}$ is referred to as the dia-surface velocity component since it measures flow rate of fluid through the surface. We can think of $w^{(\dot{\sigma})}$ as the “vertical” velocity which, when multiplied by the horizontal area element, measures the transport of fluid that crosses the surface in the normal direction.

44.4.6 Expressions for the dia-surface velocity component

Making use of various identities derived above, as well as the transformation of partial derivative operators in Section 43.11, allows us to write the dia-surface velocity component in the following equivalent forms

$$w^{(\dot{\sigma})} = \frac{\partial z}{\partial \sigma} \frac{D\sigma}{Dt} \quad (44.35a)$$

$$= \frac{\partial z}{\partial \sigma} |\nabla \sigma| \hat{\mathbf{n}} \cdot (\mathbf{v} - \mathbf{v}^{(\sigma)}) \quad (44.35b)$$

$$= \frac{\partial z}{\partial \sigma} \nabla \sigma \cdot \mathbf{v} - \frac{\partial z}{\partial \sigma} |\nabla \sigma| \hat{\mathbf{n}} \cdot (\mathbf{v} - \mathbf{v}^{(\sigma)}) \quad (44.35c)$$

$$= (\hat{\mathbf{z}} - \nabla_\sigma z) \cdot \mathbf{v} + \frac{\partial z}{\partial \sigma} \frac{\partial \sigma}{\partial t} \quad (44.35d)$$

$$= (\hat{\mathbf{z}} - \nabla_\sigma z) \cdot \mathbf{v} - \frac{\partial z}{\partial t} \quad (44.35e)$$

$$= w - (\partial_t + \mathbf{u} \cdot \nabla_\sigma) z, \quad (44.35f)$$

where $\partial z / \partial t = (\partial z / \partial t)_\sigma$ is the time derivative for the depth of the σ surface. We also made use of the identity (see equations (43.25b) and (43.25c))

$$\nabla_\sigma z = -z_\sigma \nabla_z \sigma \quad (44.36)$$

to express the slope of the σ surface as projected onto the horizontal direction plane, as well as the corresponding identity (43.25a) for the time derivative

$$\left[\frac{\partial z}{\partial t} \right]_{\sigma} = - \frac{[\partial \sigma / \partial t]_z}{[\partial \sigma / \partial z]}. \quad (44.37)$$

The form given by equation (44.35f) directly relates the vertical component to the fluid particle velocity to the dia-surface velocity component

$$w = \frac{Dz}{Dt} \longleftrightarrow w^{(\dot{\sigma})} = \frac{\partial z}{\partial \sigma} \frac{D\sigma}{Dt} = w - (\partial_t + \mathbf{u} \cdot \nabla_{\sigma})z. \quad (44.38)$$

When the GVC surface is static, so that it occupies a constant vertical position $\partial z / \partial t = 0$, then the dia-surface velocity component reduces to

$$w^{(\dot{\sigma})} = w - \mathbf{u} \cdot \nabla_{\sigma} z \quad \text{static surface}, \quad (44.39)$$

whereas if the GVC surface is flat, then the dia-surface velocity component measures the flux of fluid moving vertically relative to the motion of the GVC surface. Finally, if the surface is flat and static, the dia-surface velocity component becomes the vertical velocity component

$$w^{(\dot{\sigma})} = w = \frac{Dz}{Dt} \quad \text{GVC surface static and flat}, \quad (44.40)$$

which is the case for the geopotential vertical coordinate. This relation reveals the kinematic distinction between w and $w^{(\dot{\sigma})}$, with the two differing in the presence of GVC transients and horizontal velocities that project onto a non-horizontal GVC surface. Equation (44.35f) thus offers a useful means to distinguish w from $w^{(\dot{\sigma})}$.

44.4.7 An alternative definition of dia-surface velocity component

In some literature presentations, the dia-surface velocity component is taken to be

$$w^{\text{dia}} = \hat{\mathbf{n}} \cdot (\mathbf{v} - \mathbf{v}^{(\sigma)}) = \frac{1}{|\nabla \sigma|} \frac{D\sigma}{Dt}. \quad (44.41)$$

For example, [Groeskamp et al. \(2019\)](#) prefer this definition for watermass analysis. As seen in Chapter 52, the reason to prefer expression (44.41) for watermass analysis is that we do not wish to assume vertically stable stratification for surfaces of constant σ . Dropping that assumption allows us to consider transformation between arbitrarily oriented elements of seawater, even those that are gravitationally unstable.

44.4.8 Area integrated dia-surface transport for incompressible flows

We close this section by further emphasizing the distinction in time dependent flows between dia-surface transport and flow normal to a surface. For this purpose consider an incompressible flow whereby $\nabla \cdot \mathbf{v} = 0$. Incompressibility means that for any closed surface within the fluid interior, the following identity holds via the divergence theorem

$$0 = \int_{\mathcal{R}} \nabla \cdot \mathbf{v} dV = \oint_{\partial \mathcal{R}} \hat{\mathbf{n}} \cdot \mathbf{v} d\mathcal{S}. \quad (44.42)$$

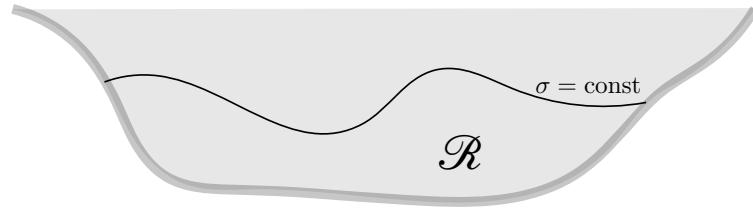


FIGURE 44.4: A constant GVC surface, $\sigma = \text{constant}$, within an ocean basin that intersects the bottom. The region \mathcal{R} is bounded above by the σ surface and below by the solid-earth. Along the constant σ surface an incompressible flow satisfies $\int_{\sigma=\text{const}} \hat{\mathbf{n}} \cdot \mathbf{v} d\mathcal{S} = 0$.

Notably, only in the case of a static surface do we conclude there is no net flow across the surface (see Exercise 17.4). For surfaces that move, there is generally a nonzero net dia-surface transport. We clarify this rather puzzling statement in the following.

As a specific example, consider a fluid region such as shown in Figure 44.4, which is bounded by the solid-earth bottom and a constant GVC surface. Since the solid-earth bottom is static and there is no-normal flow through the bottom, the identity (44.42) means that the area integrated flow normal to the GVC vanishes

$$\int_{\sigma=\text{const}} \hat{\mathbf{n}} \cdot \mathbf{v} d\mathcal{S} = 0. \quad (44.43)$$

But what does this identity imply about the area integrated dia-surface velocity? For the case of a geopotential vertical coordinate, $\sigma = z$, it means that the area integrated vertical velocity vanishes across any geopotential surface below the ocean free surface, $\int_{z=\text{const}} w dA = 0$ (see Exercise 17.4). What about other GVCs?

To address this question consider the general result

$$\int_{\sigma=\text{const}} \hat{\mathbf{n}} \cdot (\mathbf{v} - \mathbf{v}^{(\sigma)}) d\mathcal{S} = \int_{\sigma=\text{const}} w^{\text{dia}} d\mathcal{S} = \int_{\sigma=\text{const}} w^{(\dot{\sigma})} dA, \quad (44.44)$$

where again $dA = dx dy$. Now make use of the property (44.43) for incompressible flows as well as the identity (44.25) to render

$$\int_{\sigma=\text{const}} w^{(\dot{\sigma})} dA = 0 - \int_{\sigma=\text{const}} \hat{\mathbf{n}} \cdot \mathbf{v}^{(\sigma)} d\mathcal{S} \quad (44.45a)$$

$$= \int_{\sigma=\text{const}} \frac{\partial \sigma / \partial t}{|\nabla \sigma|} d\mathcal{S} \quad (44.45b)$$

$$= \int_{\sigma=\text{const}} \frac{\partial \sigma}{\partial t} \left| \frac{\partial z}{\partial \sigma} \right| dA \quad (44.45c)$$

$$= - \int_{\sigma=\text{const}} \left[\frac{\partial z}{\partial t} \right]_\sigma dA. \quad (44.45d)$$

The final equality holds if $\partial z / \partial \sigma > 0$, whereas we swap signs when the vertical stratification is $\partial z / \partial \sigma < 0$. We can go one further step by noting that the time derivative is computed with σ constant, as is the horizontal area integral. Hence, we can pull the time derivative outside the integral to render

$$\int_{\sigma=\text{const}} w^{(\dot{\sigma})} dA = - \left[\frac{\partial}{\partial t} \right]_\sigma \int_{\sigma=\text{const}} z dA. \quad (44.46)$$

This identity means that for an incompressible flow, the integrated dia-surface transport across the GVC surface equals to minus the time tendency for the area integrated vertical position of that

surface. Hence, there is an area integrated dia-surface transport across the GVC surface so long as there is a volume change for the region beneath the surface.

For the case of an isopycnal surface in an adiabatic fluid, there is no change in the volume beneath any interior isopycnal since no flow crosses the isopycnal, in which case we recover the expected result $\int_{\sigma=\text{const}} w^{(\dot{\sigma})} dA = 0$. However, this result does not hold for other coordinates, such as the rescaled vertical coordinate, $\sigma = z^*$ defined by equation (44.112). In this case

$$z^* = H \frac{z - \eta}{H + \eta} \quad (44.47a)$$

$$\frac{\partial z}{\partial z^*} = 1 + H/\eta > 0 \quad (44.47b)$$

$$\left[\frac{\partial z}{\partial t} \right]_{z^*} = \frac{\partial \eta}{\partial t} (1 + z^*/H), \quad (44.47c)$$

so that

$$\int_{z^*=\text{const}} w^{(z^*)} dA = \int_{z^*=\text{const}} (\partial \eta / \partial t) (1 + z^*/H) dA, \quad (44.48)$$

which is generally nonzero. For example, consider a flat bottom so that

$$\int_{z^*=\text{const}} w^{(z^*)} dA = (1 + z^*/H) \int_{z^*=\text{const}} (\partial \eta / \partial t) dA = (1 + z^*/H) \int_{z^*=\text{const}} (Q_m / \rho_0) dA, \quad (44.49)$$

where Q_m is the surface mass flux and we made use of the incompressible free surface equation (17.8). In this case the area integrated dia-surface transport across a z^* surface is proportional to the area integrated surface mass flux.

44.5 Material time derivative

The expression (44.31) for $w^{(\dot{\sigma})}$ brings the material time derivative operator into the following equivalent forms

$$\frac{D}{Dt} = \left[\frac{\partial}{\partial t} \right]_z + \mathbf{u} \cdot \nabla_z + w \frac{\partial}{\partial z} \quad (44.50a)$$

$$= \left[\frac{\partial}{\partial t} \right]_\sigma + \mathbf{u} \cdot \nabla_\sigma + \frac{D\sigma}{Dt} \frac{\partial}{\partial \sigma} \quad (44.50b)$$

$$= \left[\frac{\partial}{\partial t} \right]_\sigma + \mathbf{u} \cdot \nabla_\sigma + w^{(\dot{\sigma})} \frac{\partial}{\partial z}. \quad (44.50c)$$

Note that the chain-rule means that

$$\frac{\partial}{\partial \sigma} = \frac{\partial z}{\partial \sigma} \frac{\partial}{\partial z}, \quad (44.51)$$

thus providing a relationship between the two vertical coordinate partial derivatives. Furthermore, recall that subscripts in the above derivative operators denote variables held fixed when taking the partial derivatives.

We highlight the special case of no fluid particles crossing the generalized coordinate surface. This situation occurs in the case of adiabatic flows with σ equal to the buoyancy or isopycnal coordinate. For adiabatic flow, the material time derivative in equation (44.50c) only has a horizontal two-dimensional advective component, $\mathbf{u} \cdot \nabla_\sigma$. This result should not be interpreted to mean that the fluid particle velocity in an adiabatic flow is strictly horizontal. Indeed, it generally is not, as the form given by equation (44.50a) makes clear. Rather, it means that the advective transport of fluid properties occurs along surfaces of constant buoyancy, and such transport is measured by the convergence of horizontal advective fluxes as measured along these constant buoyancy surfaces.

44.6 Vertical velocity and dia-surface velocity

Making use of the material time derivative operator (44.50c) affords us an opportunity to emphasize both the differences and similarities between the vertical velocity component and the dia-surface velocity component. Namely, the vertical velocity component takes on the equivalent forms

$$w = \frac{Dz}{Dt} = \left[\frac{\partial z}{\partial t} \right]_\sigma + \mathbf{u} \cdot \nabla_\sigma z + w^{(\dot{\sigma})} = \frac{\partial z}{\partial \sigma} \left[-\frac{\partial \sigma}{\partial t} - \mathbf{u} \cdot \nabla_z \sigma + \frac{D\sigma}{Dt} \right], \quad (44.52)$$

and the corresponding expressions for the dia-surface velocity component are given by

$$w^{(\dot{\sigma})} = \frac{\partial z}{\partial \sigma} \frac{D\sigma}{Dt} = \frac{\partial z}{\partial \sigma} \left[\frac{\partial \sigma}{\partial t} + \mathbf{u} \cdot \nabla_z \sigma + w \frac{\partial \sigma}{\partial t} \right] = - \left[\frac{\partial z}{\partial t} \right]_\sigma - \mathbf{u} \cdot \nabla_\sigma z + w. \quad (44.53)$$

Whereas the vertical velocity component, w , measures the transport crossing z surfaces, which are static and horizontal, the dia-surface velocity component, $w^{(\dot{\sigma})}$, measures the transport crossing σ surfaces, which are generally moving and sloped. It is notable that the area normalization used in equation (44.33) for the dia-surface velocity component means that it appears only in the expression for the vertical velocity. However, as we will see in the following, the appearance of $w^{(\dot{\sigma})}$ in the w equation does not necessarily mean that $w^{(\dot{\sigma})}$ corresponds to vertical particle motion. Instead, when it arises from mixing, $w^{(\dot{\sigma})}$ can lead to vertical motion of the σ surface while maintaining a fixed position for the fluid particle.

44.6.1 Decomposing the vertical velocity

The expression

$$w = \left[\frac{\partial z}{\partial t} \right]_\sigma + \mathbf{u} \cdot \nabla_\sigma z + w^{(\dot{\sigma})} \quad (44.54)$$

decomposes the vertical velocity of a fluid particle into (i) changes to the vertical position of the σ -surface at a particular horizontal point, (ii) lateral particle motion projected onto a sloped σ -surface, (iii) motion that crosses a σ -surface. Importantly, the three terms are coupled. For example, consider the case of σ defined by isopycnals, in which case irreversible mixing ($w^{(\dot{\sigma})} \neq 0$) changes the configuration of σ surfaces by changing both their height, $(\partial z / \partial t)_\sigma$, and slope $\nabla_\sigma z$.

44.6.2 Another form of the vertical velocity decomposition

Consider the velocity for a point on the surface, $\mathbf{v}^{(\sigma)}$, which satisfies (Section 44.4.2)

$$\frac{\partial \sigma}{\partial t} + \mathbf{v}^{(\sigma)} \cdot \nabla \sigma = 0. \quad (44.55)$$

Making use of the triple product identities from Section 43.4

$$\frac{\partial z}{\partial \sigma} \nabla \sigma = -\nabla_\sigma z + \hat{\mathbf{z}} \quad \text{and} \quad \frac{\partial z}{\partial \sigma} \left[\frac{\partial \sigma}{\partial t} \right]_z = - \left[\frac{\partial z}{\partial t} \right]_\sigma \quad (44.56)$$

brings equation (44.55) into the form

$$\left[\frac{\partial z}{\partial t} \right]_\sigma = (\hat{\mathbf{z}} - \nabla_\sigma z) \cdot \mathbf{v}^{(\sigma)} \implies \hat{\mathbf{z}} \cdot \mathbf{v}^{(\sigma)} = \left[\frac{\partial z}{\partial t} \right]_\sigma + \mathbf{u}^{(\sigma)} \cdot \nabla_\sigma z, \quad (44.57)$$

where $\mathbf{u}^{(\sigma)}$ is the horizontal component to the surface velocity $\mathbf{v}^{(\sigma)}$. This equation shows that the vertical component to the σ -surface velocity is given by the sum of the changes to the vertical position of the surface plus the projection of the horizontal motion of the surface onto the slope of the surface. Additionally, even if the σ -surface has no component of velocity in the vertical, the depth of the σ -surface measured at a horizontal point generally changes if the surface is sloped and moves horizontally past that point

$$\left[\frac{\partial z}{\partial t} \right]_{\sigma} = -\mathbf{u}^{(\sigma)} \cdot \nabla_{\sigma} z \quad \text{if } \hat{\mathbf{z}} \cdot \mathbf{v}^{(\sigma)} = 0. \quad (44.58)$$

Returning to the general result (44.57) allows us to write

$$\left[\frac{\partial z}{\partial t} \right]_{\sigma} + \mathbf{u} \cdot \nabla_{\sigma} z = \hat{\mathbf{z}} \cdot \mathbf{v}^{(\sigma)} + (\mathbf{u} - \mathbf{u}^{(\sigma)}) \cdot \nabla_{\sigma} z. \quad (44.59)$$

Furthermore, return to the fundamental definition of the dia-surface velocity component detailed in Section 44.4, in which we showed that

$$w^{(\dot{\sigma})} = \frac{\partial z}{\partial \sigma} \frac{D\sigma}{Dt} = \frac{\partial z}{\partial \sigma} \nabla_{\sigma} \cdot (\mathbf{v} - \mathbf{v}^{(\sigma)}) = (-\nabla_{\sigma} z + \hat{\mathbf{z}}) \cdot (\mathbf{v} - \mathbf{v}^{(\sigma)}). \quad (44.60)$$

This expression, along with equation (44.59), leads to the rather elaborate decomposition of the vertical velocity component according to motion of a generalized vertical coordinate surface

$$w = \underbrace{\left[\hat{\mathbf{z}} \cdot \mathbf{v}^{(\sigma)} + (\mathbf{u} - \mathbf{u}^{(\sigma)}) \cdot \nabla_{\sigma} z \right]}_{(\partial_t + \mathbf{u} \cdot \nabla_{\sigma}) z} + \underbrace{\left[\hat{\mathbf{z}} \cdot \mathbf{v} - \hat{\mathbf{z}} \cdot \mathbf{v}^{(\sigma)} - (\mathbf{u} - \mathbf{u}^{(\sigma)}) \cdot \nabla_{\sigma} z \right]}_{w^{(\dot{\sigma})}}. \quad (44.61)$$

Terms in the first bracket compute vertical particle motion relative to the σ -surface. The dia-surface contribution from the second bracket removes the contribution from σ -surface motion to leave just the vertical motion of the particle. All terms on the right hand side cancel, except for $\hat{\mathbf{z}} \cdot \mathbf{v} = w$, thus trivially revealing $w = w$. The decomposition of w is rather pedantic when viewed in the unpacked form of equation (44.61). Even so, let us consider some special cases to offer further interpretation.

- NO HORIZONTAL CONTRIBUTION: Consider the case where the horizontal velocity of a fluid particle matches that of the σ -surface: $\mathbf{u} = \mathbf{u}^{(\sigma)}$. Alternatively, consider the case with flat σ -surfaces so that $\nabla_{\sigma} z = 0$. In either case the vertical velocity is given by

$$w = \underbrace{\left[\hat{\mathbf{z}} \cdot \mathbf{v}^{(\sigma)} \right]}_{(\partial_t + \mathbf{u} \cdot \nabla_{\sigma}) z} + \underbrace{\left[\hat{\mathbf{z}} \cdot (\mathbf{v} - \mathbf{v}^{(\sigma)}) \right]}_{w^{(\dot{\sigma})}}. \quad (44.62)$$

The first contribution is from vertical motion of the σ -surface. The second contribution adjusts for the vertical motion of the particle relative to the σ -surface, leaving behind just the vertical motion of the particle. This rather trivial case exemplifies the contributions from the two pieces of the vertical velocity.

- ZERO VERTICAL PARTICLE MOTION: Consider the case where $w = 0$ so that

$$w = 0 \quad (44.63a)$$

$$= \left[\frac{\partial z}{\partial t} \right]_{\sigma} + \mathbf{u} \cdot \nabla_{\sigma} z + w^{(\dot{\sigma})} \quad (44.63b)$$

$$= \underbrace{\left[\hat{\mathbf{z}} \cdot \mathbf{v}^{(\sigma)} + (\mathbf{u} - \mathbf{u}^{(\sigma)}) \cdot \nabla_{\sigma} z \right]}_{(\partial_t + \mathbf{u} \cdot \nabla_{\sigma}) z} + \underbrace{\left[-\hat{\mathbf{z}} \cdot \mathbf{v}^{(\sigma)} - (\mathbf{u} - \mathbf{u}^{(\sigma)}) \cdot \nabla_{\sigma} z \right]}_{w^{(\dot{\sigma})}}. \quad (44.63c)$$

The final expression is trivial since each term in one bracket identically cancels terms in the other bracket. The penultimate expression reveals the balance between dia-surface transport and motion relative to the σ surface

$$-w^{(\dot{\sigma})} = \left[\frac{\partial z}{\partial t} \right]_{\sigma} + \mathbf{u} \cdot \nabla_{\sigma} z \quad \text{if } w = 0. \quad (44.64)$$

A particularly simple realization of this balance holds for σ given by isopycnals and where the isosurfaces are horizontal. In the presence of uniform mixing, the flat isopycnals stay flat and there is correspondingly no vertical motion of fluid particles even as the vertical stratification is modified. In contrast, the vertical position of an isopycnal surface changes according to the dia-surface velocity component $(\partial z / \partial t)_{\sigma} = -w^{(\dot{\sigma})} \neq 0$. This case illustrates that $w^{(\dot{\sigma})} \neq 0$ can still occur even when there is zero fluid particle motion since $w^{(\dot{\sigma})} \neq 0$ can arise from motion of a σ -surface alone.

44.7 The velocity vector and fluid particle trajectories

Recall from Section 44.6 the alternative forms for the vertical velocity component given by equation (44.52). We focus on the form

$$w = \left[\frac{\partial z}{\partial t} \right]_{\sigma} + \mathbf{u} \cdot \nabla_{\sigma} z + w^{(\dot{\sigma})} \quad (44.65)$$

so that the velocity vector is written¹

$$\mathbf{v} = u \hat{x} + v \hat{y} + w \hat{z} \quad (44.66a)$$

$$= u \hat{x} + v \hat{y} + \left[(\partial z / \partial t)_{\sigma} + \mathbf{u} \cdot \nabla_{\sigma} z + w^{(\dot{\sigma})} \right] \hat{z} \quad (44.66b)$$

$$= u [\hat{x} + \hat{z} (\partial z / \partial x)_{\sigma}] + v [\hat{y} + \hat{z} (\partial z / \partial y)_{\sigma}] + \left[(\partial z / \partial t)_{\sigma} + w^{(\dot{\sigma})} \right] \hat{z}. \quad (44.66c)$$

To help further understand these velocity expressions we consider the following three cases, each of which are illustrated in Figure 44.5.

- STEADY AND MATERIAL σ -SURFACE: The velocity vector is aligned with the instantaneous σ -surface ($\mathbf{v} \cdot \nabla \sigma = 0$) when the σ -surface is steady ($\partial \sigma / \partial t = 0$) and material ($D\sigma / Dt = 0$). Hence, we can diagnose the vertical velocity component in terms of the horizontal via

$$w \partial \sigma / \partial z = -\mathbf{u} \cdot \nabla_z \sigma \implies w = \mathbf{u} \cdot \nabla_{\sigma} z, \quad (44.67)$$

where we used the triple product identities (43.25b) and (43.25c) for the final equality. The velocity vector thus takes on the form

$$\mathbf{v} = u [\hat{x} + \hat{z} (\partial z / \partial x)_{\sigma}] + v [\hat{y} + \hat{z} (\partial z / \partial y)_{\sigma}] \quad \partial \sigma / \partial t = 0 \text{ and } D\sigma / Dt = 0. \quad (44.68)$$

In this case, the velocity vector is determined only by the horizontal velocity plus the slope of the σ surface.

¹ As discussed in Section 43.6, we can connect these expressions to the contravariant representation of the velocity vector using GVCs.

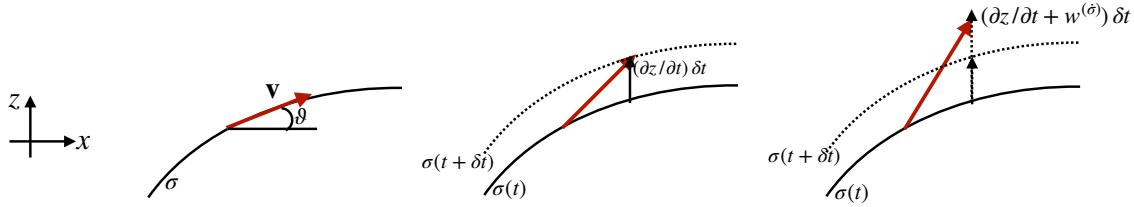


FIGURE 44.5: This schematic shows the various contributions to the fluid particle velocity (red vector) when written relative to motion of a particular generalized vertical coordinate surface. The fluid particle sits at the tail of the velocity vector at time t and at the head at time $t + \delta t$. The left panel is for the case of a static and material σ -surface so that the particle remains on the σ -surface and has a velocity vector given by equation (44.68). The slope of the σ -surface in the \hat{x} -direction is given by $\tan \vartheta = (\partial z / \partial x)_\sigma$. The middle panel is for a non-steady material σ -surface whereby the velocity of a particle takes on the form (44.69), with the particle remaining on the moving σ -surface. The right panel shows the case of a non-steady and non-material σ -surface with velocity (44.70). In this final case the particle position departs from the original σ -surface due to the nonzero dia-surface velocity component, $w^{(\dot{\sigma})} \neq 0$. However, it is not known *a priori* whether this departure is due to particle motion or motion of the surface. Notably, the horizontal position of the particle remains identical for each of the three cases. It is only the vertical position that is modified according to the slope of the σ -surface (left panel), motion of the σ -surface (middle panel), and motion crossing the σ -surface (right panel).

- NON-STEADY AND MATERIAL σ -SURFACE: Next consider material σ surfaces ($D\sigma/Dt = 0$) that move ($\partial_t \sigma \neq 0$), in which case the velocity vector is

$$\mathbf{v} = u [\hat{x} + \hat{z} (\partial z / \partial x)_\sigma] + v [\hat{y} + \hat{z} (\partial z / \partial y)_\sigma] + (\partial z / \partial t)_\sigma \hat{z} \quad D\sigma/Dt = 0. \quad (44.69)$$

To remain on the moving surface, the fluid particle must move vertically by the extra amount $(\partial z / \partial t)_\sigma \delta t \hat{z}$ relative to the case of a static σ -surface.

- NON-STEADY AND NON-MATERIAL σ -SURFACE: The general case with a non-material and non-steady σ also requires the dia-surface velocity component, $w^{(\dot{\sigma})}$, which is diagnosed based on the material time derivative of σ and the inverse stratification, $w^{(\dot{\sigma})} = (\partial z / \partial \sigma) D\sigma/Dt$:

$$\mathbf{v} = u [\hat{x} + \hat{z} (\partial z / \partial x)_\sigma] + v [\hat{y} + \hat{z} (\partial z / \partial y)_\sigma] + [(\partial z / \partial t)_\sigma + w^{(\dot{\sigma})}] \hat{z}. \quad (44.70)$$

The contribution $w^{(\dot{\sigma})}$ measures the vertical motion of the particle relative to the moving σ -surface. Hence, the sum, $(\partial z / \partial t)_\sigma + w^{(\dot{\sigma})}$, measures the vertical motion of the particle relative to a fixed origin. As emphasized in Section 44.6, a non-zero $w^{(\dot{\sigma})}$ arises from motion of the fluid particle relative to the σ -surface, and this relative motion does not necessarily mean that the particle moves; e.g., recall the example discussed in Section 44.6.2 with a static particle and moving σ -surface.

44.8 Subduction across the mixed layer base

Consider the generalized vertical coordinate defined according to the mixed layer base as in equation (44.3). The dia-surface mass transport across this surface leads us to define the subduction

$$-\mathcal{S}^{(\text{subduction})} \equiv \rho dA \left[\frac{D(z - \eta^{(\text{mld})})}{Dt} \right] \quad \text{at } z = \eta^{(\text{mld})}(x, y, t), \quad (44.71)$$

where the mass transport $\mathcal{S}^{(\text{subduction})}$ (dimensions of mass per time) is positive for fluid moving downward beneath the mixed layer base into the pycnocline (subduction) and negative for water

moving into the mixed layer (obduction). The area element dA is the horizontal projection of the area on the mixed layer base. Expanding the material time derivative leads to

$$-\left[\frac{\mathcal{S}^{(\text{subduction})}}{\rho dA}\right] = w - [\partial_t + \mathbf{u} \cdot \nabla] \eta^{(\text{mld})} \quad \text{at } z = \eta^{(\text{mld})}(x, y, t), \quad (44.72)$$

where again we define

$$\mathcal{S}^{(\text{subduction})} > 0 \quad \text{subduction} \quad (44.73)$$

$$\mathcal{S}^{(\text{subduction})} < 0 \quad \text{obduction.} \quad (44.74)$$

We illustrate this definition in Figure 44.6, and note that this definition is consistent with that introduced by *Cushman-Roisin* (1987).

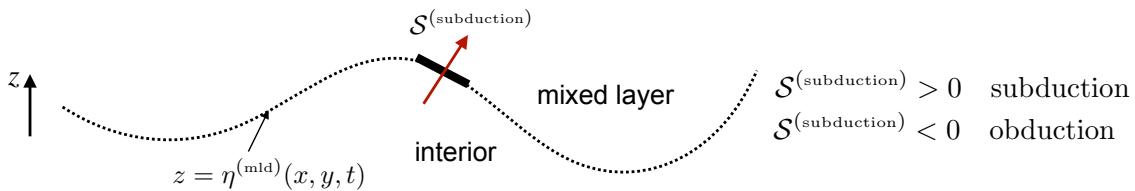


FIGURE 44.6: Illustrating the subduction rate as defined by equation (44.72), which measures the mass transport across the base of the ocean mixed layer. When water enters the ocean interior, $\mathcal{S}^{(\text{subduction})} > 0$, and we say that water subducts from the mixed layer to the ocean interior. Conversely, when water enters the mixed layer from below, $\mathcal{S}^{(\text{subduction})} < 0$ and we say that water is obducted from the interior to the mixed layer.

44.9 Mass continuity

We here derive the Eulerian expression for mass continuity (15.6) using generalized vertical coordinates. We then specialize to incompressible flows, in which mass conservation is converted to volume conservation. To start, recall that mass conservation for a fluid element states that

$$\rho \delta V = \rho \delta x \delta y \delta z = \rho \delta x \delta y z_\sigma \delta \sigma \quad (44.75)$$

is constant following a fluid element.² To develop the Eulerian expressions we first consider the case of Cartesian coordinates.

44.9.1 Cartesian coordinates

Consider the expression

$$\frac{1}{\rho \delta V} \frac{D(\rho \delta V)}{Dt} = \frac{1}{\rho} \frac{D\rho}{Dt} + \frac{1}{\delta V} \frac{D(\delta V)}{Dt}. \quad (44.76)$$

²Recall that we write δ as an infinitesimal increment following a fluid element.

Now make use of Cartesian coordinates to write the volume

$$\frac{1}{\delta V} \frac{D(\delta V)}{Dt} = \frac{1}{\delta x \delta y \delta z} \frac{D(\delta x \delta y \delta z)}{Dt} \quad (44.77a)$$

$$= \frac{1}{\delta x} \frac{D(\delta x)}{Dt} + \frac{1}{\delta y} \frac{D(\delta y)}{Dt} + \frac{1}{\delta z} \frac{D(\delta z)}{Dt} \quad (44.77b)$$

$$= \frac{\delta u}{\delta x} + \frac{\delta v}{\delta y} + \frac{\delta w}{\delta z} \quad (44.77c)$$

$$= \nabla \cdot \mathbf{v}. \quad (44.77d)$$

Setting $D(\rho \delta V)/Dt = 0$ leads to the familiar expression for the continuity equation

$$\frac{D\rho}{Dt} = -\rho \nabla \cdot \mathbf{v}. \quad (44.78)$$

44.9.2 Generalized vertical coordinates

We follow the above procedure but now with generalized vertical coordinates so that

$$\frac{1}{\delta V} \frac{D(\delta V)}{Dt} = \frac{1}{\delta x \delta y z_\sigma \delta \sigma} \frac{D(\delta x \delta y z_\sigma \delta \sigma)}{Dt} \quad (44.79a)$$

$$= \frac{1}{\delta x} \frac{D(\delta x)}{Dt} + \frac{1}{\delta y} \frac{D(\delta y)}{Dt} + \frac{1}{z_\sigma} \frac{D(z_\sigma)}{Dt} + \frac{1}{\delta \sigma} \frac{D(\delta \sigma)}{Dt} \quad (44.79b)$$

$$= \frac{\delta u}{\delta x} + \frac{\delta v}{\delta y} + \frac{1}{z_\sigma} \frac{D(z_\sigma)}{Dt} + \frac{\delta(\dot{\sigma})}{\delta \sigma} \quad (44.79c)$$

$$= \nabla_\sigma \cdot \mathbf{u} + \frac{1}{z_\sigma} \frac{D(z_\sigma)}{Dt} + \frac{\partial \dot{\sigma}}{\partial \sigma} \quad (44.79d)$$

where we introduced the shorthand $\dot{\sigma} = D\sigma/Dt$. Note that we set

$$\frac{\delta u}{\delta x} + \frac{\delta v}{\delta y} = \nabla_\sigma \cdot \mathbf{u} \quad (44.80)$$

since we are working with generalized vertical coordinates so that we consider infinitesimal displacements occurring on constant σ surfaces. We are thus led to

$$\frac{1}{\rho \delta V} \frac{D(\rho \delta V)}{Dt} = \nabla_\sigma \cdot \mathbf{u} + \frac{1}{z_\sigma} \frac{Dz_\sigma}{Dt} + \frac{\partial \dot{\sigma}}{\partial \sigma} + \frac{1}{\rho} \frac{D\rho}{Dt} = 0. \quad (44.81)$$

Now use the material time derivative in the form (44.50b) to derive the Eulerian expression of mass conservation

$$\frac{\partial(\rho z_\sigma)}{\partial t} + \nabla_\sigma \cdot (\rho z_\sigma \mathbf{u}) + \frac{\partial(\rho z_\sigma \dot{\sigma})}{\partial \sigma} = 0, \quad (44.82)$$

where the time derivative is computed holding σ fixed. We can furthermore introduce the dia-surface velocity component

$$w^{(\dot{\sigma})} = z_\sigma \dot{\sigma} \quad (44.83)$$

so that mass continuity takes the form

$$\frac{\partial(\rho z_\sigma)}{\partial t} + \nabla_\sigma \cdot (\rho z_\sigma \mathbf{u}) + \frac{\partial(\rho w^{(\dot{\sigma})})}{\partial \sigma} = 0. \quad (44.84)$$

Alternatively, we can reintroduce the material time derivative operator to write the mass continuity equation (44.82) in the form

$$\frac{1}{\rho z_\sigma} \frac{D(\rho z_\sigma)}{Dt} = -(\nabla_\sigma \cdot \mathbf{u} + \partial\dot{\sigma}/\partial\sigma), \quad (44.85)$$

where we used equation (44.50b) to write

$$\frac{D}{Dt} = \left[\frac{\partial}{\partial t} \right]_\sigma + \mathbf{u} \cdot \nabla_\sigma + \dot{\sigma} \frac{\partial}{\partial \sigma}. \quad (44.86)$$

44.10 Layer integrated mass continuity

The formulation thus far has been continuous, with the only assumption made that the specific thickness, $h = \partial z / \partial \sigma$, is single signed. We here consider a discrete increment in the generalized vertical coordinate,

$$\sigma - \delta\sigma/2 \leq \sigma' \leq \sigma + \delta\sigma/2, \quad (44.87)$$

and formulate the mass budget over this layer whose thickness is given by

$$h \equiv \int_{z(\sigma-\delta\sigma/2)}^{z(\sigma+\delta\sigma/2)} dz = \int_{\sigma-\delta\sigma/2}^{\sigma+\delta\sigma/2} \frac{\partial z}{\partial \sigma} d\sigma, \quad (44.88)$$

and whose mass per horizontal area is

$$\delta m = \int_{z(\sigma-\delta\sigma/2)}^{z(\sigma+\delta\sigma/2)} \rho dz = \int_{\sigma-\delta\sigma/2}^{\sigma+\delta\sigma/2} \rho z_\sigma d\sigma = \bar{\rho} h, \quad (44.89)$$

where $\bar{\rho}$ is the layer averaged density. Note that for Boussinesq fluids the mass per area equals to the layer thickness times the reference density

$$\delta m = \rho_0 h \quad \text{Boussinesq.} \quad (44.90)$$

As defined by equation (44.88) and illustrated in Figure 44.2, the thickness of a layer is relatively large in regions where $\partial\sigma/\partial z$ is small; i.e., in regions where σ is weakly stratified in the vertical. Conversely, the layer thickness is relatively small where the vertical stratification is large. Furthermore, if the specific thickness is negative, then the layer thickness remains positive by choosing $\delta\sigma < 0$. For example, in a stably stratified fluid with σ given by potential density, $\partial\sigma/\partial z = -(g/\rho_0) N^2 < 0$ so that we take $\delta\sigma < 0$ to move vertically upward in the water column to regions of lower potential density. The same situation holds when σ is the hydrostatic pressure in which $\partial p/\partial z = -\rho g$ (Section 44.10.2).

The formulation in this section, and its companion for tracers in Section 44.11, holds across all generalized vertical coordinates, even incorporating the trivial case of geopotential coordinates ($\sigma = z$) whereby the specific thickness is unity. Application of the resulting layer integrated kinematics include the development of discrete equations for numerical layered models (see [Griffies et al. \(2020\)](#) for a review), as well as the shallow water models discussed in Part V of this book.

44.10.1 Compressible fluids

Performing a layer integral of the specific thickness equation (44.84) renders

$$\int_{\sigma-\delta\sigma/2}^{\sigma+\delta\sigma/2} \left[\frac{\partial(\rho z_\sigma)}{\partial t} + \nabla_\sigma \cdot (\rho z_\sigma \mathbf{u}) + \frac{\partial(\rho w^{(\dot{\sigma})})}{\partial \sigma} \right] d\sigma = 0. \quad (44.91)$$

The dia-surface term integrates to a finite difference across the layer

$$\int_{\sigma-\delta\sigma/2}^{\sigma+\delta\sigma/2} \left[\frac{\partial(\rho z_\sigma)}{\partial t} + \nabla_\sigma \cdot (\rho z_\sigma \mathbf{u}) \right] = -\Delta_\sigma(\rho w^{(\dot{\sigma})}), \quad (44.92)$$

where we introduced the finite difference operator for properties defined at the layer interface

$$\Delta_\sigma(A) = A(\sigma + \delta\sigma/2) - A(\sigma - \delta\sigma/2). \quad (44.93)$$

The time derivative and horizontal space derivative commute with the layer integral, since the limits are specified fixed values for the layer increment, $\delta\sigma$, and the derivatives are computed with σ fixed. Hence, layer mass continuity takes the form

$$\left[\frac{\partial}{\partial t} \right]_\sigma \int_{\sigma-\delta\sigma/2}^{\sigma+\delta\sigma/2} \rho z_\sigma d\sigma + \nabla_\sigma \cdot \int_{\sigma-\delta\sigma/2}^{\sigma+\delta\sigma/2} \rho \mathbf{u} z_\sigma d\sigma = -\Delta_\sigma(\rho w^{(\dot{\sigma})}). \quad (44.94)$$

The first term involves the layer averaged density times the layer thickness as per equation (44.89). The second term involves the layer averaged density-weighted velocity, which is the layer averaged horizontal mass flux

$$\int_{\sigma-\delta\sigma/2}^{\sigma+\delta\sigma/2} \rho \mathbf{u} z_\sigma d\sigma = h \bar{\rho} \bar{\mathbf{u}}. \quad (44.95)$$

We are thus led to the layer integrated continuity equation

$$\left[\frac{\partial(h \bar{\rho})}{\partial t} \right]_\sigma + \nabla_\sigma \cdot (h \bar{\rho} \bar{\mathbf{u}}) + \Delta_\sigma(\rho w^{(\dot{\sigma})}) = 0. \quad (44.96)$$

When evolving the fields in a discrete numerical model, we have information only about layer averaged fields. So how do we estimate the depth average of the horizontal advective flux, $\bar{\rho} \bar{\mathbf{u}}$, appearing in equation (44.96)? One method interprets all fields as their layer averaged values so that $\bar{\rho} \bar{\mathbf{u}} = \bar{\rho} \bar{\mathbf{u}}$, thus considering uncomputed sub-layer correlations $\bar{\rho}' \bar{\mathbf{u}'}$ as part of the truncation error. Alternately, we note that compressible hydrostatic flows can be described by a pressure-based vertical coordinate in which case the layer mass per horizontal area is proportional to a prescribed increment in pressure

$$\delta m = \int_{\sigma-\delta\sigma/2}^{\sigma+\delta\sigma/2} \rho z_\sigma d\sigma = \bar{\rho} h = -g^{-1} \delta p. \quad (44.97)$$

Correspondingly, the layer integrated horizontal mass flux equals to the mass increment times the pressure-layer averaged velocity

$$\int_{\sigma-\delta\sigma/2}^{\sigma+\delta\sigma/2} \rho \mathbf{u} z_\sigma d\sigma = -g^{-1} \int_{p-\delta p/2}^{p+\delta p/2} \mathbf{u} dp = -g^{-1} \bar{\mathbf{u}} \delta p = h \bar{\rho} \bar{\mathbf{u}}. \quad (44.98)$$

With either of the above two methods, we are led to the same layer integrated continuity equation, which we write in the generic form that drops overbars

$$\left[\frac{\partial(h \rho)}{\partial t} \right]_\sigma + \nabla_\sigma \cdot (h \rho \mathbf{u}) + \Delta_\sigma(\rho w^{(\dot{\sigma})}) = 0. \quad (44.99)$$

We illustrate contributions to this layer mass budget in Figure 44.7.

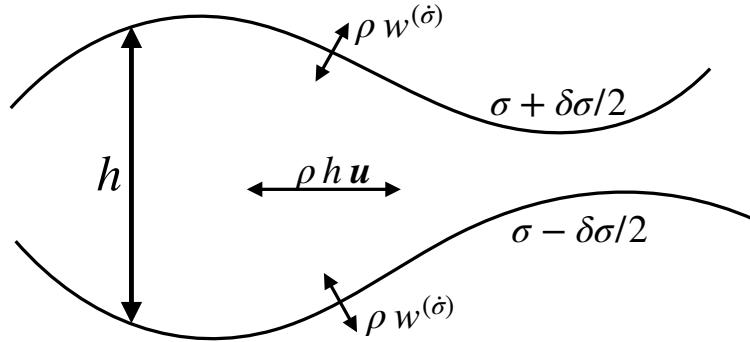


FIGURE 44.7: Illustrating the terms contributing to changes in layer mass according to the layer integrated continuity equation (44.99). The discrete layer is shown here with bounding interfaces at $\sigma - \delta\sigma/2$ and $\sigma + \delta\sigma/2$. Within a layer there is a horizontal redistribution due to horizontal advective transport. Additionally, matter can cross the layer due to dia-surface transport via $w^{(\dot{\sigma})}$.

44.10.2 Mass continuity using pressure coordinates

Let us here consider in some detail the special case of pressure coordinates in a hydrostatic fluid, and thus derive mass continuity using these coordinates.

Method I

The thickness of a hydrostatic pressure layer (equation (44.88)) takes on the following form

$$h = \int_{p-\delta p/2}^{p+\delta p/2} \frac{\partial z}{\partial p} dp = - \int_{p-\delta p/2}^{p+\delta p/2} \frac{dp}{\rho g}, \quad (44.100)$$

so that its mass per unit area is

$$\int_{p-\delta p/2}^{p+\delta p/2} \rho \frac{\partial z}{\partial p} dp = -\delta p/g. \quad (44.101)$$

The mass continuity equation (44.99) thus becomes

$$\frac{\partial(\delta p)}{\partial t} + \nabla_p \cdot (\mathbf{u} \delta p) + \Delta_p (\dot{p}) = 0. \quad (44.102)$$

The partial time derivative vanishes since it is computed by holding pressure fixed so that the pressure increment has a zero time tendency

$$\left[\frac{\partial(\delta p)}{\partial t} \right]_p = 0. \quad (44.103)$$

Likewise, $\nabla_p(\delta p) = 0$. Thus, we can divide by δp to render the continuity equation

$$\nabla_p \cdot \mathbf{u} + \frac{\partial \dot{p}}{\partial p} = 0 \quad \text{compressible hydrostatic.} \quad (44.104)$$

This equation is isomorphic to the incompressible continuity equation (for both hydrostatic and non-hydrostatic fluids) written using geopotential coordinates

$$\nabla_z \cdot \mathbf{u} + \frac{\partial \dot{z}}{\partial z} = \nabla_z \cdot \mathbf{u} + \frac{\partial w}{\partial z} 0 \quad \text{incompressible,} \quad (44.105)$$

where $w = \dot{z}$ is the vertical component to the velocity vector. For both compressible pressure coordinates and incompressible depth coordinates, the continuity equation is a diagnostic relation (i.e., no time derivatives) rather than prognostic (i.e., containing time derivatives).

Method II

For the second method we make use of the approach detailed in Section 44.9.2, which starts from

$$\frac{D(\rho \delta V)}{Dt} = 0. \quad (44.106)$$

In pressure coordinates the volume of the fluid element takes the form

$$\delta V = \delta x \delta y \delta z = \delta x \delta y \left[\frac{\partial z}{\partial p} \right] \delta p = -(\rho g)^{-1} \delta x \delta y \delta p. \quad (44.107)$$

Consequently,

$$0 = \frac{D(\rho \delta V)}{Dt} = g^{-1} \left(\frac{D(\delta x \delta y \delta p)}{Dt} \right), \quad (44.108)$$

so that

$$0 = \frac{1}{\delta x \delta y \delta p} \left(\frac{D(\delta x \delta y \delta p)}{Dt} \right) = \nabla_p \cdot \mathbf{u} + \frac{\partial \dot{p}}{\partial p}. \quad (44.109)$$

The second step made use of the isomorphism between this result and that for equation (14.72) that holds for a geopotential vertical coordinate.

44.10.3 Incompressible flow

Specializing to an incompressible volume conserving flow (see Chapters 17 and 24) yields the incompressible layer thickness equation

$$\frac{\partial h}{\partial t} + \nabla_\sigma \cdot (h \mathbf{u}) + \Delta_\sigma w^{(\dot{\sigma})} = 0. \quad (44.110)$$

Further specializing to the case of zero dia-surface transport leads to

$$\frac{\partial h}{\partial t} + \nabla_\sigma \cdot (h \mathbf{u}) = 0 \quad \text{no dia-surface transport.} \quad (44.111)$$

This case is commonly studied for adiabatic fluids using isopycnal coordinates, in which isopycnal surfaces are material (Section 46.1).

44.10.4 Rescaled geopotential coordinates

The rescaled geopotential coordinate

$$z^* = \frac{H(z - \eta)}{H + \eta} = \frac{\eta_b(z - \eta)}{\eta_b - \eta} \quad \text{and} \quad \eta_b(x, y) \leq z^* \leq 0, \quad (44.112)$$

is commonly used in Boussinesq ocean models, where $z = \eta(x, y, t)$ is the ocean free surface and $z = \eta_b(x, y) = -H(x, y)$ is the ocean bottom. The thickness of a coordinate layer is given by

$$h = dz = \frac{\partial z}{\partial z^*} dz^* = (1 + \eta/H) dz^* = (1 - \eta/\eta_b) dz^*. \quad (44.113)$$

The depth integrated column thickness and depth integrated coordinate thickness are given by

$$\int_{\eta_b}^{\eta} dz = \eta - \eta_b = \eta + H \quad \text{and} \quad \int_{z^*(\eta_b)}^{z^*(\eta)} dz^* = -\eta_b = H. \quad (44.114)$$

Correspondingly, the depth integrated thickness equation is given by the depth integrated volume budget derived in Section 17.3

$$\frac{\partial \eta}{\partial t} + \nabla \cdot \mathbf{U} + [w_{z^*=0}^{(\dot{\sigma})} - w_{z^*=\eta_b}^{(\dot{\sigma})}] = 0. \quad (44.115)$$

We assume no volume flow through the ocean bottom so that $w_{z^*=\eta_b}^{(\dot{\sigma})} = 0$, whereas

$$-\rho_0 w_{z^*=0}^{(\dot{\sigma})} = Q_m \quad (44.116)$$

is the mass flux crossing the ocean free surface (Section 15.4.3).

44.11 Layer integrated tracer equation

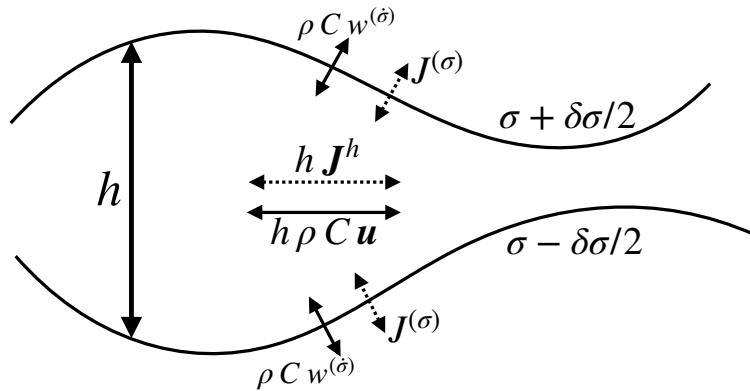


FIGURE 44.8: Illustrating the terms contributing to changes in layer tracer content according to the layer integrated tracer equation (44.121). The layer is shown here with bounding interfaces at $\sigma - \delta\sigma/2$ and $\sigma + \delta\sigma/2$. Within a layer there is a redistribution of tracer due to horizontal advective and subgrid scale tracer fluxes. Additionally, matter can cross the layer due to dia-surface transport via $\rho C w^{(\dot{\sigma})}$ and subgrid tracer transport $J^{(\sigma)}$.

The tracer equation from Section 16.1.3 is given by

$$\rho \frac{DC}{Dt} = -\nabla \cdot \mathbf{J}, \quad (44.117)$$

where \mathbf{J} is a subgrid scale flux. Now introduce the material time derivative operator in the form (44.50b) to have

$$\rho \left[\frac{\partial C}{\partial t} + \mathbf{u} \cdot \nabla_\sigma C + \dot{\sigma} \partial_\sigma C \right] = -\nabla \cdot \mathbf{J}, \quad (44.118)$$

Multiplying by the specific thickness and making use of the mass conservation equation (44.84) renders the flux-form Eulerian equation

$$\frac{\partial(z_\sigma \rho C)}{\partial t} + \nabla_\sigma \cdot (z_\sigma \rho C \mathbf{u}) + \frac{\partial(\rho C w^{(\dot{\sigma})})}{\partial \sigma} = - \left[\nabla_\sigma \cdot (z_\sigma \mathbf{J}^h) + \frac{\partial(z_\sigma \nabla \sigma \cdot \mathbf{J})}{\partial \sigma} \right], \quad (44.119)$$

where we made use of expression (43.82) for the subgrid scale operator. Now perform a layer integral as detailed in Section 44.10 and use the layer mass continuity equation (44.99) to yield the layer integrated tracer equation

$$\frac{\partial(h \rho C)}{\partial t} + \nabla_\sigma \cdot (h \rho C \mathbf{u}) + \Delta_\sigma(\rho C w^{(\dot{\sigma})}) = - \left[\nabla_\sigma \cdot (h \mathbf{J}^h) + \Delta_\sigma(z_\sigma \nabla \sigma \cdot \mathbf{J}) \right]. \quad (44.120)$$

Alternatively, we can bring all terms to the left hand side to yield

$$\frac{\partial(h \rho C)}{\partial t} + \nabla_\sigma \cdot (h \rho C \mathbf{u} + h \mathbf{J}^h) + \Delta_\sigma(\rho C w^{(\dot{\sigma})} + J^{(\sigma)}) = 0 \quad (44.121)$$

where we wrote

$$J^{(\sigma)} = z_\sigma \nabla \sigma \cdot \mathbf{J}. \quad (44.122)$$

We illustrate contributions to the layer tracer budget (44.121) in Figure 44.8. Note that we interpret these layer integrated fields and fluxes as per the discussion in Section 44.10.1.



Dynamical equations

In this chapter we derive the dynamical equations for momentum, vorticity, and potential vorticity using generalized vertical coordinates. These equations provide the foundations for many numerical models of the atmosphere and ocean. Besides being essential for developing methods for numerical simulations, understanding the physical and mathematical basis of these equations supports the analysis of simulations.

READER'S GUIDE TO THIS CHAPTER

We assume a working knowledge of the mathematics of generalized vertical coordinates as detailed in Chapter 43 and the corresponding kinematics in Chapter 44. We make particular use of the layer integrated notions introduced for mass continuity and the tracer equations in Sections 44.10 and 44.11. We also make use of the dynamical equations derived in Chapter 18. For most purposes in this chapter we find Cartesian horizontal coordinates sufficient. However, we note some places where spherical coordinates warrant special consideration.

45.1	Equations of motion	1028
45.1.1	Notation	1028
45.1.2	Mass and tracer equations	1028
45.1.3	Momentum equation	1028
45.1.4	Eulerian flux-form horizontal momentum equation	1029
45.1.5	Vector-invariant horizontal momentum equation	1029
45.1.6	Hydrostatic balance with constant gravitational acceleration	1030
45.2	Concerning the pressure force	1031
45.2.1	Computing the horizontal pressure gradient	1031
45.2.2	Integrated pressure force on the cell faces	1032
45.2.3	Net vertical pressure force	1033
45.2.4	Net horizontal pressure force	1033
45.2.5	Comments	1034
45.3	Hydrostatic vorticity and potential vorticity	1034
45.3.1	Basic manipulations	1034
45.3.2	Vorticity and potential vorticity equation	1035
45.3.3	Boussinesq fluid	1036
45.4	Exercise	1036

45.1 Equations of motion

We here derive the equations of motion based on generalized vertical coordinates. The scalar equations were already discussed in Sections 44.9, 44.10, and 44.11, so our main focus concerns the momentum equation.

45.1.1 Notation

For much of this chapter we focus on the continuous formulation of the generalized vertical coordinates. Following the discussion in Section 44.3, we encounter the specific thickness throughout the equations

$$\hbar = \frac{\partial z}{\partial \sigma} = z_\sigma. \quad (45.1)$$

Using this notation we write the dia-surface transport velocity as

$$w^{(\dot{\sigma})} = \frac{\partial z}{\partial \sigma} \dot{\sigma} = \hbar \dot{\sigma}, \quad (45.2)$$

and the dia-surface advection operator is

$$w^{(\dot{\sigma})} \partial_z = \dot{\sigma} \partial_\sigma. \quad (45.3)$$

45.1.2 Mass and tracer equations

The mass and tracer equations were derived in Sections 44.9, 44.10, and 44.11, with their continuous vertical coordinate formulation given by

$$\frac{\partial(\rho \hbar)}{\partial t} + \nabla_\sigma \cdot (\rho \hbar \mathbf{u}) + \partial_\sigma(\rho \hbar \dot{\sigma}) = 0 \quad (45.4a)$$

$$\frac{\partial(\hbar \rho C)}{\partial t} + \nabla_\sigma \cdot (\hbar \rho C \mathbf{u} + \hbar \mathbf{J}^h) + \partial_\sigma(\rho \hbar \dot{\sigma} C + \hbar \nabla_\sigma \cdot \mathbf{J}) = 0. \quad (45.4b)$$

Compatibility is maintained between the mass continuity equation (45.4a) and the tracer equation (45.4b) so long as the tracer equation reduces to the mass equation upon setting the tracer concentration to a spatial constant. Hence, for compatibility we must have the subgrid fluxes, \mathbf{J} , vanish when the tracer is a spatial constant. For example, diffusive fluxes, which are proportional to the tracer gradient, respect this constraint. These properties originate from our discussion of mass budgets and the barycentric velocity in Section 16.1.

45.1.3 Momentum equation

From Section 22.10, the horizontal and vertical components to the momentum equation are

$$\rho \frac{D\mathbf{u}}{Dt} + 2\rho \boldsymbol{\Omega} \wedge \mathbf{u} = -\rho \nabla_z \Phi - \nabla_z p + \rho \mathbf{F}^h \quad (45.5a)$$

$$\rho \frac{Dw}{Dt} = -\rho \frac{\partial \Phi}{\partial z} - \frac{\partial p}{\partial z} + \rho F^z. \quad (45.5b)$$

For the simple form of the geopotential, $\Phi = g z$ (Section 10.12), so that the horizontal gradient of the geopotential vanishes

$$\Phi = g z \implies \nabla_z \Phi = 0. \quad (45.6)$$

This gradient is nonzero in the presence of astronomical tide forcing (Chapter 29).

Horizontal momentum equation

We transform the horizontal derivatives from geopotential coordinates to generalized vertical coordinates according to (see equation (43.64))

$$\nabla_z = \nabla_\sigma - (\nabla_\sigma z) \partial_z \quad (45.7)$$

thus leading to the horizontal momentum equation

$$\rho \frac{D\mathbf{u}}{Dt} + 2\rho \boldsymbol{\Omega} \wedge \mathbf{u} = -\rho [\nabla_\sigma - (\nabla_\sigma z) \partial_z] \Phi - [\nabla_\sigma - (\nabla_\sigma z) \partial_z] p + \rho \mathbf{F}^h. \quad (45.8)$$

In Section 45.1.6 we present some special cases for this equation that simplify the pressure and geopotential terms.

Vertical momentum equation

The vertical momentum equation is transformed into

$$\rho \frac{Dw}{Dt} = -\frac{\partial \sigma}{\partial z} \left[\rho \frac{\partial \Phi}{\partial \sigma} + \frac{\partial p}{\partial \sigma} \right] + \rho F^z, \quad (45.9)$$

with the hydrostatic form given by

$$\frac{\partial p}{\partial \sigma} = -\rho \frac{\partial \Phi}{\partial \sigma}. \quad (45.10)$$

45.1.4 Eulerian flux-form horizontal momentum equation

Using Cartesian horizontal coordinates and generalized vertical coordinates, the horizontal momentum equation includes a contribution from the acceleration that has a form similar to that for a tracer (Section 44.11)

$$\hbar \rho \frac{Du}{Dt} = \left[\frac{\partial(\hbar \rho u)}{\partial t} \right]_\sigma + \nabla_\sigma \cdot (\hbar \rho u \mathbf{u}) + \partial_\sigma(\hbar \rho u \dot{\sigma}) \quad (45.11a)$$

$$\hbar \rho \frac{Dv}{Dt} = \left[\frac{\partial(\hbar \rho v)}{\partial t} \right]_\sigma + \nabla_\sigma \cdot (\hbar \rho v \mathbf{u}) + \partial_\sigma(\hbar \rho v \dot{\sigma}). \quad (45.11b)$$

We provide a σ subscript on the time derivative operator to signal that this derivative is taken with σ held fixed. With spherical coordinates there are additional metric terms appearing on the right hand side, as detailed in Section 18.2.1. In particular, there is a metric term that contains the vertical velocity component, $w = Dz/Dt$. The appearance of w is awkward since the vertical velocity is not naturally computed using generalized vertical coordinates. This limitation is overcome through use of the vector-invariant velocity equation derived in Section 45.1.5.

45.1.5 Vector-invariant horizontal momentum equation

As noted in Section 18.2.2, the *vector-invariant* form of the momentum equation eliminates the metric terms that appear in the non-Cartesian flux-form equations. The vector-invariant form is also suited for deriving the vorticity equation (see Section 45.3). Here, we start with the material time derivative in the form (44.50c) appropriate for generalized vertical coordinates, in which case the horizontal acceleration is given by

$$\frac{D\mathbf{u}}{Dt} = \left[\frac{\partial \mathbf{u}}{\partial t} \right]_\sigma + (\mathbf{u} \cdot \nabla_\sigma) \mathbf{u} + (\dot{\sigma} \partial_\sigma) \mathbf{u}. \quad (45.12)$$

Now make use of the vector identity (see Section 2.3.4)

$$(\mathbf{u} \cdot \nabla_\sigma) \mathbf{u} = \nabla_\sigma K + (\nabla_\sigma \wedge \mathbf{u}) \wedge \mathbf{u}, \quad (45.13)$$

where

$$K = \mathbf{u} \cdot \mathbf{u}/2 \quad (45.14)$$

is the kinetic energy per mass of the horizontal flow. Introducing the GVC version of the relative vorticity (see Section 46.2.1)

$$\tilde{\zeta} \equiv \hat{\mathbf{z}} \cdot (\nabla_\sigma \wedge \mathbf{u}) = \left[\frac{\partial v}{\partial x} \right]_\sigma - \left[\frac{\partial u}{\partial y} \right]_\sigma \quad (45.15)$$

renders

$$\frac{D\mathbf{u}}{Dt} = \left[\frac{\partial \mathbf{u}}{\partial t} \right]_\sigma + \nabla_\sigma K + \tilde{\zeta} \hat{\mathbf{z}} \wedge \mathbf{u} + \dot{\sigma} \partial_\sigma \mathbf{u}, \quad (45.16)$$

so that the horizontal momentum equation takes the form

$$\left[\frac{\partial \mathbf{u}}{\partial t} \right]_\sigma + \dot{\sigma} \frac{\partial \mathbf{u}}{\partial \sigma} + (2\boldsymbol{\Omega} + \hat{\mathbf{z}} \tilde{\zeta}) \wedge \mathbf{u} = -\nabla_\sigma K - \nabla_z \Phi - (1/\rho) \nabla_z p + \mathbf{F}^h, \quad (45.17)$$

where again $\nabla_z = \nabla_\sigma - (\nabla_\sigma z) \partial_z$ as per equation (45.7). This equation is form-invariant regardless the horizontal coordinates, thus motivating the name *vector-invariant*.¹

45.1.6 Hydrostatic balance with constant gravitational acceleration

There are many special cases that simplify various terms in the momentum equation. For example, when considering a geopotential in the form $\Phi = g z$ (Section 10.12) with g assumed to be a constant effective gravitational acceleration, then the horizontal momentum equation (45.8) becomes

$$\rho \frac{D\mathbf{u}}{Dt} + 2\rho \boldsymbol{\Omega} \wedge \mathbf{u} = -[\nabla_\sigma - (\nabla_\sigma z) \partial_z] p + \rho \mathbf{F}^h. \quad (45.18)$$

Furthermore, assuming an approximate hydrostatic balance (and corresponding simplification of the Coriolis acceleration as per Section 23.1.3) allows us to write $\partial p / \partial z = -g \rho$ so that

$$\rho \frac{D\mathbf{u}}{Dt} + \rho f \hat{\mathbf{z}} \wedge \mathbf{u} = -[\nabla_\sigma p + \rho \nabla_\sigma \Phi] + \rho \mathbf{F}^h, \quad (45.19)$$

which also takes on the vector-invariant form

$$\left[\frac{\partial \mathbf{u}}{\partial t} \right]_\sigma + \dot{\sigma} \frac{\partial \mathbf{u}}{\partial \sigma} + (f + \tilde{\zeta}) \hat{\mathbf{z}} \wedge \mathbf{u} = -\nabla_\sigma (K + \Phi) - (1/\rho) \nabla_\sigma p + \mathbf{F}^h. \quad (45.20)$$

This form is commonly used for hydrostatic models of the ocean and atmosphere, such as discussed in [Griffies et al. \(2020\)](#).

¹See Section 4.4.4 of [Griffies \(2004\)](#) for a detailed derivation using arbitrary horizontal coordinates.

45.2 Concerning the pressure force

As seen in Section 19.1.3, the pressure force acting on a fluid region is given by the integral

$$\mathbf{F}^{\text{pressure}} = - \oint_{\partial\mathcal{R}} p \hat{\mathbf{n}} d\mathcal{S} = - \int_{\mathcal{R}} \nabla p dV, \quad (45.21)$$

where the second equality follows from Gauss's divergence theorem applied to a scalar field (Section 2.7.2). We refer to the right-most expression as the pressure gradient body force, and this expression is the basis for the discussion in Sections 45.1.6 and 45.2.1. In this formulation, the pressure force at a point is oriented down the pressure gradient, so that the net pressure force acting on a region is the volume integral of pressure gradient.

The middle expression in equation (45.21) formulates the pressure force as the area integrated pressure contact force acting on the region boundaries. In this form, the pressure acting on a region is computed as the integral of pressure over the area bounding the region, with the orientation determined by the inward normal at each point on the boundary. Much of this section is concerned with the contact force expression as a basis for computing the pressure force acting on a finite region as shown in Figure 45.2. The contact force perspective was taken by [Lin \(1997\)](#) and [Adcroft et al. \(2008\)](#) in their finite volume approach to computing the pressure force acting on a numerical model grid cell.

45.2.1 Computing the horizontal pressure gradient

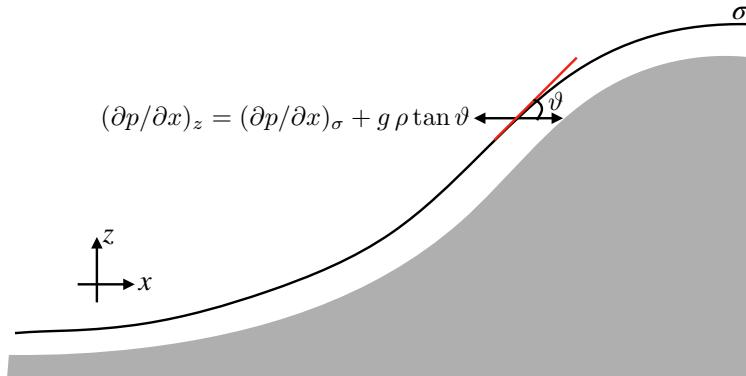


FIGURE 45.1: Illustrating how the horizontal pressure gradient is decomposed into two terms, one aligned with the surface of constant σ , and another associated with the slope of the σ -surface relative to the horizontal, $\tan \vartheta = (\partial z / \partial x)_\sigma$. We here consider the decomposition using terrain following vertical coordinates, where the vertical coordinate is aligned according to the solid-earth bottom (shaded region). Specifically, for terrain following ocean models we have $\sigma = (z - \eta) / (-\eta_b + \eta)$, where $z = \eta(x, y, t)$ is the ocean free surface and $z = \eta_b(x, y)$ is the ocean bottom. Terrain-following atmospheric have a similar definition, often using pressure rather than geopotential so that $\sigma = (p - p_a) / (p_b - p_a)$, where p is the pressure, $p_a = p_a(x, y, t)$ is the pressure applied at the top of the atmosphere (often assumed to be zero), and $p_b = p_b(x, y, t)$ is the pressure at the bottom of the atmosphere.

The horizontal pressure gradient is aligned perpendicular to the local gravitational direction. It is generally among the dominant horizontal forces acting on a fluid element. Hence, its accurate representation in numerical models is crucial for the physical integrity of a simulation. Unfortunately, decomposition of the horizontal pressure gradient into two terms according to the transformation (45.7) can lead to numerical difficulties. For example, with a simple geopotential and a hydrostatic

fluid, equation (45.19) shows that the horizontal pressure gradient takes the form

$$\nabla_z p = \nabla_\sigma p + \rho \nabla_\sigma \Phi = \nabla_\sigma p + g \rho \nabla_\sigma z, \quad (45.22)$$

with this decomposition illustrated in Figure 45.1 for the case of terrain following vertical coordinates. Numerical difficulties occur when the two terms on the right hand side have comparable magnitude but distinct signs. We are thus confronted with computing the small difference between two large numbers, and that situation generally exposes a numerical simulation to nontrivial truncation errors. Unfortunately, these errors can corrupt the integrity of the computed pressure forces and in turn contribute to spurious flow. An overview of this issue for ocean models is given by [Haney \(1991\)](#), [Mellor et al. \(1998\)](#), [Griffies et al. \(2000\)](#), with advances offered by [Lin \(1997\)](#), [Shchepetkin and McWilliams \(2002\)](#), and [Adcroft et al. \(2008\)](#). In the remainder of this section, we outline a finite volume method for computing the pressure force as proposed by [Lin \(1997\)](#) for atmospheric models and [Adcroft et al. \(2008\)](#) for ocean models. This approach starts from the middle expression in equation (45.21) for the pressure force; i.e., it formulates the pressure force as the area integral of the pressure contact force rather than the volume integral of the pressure gradient force.

45.2.2 Integrated pressure force on the cell faces

The inward normal on the grid cell vertical side boundaries points in the horizontal direction. For example, on the left side of Figure 45.2 the pressure force acts in the positive \hat{y} direction

$$\mathbf{F}_{\text{left}}^{\text{press}} = \hat{y} \int_{\text{left}} p \, dx \, dz \quad (45.23)$$

whereas pressure on the right wall acts in the opposite direction

$$\mathbf{F}_{\text{right}}^{\text{press}} = -\hat{y} \int_{\text{right}} p \, dx \, dz. \quad (45.24)$$

Similar expressions appear for the front and back vertical boundaries acting in the \hat{x} direction.

Since the top and bottom boundaries of the grid cell are sloped, there is a pressure force acting on this surface directed in both the horizontal and vertical directions. To unpack the form of this force, write the vertical position of a point on the top interface as $z = \eta(x, y, t)$ so that the outward normal is given by

$$\hat{\mathbf{n}} = \frac{\nabla(z - \eta)}{|\nabla(z - \eta)|} = \frac{\hat{z} - \nabla\eta}{\sqrt{1 + |\nabla\eta|^2}}. \quad (45.25)$$

Following our discussion of dia-surface transport in Section 44.4.5, we know that the product of the normal direction and the area element can be written

$$\hat{\mathbf{n}} \, d\mathcal{S} = (\hat{z} - \nabla\eta) \, dA, \quad (45.26)$$

where $dA = dx \, dy$ is the horizontal projection of the area element (see Figure 44.3). Hence, the net pressure force acting on the top face is given by

$$\mathbf{F}_{\text{top}}^{\text{press}} = -\hat{z} \int_{\text{top}} p \, dx \, dy + \hat{x} \int_{\text{top}} p (\partial z / \partial x)_\sigma \, dx \, dy + \hat{y} \int_{\text{top}} p (\partial z / \partial y)_\sigma \, dx \, dy, \quad (45.27)$$

where we set $z = \eta$ in the second and third terms and placed a σ subscript to emphasize that the horizontal derivative is taken with σ held constant. Notice that the pressure acts in the positive

horizontal direction if the top surface slopes upward (surface shoaling) when moving in either of the two horizontal directions. Pressure acting on the bottom face has the same appearance yet with opposite signs

$$\mathbf{F}_{\text{bott}}^{\text{press}} = \hat{z} \int_{\text{bott}} p \, dx \, dy - \hat{x} \int_{\text{bott}} p (\partial z / \partial x)_\sigma \, dx \, dy - \hat{y} \int_{\text{bott}} p (\partial z / \partial y)_\sigma \, dx \, dy. \quad (45.28)$$

The pressure acts in the positive horizontal direction if the bottom surface slopes downward (surface deepens) when moving in either of the two horizontal directions. As discussed in Section 19.1.3, the horizontal pressure acting on a sloped surface is known as *form stress*. Here the sloped surface is defined by a constant generalized vertical coordinate.

45.2.3 Net vertical pressure force

Bringing the pieces together leads to the net vertical pressure force acting on the grid cell

$$\mathbf{F}_{\text{vertical}}^{\text{press}} = -\hat{z} \left[\int_{\text{top}} p \, dx \, dy - \int_{\text{bott}} p \, dx \, dy \right]. \quad (45.29)$$

If the fluid is in hydrostatic balance, then this vertical force is given by the weight of fluid within the cell

$$\mathbf{F}_{\text{vertical}}^{\text{press}} = \hat{z} M g, \quad (45.30)$$

where M is the mass of fluid in the grid cell. The net vertical hydrostatic pressure force acts vertically upward since hydrostatic pressure at the cell bottom is greater than at the cell top.

45.2.4 Net horizontal pressure force

The net meridional pressure force is given by the forces acting on the sides as well as those acting on the sloped top and bottom boundaries

$$\mathbf{F}_{\text{merid}}^{\text{press}} = \left[\int_{\text{left}} p \, dx \, dz - \int_{\text{right}} p \, dx \, dz \right] + \left[\int_{\text{top}} p (\partial z / \partial y)_\sigma \, dx \, dy - \int_{\text{bott}} p (\partial z / \partial y)_\sigma \, dx \, dy \right]. \quad (45.31)$$

We can write this expression in a more compact form by orienting our integration in a counter-clockwise manner around the cell boundaries, and making use of the identity $(\partial z / \partial y)_\sigma \, dy = dz$ on the top and bottom faces, so that

$$\mathbf{F}_{\text{merid}}^{\text{press}} = -\oint p \, dx \, dz. \quad (45.32)$$

For some purposes it is more convenient to work with the geopotential, $\Phi = g z$, than the pressure. In this case we can write the meridional pressure force as

$$\mathbf{F}_{\text{merid}}^{\text{press}} = -\oint p \, dx \, dz = -\oint dx [d(pz) - z \, dp] = g^{-1} \oint \Phi \, dx \, dp, \quad (45.33)$$

where $\oint dx \, d(pz) = 0$. This form is useful with compressible / non-Boussinesq models, in which pressure is a natural vertical coordinate (e.g., see the caption to Figure 45.1).

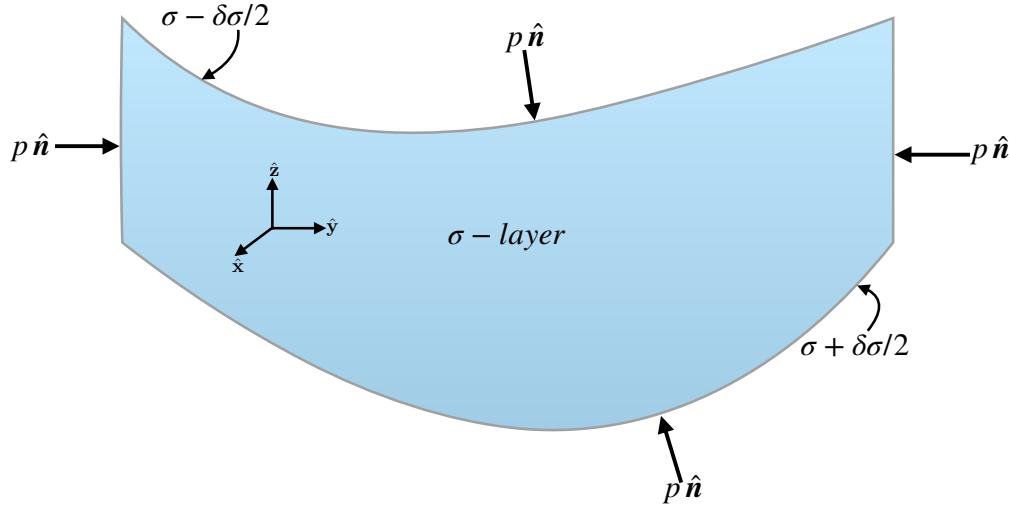


FIGURE 45.2: Schematic of pressure forces acting on the boundaries of a finite region such as a discrete model grid cell. In generalized vertical coordinate models, the side faces are vertical, so that pressure acts only in the horizontal directions. The top and bottom faces are defined by surfaces of constant generalized vertical coordinates with depth $\sigma(x, y, z, t) = \text{constant}$. We assume that these surfaces have an outward normal that has a nonzero projection into the vertical so that we can write the depth of a point on these surfaces as $z = \eta(x, y, t)$. Because of the slope of the top and bottom surfaces, pressure has both a horizontal and vertical component when acting on these surfaces. The net pressure acting on the grid cell is given by the area integral of the pressures around the grid cell boundary.

45.2.5 Comments

A numerical realization of the integrated contact pressure force requires a representation of pressure along the boundaries of the grid cell. A variety of methods are available with differing accuracies. [Adcroft et al. \(2008\)](#) are notable in proposing an analytic form that allows for an exact integration along the cell faces in special cases, and a highly accurate numerical integration in other cases. In general, this method for computing pressure forces is highly suited to generalized vertical coordinate grid cells, which was the motivation offered by [Lin \(1997\)](#) in the context of terrain following atmospheric models.

45.3 Hydrostatic vorticity and potential vorticity

Generalized vertical coordinates are most commonly used to study hydrostatic fluids. We are thus motivated to develop the equation for the vertical component of vorticity, ζ , in a hydrostatic fluid as written using generalized vertical coordinates.

45.3.1 Basic manipulations

Our starting point is the vector-invariant velocity equation given by equation (45.20)

$$\left[\frac{\partial \mathbf{u}}{\partial t} \right]_\sigma + \dot{\sigma} \frac{\partial \mathbf{u}}{\partial \sigma} + \tilde{\zeta}_a \hat{z} \wedge \mathbf{u} = -\nabla_\sigma (K + \Phi) - (1/\rho) \nabla_\sigma p + \mathbf{F}^h, \quad (45.34)$$

where $\tilde{\zeta}_a = \tilde{\zeta} + f$ is the absolute vorticity. Taking the curl of this equation and projecting onto the vertical direction leads to

$$\frac{D\tilde{\zeta}_a}{Dt} = -\tilde{\zeta}_a \nabla_\sigma \cdot \mathbf{u} + \frac{\hat{z} \cdot (\nabla_\sigma \rho \wedge \nabla_\sigma p)}{\rho^2} + \hat{z} \cdot \left[\frac{\partial \mathbf{u}}{\partial \sigma} \wedge \nabla_\sigma \dot{\sigma} + \nabla_\sigma \wedge \mathbf{F}^h \right] \quad (45.35)$$

where we noted that the planetary vorticity, f , is independent of time and vertical position.

Making use of the mass conservation equation

Mass conservation in the form of equation (44.85)

$$\frac{1}{\rho h} \frac{D(\rho h)}{Dt} = -(\nabla_\sigma \cdot \mathbf{u} + \partial \dot{\sigma} / \partial \sigma), \quad (45.36)$$

renders

$$\rho h \frac{D}{Dt} \left[\frac{\tilde{\zeta}_a}{\rho h} \right] = \frac{\hat{\mathbf{z}} \cdot (\nabla_\sigma \rho \wedge \nabla_\sigma p)}{\rho^2} + \tilde{\zeta}_a \frac{\partial \dot{\sigma}}{\partial \sigma} + \hat{\mathbf{z}} \cdot \left[\frac{\partial \mathbf{u}}{\partial \sigma} \wedge \nabla_\sigma \dot{\sigma} + \nabla_\sigma \wedge \mathbf{F}^h \right]. \quad (45.37)$$

Massaging the $\dot{\sigma}$ terms

The terms containing $\dot{\sigma}$ can be written in the form

$$\tilde{\zeta}_a \partial_\sigma \dot{\sigma} + \hat{\mathbf{z}} \cdot (\partial_\sigma \mathbf{u} \wedge \nabla_\sigma \dot{\sigma}) = \tilde{\zeta}_a \partial_\sigma \dot{\sigma} + \hat{\mathbf{z}} \cdot [-\nabla_\sigma \wedge (\dot{\sigma} \partial_\sigma \mathbf{u}) + \dot{\sigma} \nabla_\sigma \wedge \partial_\sigma \mathbf{u}] \quad (45.38a)$$

$$= \tilde{\zeta}_a \partial_\sigma \dot{\sigma} + \dot{\sigma} \partial_\sigma \tilde{\zeta}_a - \hat{\mathbf{z}} \cdot [\nabla_\sigma \wedge (\dot{\sigma} \partial_\sigma \mathbf{u})] \quad (45.38b)$$

$$= \partial_\sigma (\dot{\sigma} \tilde{\zeta}_a) - \hat{\mathbf{z}} \cdot [\nabla_\sigma \wedge (\dot{\sigma} \partial_\sigma \mathbf{u})]. \quad (45.38c)$$

45.3.2 Vorticity and potential vorticity equation

The above results allow us to write equation (45.37) in the form

$$\rho h \frac{DQ}{Dt} = \frac{\hat{\mathbf{z}} \cdot (\nabla_\sigma \rho \wedge \nabla_\sigma p)}{\rho^2} + \partial_\sigma (\dot{\sigma} \tilde{\zeta}_a) + \nabla_\sigma \cdot [\hat{\mathbf{z}} \wedge \dot{\sigma} \partial_\sigma \mathbf{u} - \hat{\mathbf{z}} \wedge \mathbf{F}^h], \quad (45.39)$$

where we introduced the potential vorticity defined according to the generalized vertical coordinates

$$Q = \frac{\tilde{\zeta}_a}{h \rho}. \quad (45.40)$$

The potential vorticity equation (45.39) has a generally nonzero baroclinicity contribution (see Section 36.5 for more on baroclinicity)

$$\frac{\hat{\mathbf{z}} \cdot (\nabla_\sigma \rho \wedge \nabla_\sigma p)}{\rho^2} \quad (45.41)$$

so that the potential vorticity (45.40) is generally not materially invariant even for a perfect fluid.

Pressure coordinates

The baroclinicity (45.41) vanishes when choosing $\sigma = p$. We already noted this property in our general discussion of baroclinicity in Section 36.5. This choice, however, does not offer us a materially invariant potential vorticity since $\dot{\sigma} = \dot{p}$ does not generally vanish for a perfect fluid. Namely, a nonzero \dot{p} merely signals vertical motion, so that $\dot{p} \neq 0$ for both real and perfect fluids. Hence, even though the baroclinicity vanishes by choosing $\sigma = p$, the $\partial_\sigma (\dot{\sigma} \tilde{\zeta}_a)$ term does not.

Eulerian flux-form equation

Expanding the material time derivative in equation (45.39) allows us to cancel the $\partial_\sigma(\dot{\sigma}\tilde{\zeta}_a)$ term since it appears on both sides of the equation. We are thus led to the Eulerian flux-form potential vorticity equation written using generalized vertical coordinates

$$\left[\frac{\partial(\rho h Q)}{\partial t} \right]_\sigma = \frac{\hat{z} \cdot (\nabla_\sigma \rho \wedge \nabla_\sigma p)}{\rho^2} - \nabla_\sigma \cdot [\rho h \mathbf{u} Q + \hat{z} \wedge \dot{\sigma} \partial_\sigma \mathbf{u} + \hat{z} \wedge \mathbf{F}^h], \quad (45.42)$$

which, since $\tilde{\zeta}_a = h \rho Q$, is equivalent to the absolute vorticity equation

$$\left[\frac{\partial \tilde{\zeta}_a}{\partial t} \right]_\sigma = \frac{\hat{z} \cdot (\nabla_\sigma \rho \wedge \nabla_\sigma p)}{\rho^2} - \nabla_\sigma \cdot [\mathbf{u} \tilde{\zeta}_a + \hat{z} \wedge \dot{\sigma} \partial_\sigma \mathbf{u} + \hat{z} \wedge \mathbf{F}^h]. \quad (45.43)$$

As a check, note that setting $\sigma = z$ so that $h = 1$ reduces the vorticity equation (45.43) to the vertical component of the vorticity equation (36.30) (see Exercise 45.1).

45.3.3 Boussinesq fluid

Recall our discussion of vorticity for the Boussinesq fluid in Section 36.8, where we noted that the vertical component to the absolute vorticity is not affected by baroclinicity. We see this property in the present context by returning to the vector-invariant velocity equation (45.34) and setting the factor $1/\rho$ multiplying the pressure gradient to $1/\rho_0$ as part of the Boussinesq approximation

$$(1/\rho) \nabla_\sigma p \longrightarrow (1/\rho_0) \nabla_\sigma p, \quad (45.44)$$

in which ρ_0 is a constant. In this case the $\nabla_\sigma \wedge$ operation annihilates pressure and we are left with no vertical component to the baroclinicity. We are thus led to define the Boussinesq potential vorticity

$$Q = \frac{\tilde{\zeta}_a}{h} \quad (45.45)$$

which satisfies the material and Eulerian evolution equations

$$h \frac{DQ}{Dt} = \partial_\sigma(\dot{\sigma} \tilde{\zeta}_a) + \nabla_\sigma \cdot [\hat{z} \wedge \dot{\sigma} \partial_\sigma \mathbf{u} - \hat{z} \wedge \mathbf{F}^h] \quad (45.46)$$

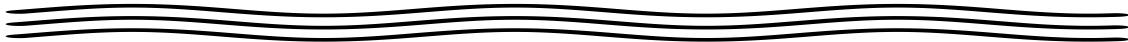
$$\left[\frac{\partial(h Q)}{\partial t} \right]_\sigma = -\nabla_\sigma \cdot [h \mathbf{u} Q + \hat{z} \wedge \dot{\sigma} \partial_\sigma \mathbf{u} + \hat{z} \wedge \mathbf{F}^h]. \quad (45.47)$$

If we choose a vertical coordinate that vanishes when there is no mixing, then this potential vorticity is materially invariant for a perfect fluid. The most common choice for this vertical coordinate is buoyancy or potential density, with the corresponding potential vorticity developed in Sections 46.2 and 46.3.

45.4 Exercise

EXERCISE 45.1: CHECKING THE VORTICITY EQUATION

Verify that for $2\Omega = f\hat{z}$ the choice $\sigma = z$ reduces the vorticity equation (45.43) to the vertical component of the vorticity equation (36.30).



Isopycnal primitive equations

For stably stratified fluids, buoyancy is a particularly useful generalized vertical coordinate. Most notably, physical processes away from turbulent boundary layers are oriented according to buoyancy surfaces, and horizontal buoyancy gradients give rise to thermal wind shears in a geostrophically balanced flow. For this reason buoyancy (or entropy) plays a key role in theoretical and numerical models of ocean and atmospheric circulation.

In this chapter we study the hydrostatic Boussinesq equations using buoyancy as the vertical coordinate. The resulting primitive equation set forms the basis for *isopycnal* models of the ocean or isentropic models of the atmosphere. We pay particular attention to the needs of vertically integrating the equations over discrete layers, as required to develop discrete numerical models. In the adiabatic limit, the isopycnal equations reduce to the stacked shallow water equations. After deriving the primitive equations using isopycnal coordinates, we derive the corresponding vorticity and potential vorticity equations. Throughout this chapter we expose details for the practitioner interested in the mathematical formalism for the purpose of analyzing ocean momentum, vorticity and potential vorticity budgets using isopycnal models.

READER'S GUIDE FOR THIS CHAPTER

We assume an understanding of the generalized vertical coordinate mathematics in Chapter 43, kinematics in Chapter 44, and dynamics in Chapter 45. Furthermore, we are concerned with details of vorticity and PV budgets in isopycnal coordinates, with the presentation building from our study of vorticity and potential vorticity in generalized vertical coordinates from Chapter 45.

46.1	Layered isopycnal primitive equations	1038
46.1.1	Montgomery potential and the pressure force	1038
46.1.2	Material time derivative	1040
46.1.3	Layer thickness and specific thickness	1040
46.1.4	Ocean equations	1041
46.1.5	Vector-invariant horizontal momentum equation	1042
46.1.6	Connection to the shallow water equations	1042
46.1.7	Diapycnal transfer	1043
46.1.8	Momentum transfer	1043
46.1.9	Allowing for layers to vanish and reappear	1043
46.2	Perfect fluid PV using isopycnal coordinates	1043
46.2.1	Derivation of the vorticity equation	1044
46.2.2	Derivation of the potential vorticity equation	1044
46.2.3	Coordinate transforming vorticity and potential vorticity	1045

46.3 Isopycnal coordinate PV with irreversible processes	1045
46.3.1 Derivation method I	1045
46.3.2 Derivation method II	1047
46.3.3 Comments	1048

46.1 Layered isopycnal primitive equations

Rather than specializing the generalized vertical coordinate equations provided in Section 45.1, we find it pedagogical to start from the equations written using the geopotential vertical coordinate (see Section 24.1.6)

$$\frac{D\mathbf{u}}{Dt} + f \hat{\mathbf{z}} \wedge \mathbf{u} = -\nabla_z \varphi + \mathbf{F} \quad \text{horizontal momentum} \quad (46.1a)$$

$$\frac{\partial \varphi}{\partial z} = b \quad \text{hydrostatic} \quad (46.1b)$$

$$\nabla_z \cdot \mathbf{u} + \frac{\partial w}{\partial z} = 0 \quad \text{continuity} \quad (46.1c)$$

$$\frac{Db}{Dt} = \dot{b} \quad \text{thermodynamics} \quad (46.1d)$$

$$\frac{DC}{Dt} = \dot{C} \quad \text{tracers,} \quad (46.1e)$$

where $\mathbf{v} = (\mathbf{u}, w)$ is the velocity field, \mathbf{u} is its horizontal component, φ is the dynamic pressure, b is the buoyancy, and C is an arbitrary tracer concentration. A discrete realization of the isopycnal layer-integrated form of these equations is depicted in Figure 46.1, with the remainder of this section detailing the formulation using isopycnal coordinates for the vertical.

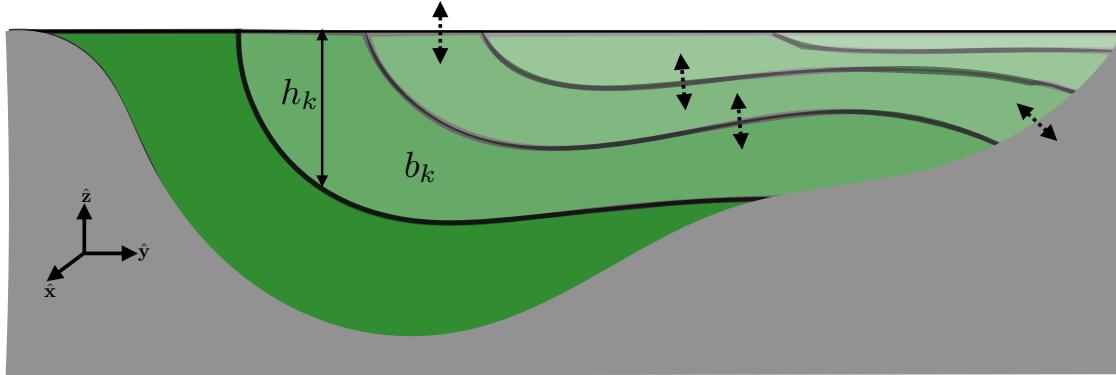


FIGURE 46.1: Schematic of an isopycnal model, formulated as stacked shallow water layers (green layers) that generally allow for the transfer of matter and energy across the layer interfaces as well as across the ocean surface and ocean bottom (as depicted by the double-headed dashed arrows). The dark gray region is land. Discrete layer thicknesses are denoted h_k with corresponding layer buoyancy, b_k .

46.1.1 Montgomery potential and the pressure force

We here consider the horizontal pressure force appearing in isopycnal models, in which we uncover the importance of the Montgomery potential.

Horizontal pressure gradient force

Throughout this chapter we make use of the horizontal derivatives on constant buoyancy surfaces (derived in Section 43.11), written here in the form

$$\nabla_b = \hat{\mathbf{x}} \left[\frac{\partial}{\partial x} \right]_b + \hat{\mathbf{y}} \left[\frac{\partial}{\partial y} \right]_b. \quad (46.2)$$

Following the discussion in Section 45.1.3, the horizontal pressure thus gradient transforms as

$$\nabla_z \varphi = \nabla_b \varphi - \frac{\partial \varphi}{\partial z} \nabla_b z \quad (46.3a)$$

$$= \nabla_b \varphi - b \nabla_b z \quad (46.3b)$$

$$= \nabla_b (\varphi - b z) \quad (46.3c)$$

$$= \nabla_b M, \quad (46.3d)$$

where

$$M = \varphi - b z \quad (46.4)$$

defines the Montgomery potential. As the contribution to the horizontal pressure force, the Montgomery potential is the geostrophic streamfunction in buoyancy coordinates (see Section 46.1.4).

The horizontal pressure gradient force for numerical models

It is notable that the horizontal pressure gradient force is determined by the horizontal isopycnal gradient of a single term, the Montgomery potential. Furthermore, as shown below, the Montgomery potential satisfies the buoyancy coordinate form of the hydrostatic balance. Hence, numerical isopycnal models do not suffer from problems with computing the horizontal pressure gradient that can occur with other generalized vertical coordinate models, such as terrain-following models (see Figure 45.1).

Equation (46.3c) is the key step in the formulation, whereby we made use of $\nabla_b b = 0$. This step is available only under certain cases that utilize an idealized equation of state for seawater. In more realistic cases, the buoyancy determining the hydrostatic pressure (i.e., the *mass buoyancy*) is defined locally whereas the generalized vertical coordinate must be defined globally. As a result, there are two terms contributing to the pressure gradient in a manner similar to terrain-following models. [Sun et al. \(1999\)](#) and [Hallberg \(2005\)](#) discuss this issue in the context of numerical ocean modeling. For present purposes we ignore this detail and continue to assume a simplified equation of state so that $\nabla_b b = 0$.

Hydrostatic balance

Supporting our use of the Montgomery potential as a pressure field, the hydrostatic balance takes the form

$$\frac{\partial M}{\partial b} = \frac{\partial \varphi}{\partial b} - b \frac{\partial z}{\partial b} - z = \frac{\partial \varphi}{\partial z} \frac{\partial z}{\partial b} - b \frac{\partial z}{\partial b} - z = -z, \quad (46.5)$$

where we made use of the hydrostatic balance $\partial \varphi / \partial z = b$ (equation (46.1b)). This result means that M is the buoyancy coordinate version of pressure.

46.1.2 Material time derivative

As seen in Section 44.5, there are two equivalent forms for the material time derivative

$$\frac{D}{Dt} = \left[\frac{\partial}{\partial t} \right]_z + \mathbf{u} \cdot \nabla_z + w \frac{\partial}{\partial z} \quad \text{geopotential form} \quad (46.6a)$$

$$= \left[\frac{\partial}{\partial t} \right]_b + \mathbf{u} \cdot \nabla_b + w^{(b)} \frac{\partial}{\partial z} \quad \text{isopycnal form,} \quad (46.6b)$$

where

$$w^{(b)} = \frac{\partial z}{\partial b} \frac{Db}{Dt} \quad (46.7)$$

is the diapycnal velocity component that measures the rate of flow crossing buoyancy surfaces (Section 44.4). Besides differences in the spatial operators, it is important to note that the time derivative operators in equations (46.6a) and (46.6b) are computed on constant geopotential and constant buoyancy surfaces, respectively. However, the horizontal velocity component is the *same* for both forms of the material time derivative

$$\mathbf{u} = \hat{\mathbf{x}} u + \hat{\mathbf{y}} v = (D/Dt) (\hat{\mathbf{x}} x + \hat{\mathbf{y}} y). \quad (46.8)$$

46.1.3 Layer thickness and specific thickness

The continuity equation, $\nabla_z \cdot \mathbf{u} + \partial_z w = 0$, is an expression of volume conservation. We already derived the generalized vertical coordinate version of this equation in Section 44.10.3, and thus quote the isopycnal layer thickness result here

$$\left[\frac{\partial h}{\partial t} \right]_b + \nabla_b \cdot (h \mathbf{u}) + \Delta_b w^{(b)} = 0. \quad (46.9)$$

The field h measures the isopycnal layer thickness (dimensions of length) and is given by the vertical integral over a layer

$$h = \int_{z(b-\delta b/2)}^{z(b+\delta b/2)} dz = \int_{b-\delta b/2}^{b+\delta b/2} \frac{\partial z}{\partial b} db = \int_{b-\delta b/2}^{b+\delta b/2} h db = \int_{b-\delta b/2}^{b+\delta b/2} N^{-2} db = \bar{h} db. \quad (46.10)$$

The specific thickness, κ , equals to the inverse squared buoyancy frequency

$$\kappa = \frac{\partial z}{\partial b} = N^{-2}, \quad (46.11)$$

and its layer averaged value is $\bar{\kappa} = h/\delta b$. Furthermore, the dia-surface transport velocity is given by

$$w^{(b)} = \kappa \dot{b}. \quad (46.12)$$

Its difference across layer interfaces

$$\Delta_b w^{(b)} = \int_{b-\delta b/2}^{b+\delta b/2} \frac{\partial w^{(b)}}{\partial b} db = w^{(b)}(b + \delta b/2) - w^{(b)}(b - \delta b/2) \quad (46.13)$$

measures the amount of fluid that diverges from the layer through cross-layer transport.

Specific thickness equation

Inserting $h = \bar{h} \delta b$ into the thickness equation (46.9) leads to

$$\left[\frac{\partial \bar{h}}{\partial t} \right]_b + \nabla_b \cdot (\bar{h} \mathbf{u}) + \partial_b w^{(b)} = 0, \quad (46.14)$$

where we pulled the buoyancy increment, δb , outside of the time and horizontal derivative operators since δb is a fixed number for a chosen layer. We also used the identity relating the difference operator to a differential operator

$$\delta b \partial_b = \Delta_b. \quad (46.15)$$

For a vertically continuous treatment, equation (46.14) can be written with h rather than the discrete layer averaged value

$$\left[\frac{\partial h}{\partial t} \right]_b + \nabla_b \cdot (h \mathbf{u}) + \partial_b w^{(b)} = 0. \quad (46.16)$$

It is generally more convenient to use the specific thickness when working with the vertically continuous equations, whereas the finite layer thickness, h , is more suitable for the layer integrated equations.

Adiabatic limit

When $w^{(b)} \neq 0$, the three terms in the thickness equation (46.9) or the specific thickness equation (46.14) are coupled. We discussed this coupling in Section 44.6 as part of our broader study of the vertical velocity and the dia-surface velocity. When considering perfect fluids, we can set $w^{(b)} = h \dot{b} = 0$ since the fluid has no mixing. In this case the layer thickness is altered only through horizontal rearrangements of volume within a layer according to the adiabatic thickness equation

$$\left[\frac{\partial h}{\partial t} \right]_b + \nabla_b \cdot (h \mathbf{u}) = 0. \quad (46.17)$$

As further discussed in Section 46.1.6, the adiabatic limit brings the discrete isopycnal model into accord with the immiscible stacked shallow water models discussed in Part V of this book.

46.1.4 Ocean equations

Bringing the pieces together leads to the isopycnal version of the hydrostatic Boussinesq equations

$$\left[\frac{\partial \mathbf{u}}{\partial t} \right]_b + (\mathbf{u} \cdot \nabla_b) \mathbf{u} + (w^{(b)} \partial_z) \mathbf{u} + f \hat{z} \wedge \mathbf{u} = -\nabla_b M + \mathbf{F}^h \quad (46.18a)$$

$$\frac{\partial M}{\partial b} = -z \quad (46.18b)$$

$$\left[\frac{\partial h}{\partial t} \right]_b + \nabla_b \cdot (h \mathbf{u}) + \Delta_b w^{(b)} = 0 \quad (46.18c)$$

$$\left[\frac{\partial (h C)}{\partial t} \right]_b + \nabla_b \cdot (h C \mathbf{u} + h \mathbf{J}^h) + \Delta_b (C w^{(b)} + J^{(b)}) = 0, \quad (46.18d)$$

where the tracer equation includes possible subgrid scale flux contributions as well as advective transport. Notice how the advective transport is two-dimensional in the adiabatic case with $\dot{b} = 0$,

in which case layer-integrated scalar properties, such as volume and tracer content, are constant within buoyancy layers. Also note that geostrophic balance in the horizontal momentum equation (46.18a) gives

$$f \hat{z} \wedge \mathbf{u}_g = -\nabla_b M \implies f u_g = - \left[\frac{\partial M}{\partial y} \right]_b \quad \text{and} \quad f v_g = \left[\frac{\partial M}{\partial x} \right]_b. \quad (46.19)$$

Hence, the Montgomery potential is the streamfunction for geostrophic flow as represented using buoyancy coordinates.

46.1.5 Vector-invariant horizontal momentum equation

It is common for isopycnal models to make use of the vector-invariant form of the momentum equation derived in Section 45.1.5. Introducing the isopycnal version of the relative vorticity

$$\hat{z} \tilde{\zeta} \equiv \nabla_b \wedge \mathbf{u} = \left[\frac{\partial v}{\partial x} \right]_b - \left[\frac{\partial u}{\partial y} \right]_b \quad (46.20)$$

renders the vector-invariant horizontal momentum equation

$$\left[\frac{\partial \mathbf{u}}{\partial t} \right]_b + w^{(b)} \partial_z \mathbf{u} + \tilde{\zeta}_a \hat{z} \wedge \mathbf{u} = -\nabla_b \mathcal{B} + \mathbf{F}^h, \quad (46.21)$$

where

$$\mathcal{B} = M + \mathbf{u} \cdot \mathbf{u}/2 = \varphi - bz + \mathbf{u} \cdot \mathbf{u}/2 \quad (46.22)$$

is the Bernoulli potential for a hydrostatic Boussinesq fluid (see Section 22.6.2), and

$$\tilde{\zeta}_a = \tilde{\zeta} + f \quad (46.23)$$

is the vertical component to the absolute vorticity using isopycnal coordinates. Note that we can further introduce the isopycnal potential vorticity (Section 46.2.2)

$$h Q = \tilde{\zeta}_a \quad (46.24)$$

to bring the momentum equation to the form

$$\left[\frac{\partial \mathbf{u}}{\partial t} \right]_b + w^{(b)} \partial_z \mathbf{u} + Q \hat{z} \wedge (h \mathbf{u}) = -\nabla_b \mathcal{B} + \mathbf{F}^h. \quad (46.25)$$

This form is commonly used as the starting point for certain theoretical analyses, particularly when considering the adiabatic limit in which $w^{(b)} = 0$.

46.1.6 Connection to the shallow water equations

We can make use of the material time derivative operator (46.6b) to write the material form of the adiabatic and inviscid equations (46.18a)-(46.18c)

$$\frac{D\mathbf{u}}{Dt} + f \hat{z} \wedge \mathbf{u} = -\nabla_b M \quad (46.26a)$$

$$\frac{\partial M}{\partial b} = -z \quad (46.26b)$$

$$\frac{Dh}{Dt} + h \nabla_b \cdot \mathbf{u} = 0. \quad (46.26c)$$

These isopycnal momentum and thickness equations are isomorphic to those for a single layer of adiabatic shallow water fluid (see Section 31.2). This isomorphism allows us to derive the vorticity and potential vorticity equations in Section 46.2, making use of the shallow water manipulations from Section 35.4.

46.1.7 Diapycnal transfer

At ocean boundaries, the diapycnal term, $w^{(b)}$, accounts for the transfer of matter across the ocean boundaries via precipitation, evaporation, ice melt/form, and river runoff. Notably, this matter transfer also generally gives rise to a transfer of trace matter (tracers), heat (evaporation and precipitation carry a heat content), and momentum (precipitation generally has nonzero momentum). In the ocean interior, $w^{(b)}$ affects the transfer of volume, tracer, and momentum between layers as induced by irreversible processes such as mixing.

46.1.8 Momentum transfer

Pressure form stress mechanically couples isopycnal layers even in the absence of diapycnal matter transfer. We discussed the physics of form stress for the shallow water system in Section 32.4 and more generally in Section 20.1. Furthermore, there are a suite of unresolved processes giving rise to lateral and vertical stresses. Typical ocean model treatments incorporate a turbulent friction in the ocean interior, with lateral stresses acting within a layer and diapycnal stresses acting across isopycnal layer interfaces. A bottom drag is typically applied at the ocean bottom and a turbulent stress applied at the ocean surface. Details for the boundary stresses involve the physics of boundary layer turbulence, which is a topic outside of our scope.

46.1.9 Allowing for layers to vanish and reappear

Isopycnal layers have a transient existence at any particular horizontal position since a layer can incrop at the ocean bottom and outcrop at the ocean surface (see Figure 46.1). The seasonal cycle of warming and cooling is a canonical example of layer outcropping at the surface ocean. A formulation expedient to handle vanishing layers is to assume that all layers exist everywhere horizontally across the ocean domain, but to allow for zero layer thickness where a layer has zero volume. We made use of this approach when discussing available potential energy in Section 24.8. To admit this feature in a discrete model requires a careful realization of L'Hôpital's rule of differential calculus, thus ensuring the discrete model conserves properties in the presence of layers that can appear and disappear at any particular point in the domain.

46.2 Perfect fluid PV using isopycnal coordinates

In Section 38.5, we showed that the absolute vorticity in a Boussinesq hydrostatic fluid with a simplified seawater equation of state (Section 39.2), when projected into the direction normal to constant buoyancy surface, $\omega_a \cdot \nabla b$, is not affected by baroclinicity; i.e., that projection annihilates the baroclinicity vector. From that property we conclude that $\omega_a \cdot \nabla b$ is the potential vorticity for the Boussinesq hydrostatic fluid.

For a Boussinesq hydrostatic fluid, isopycnal coordinates build in the above feature of buoyancy surfaces. Indeed, as shown in Section 45.3.3, the vertical component to baroclinicity vanishes for any generalized vertical coordinate representation of a Boussinesq fluid. Hence, buoyancy coordinates are not special from this perspective. Instead, they are special since in the case of a perfect fluid, the buoyancy based potential vorticity is materially invariant. Deriving this result is one purpose of this section. Note that in Section 45.3, we derived the potential vorticity equation for a hydrostatic fluid represented with generalized vertical coordinates. We could merely specialize that result to the current case. However, we prefer to here work through the maths to help further our experience

performing certain of the key manipulations arising with vorticity in rotating and stratified fluids. Hence, consider this section, as well as Section 46.3, to be extended worked exercises.

46.2.1 Derivation of the vorticity equation

Acting with the vertical projection of the curl, $\hat{z} \cdot (\nabla_b \wedge)$, onto the adiabatic and inviscid form of the vector-invariant velocity equation (46.21) leads to the isopycnal vorticity equation

$$\left[\frac{\partial \tilde{\zeta}_a}{\partial t} \right]_b + (\mathbf{u} \cdot \nabla_b) \tilde{\zeta}_a = -\tilde{\zeta}_a \nabla_b \cdot \mathbf{u} \quad (46.27)$$

where

$$\tilde{\zeta}_a = f + \hat{z} \cdot (\nabla_b \wedge \mathbf{u}) = f + \tilde{\zeta} \quad (46.28)$$

is the absolute vorticity, written as the planetary vorticity plus the isopycnal relative vorticity. The left hand side of equation (46.27) is the material time derivative of absolute vorticity (see equation (46.6b)), so that we can write

$$\frac{D\tilde{\zeta}_a}{Dt} = -\tilde{\zeta}_a \nabla_b \cdot \mathbf{u}. \quad (46.29)$$

As advertised above, there is no baroclinicity vector on the right hand side of this vorticity equation. Again, that property results from our choice to use isopycnal coordinates.

46.2.2 Derivation of the potential vorticity equation

We now make use of the thickness equation derived in Section 46.1.3, here realized in its material form to eliminate the convergence $\nabla_b \cdot \mathbf{u}$ on the right hand side of equation (46.29), thus leading to

$$\frac{D\tilde{\zeta}_a}{Dt} - \frac{\tilde{\zeta}_a}{h} \frac{Dh}{Dt} = 0. \quad (46.30)$$

Introducing the isopycnal potential vorticity

$$Q = \frac{\tilde{\zeta}_a}{h} = \frac{f + \tilde{\zeta}}{h} \quad (46.31)$$

leads to

$$\frac{DQ}{Dt} = 0. \quad (46.32)$$

Expanding the material time derivative into its components according to equation (46.6b), and making use of the adiabatic form of the thickness equation leads to the Eulerian flux-form equation

$$\left[\frac{\partial (h Q)}{\partial t} \right]_b + \nabla_b \cdot (h Q \mathbf{u}) = 0. \quad (46.33)$$

As noted in Section 46.1.3, when formulating the vertically continuous equations rather than finite thickness layered equations, it is more convenient to make use of the specific thickness, \hbar , rather than the layer thickness, h . In this case we are motivated to define the potential vorticity as

$$Q = \frac{\tilde{\zeta}_a}{\hbar} = \frac{f + \tilde{\zeta}}{\hbar}. \quad (46.34)$$

The corresponding PV equation is identical to equation (46.33), only now with h replaced by \hbar .

46.2.3 Coordinate transforming vorticity and potential vorticity

As just shown, PV for a hydrostatic Boussinesq fluid can be written in the relatively simple form of a shallow water PV when choosing isopycnal coordinates. Here is a direct transformation from Cartesian to isopycnal coordinates that also reveals this form

$$(\boldsymbol{\omega}^{\text{hy}} + f \hat{\mathbf{z}}) \cdot \nabla b = -\frac{\partial v}{\partial z} \frac{\partial b}{\partial x} + \frac{\partial u}{\partial z} \frac{\partial b}{\partial y} + \left(\frac{\partial v}{\partial x} - \frac{\partial u}{\partial y} + f \right) \frac{\partial b}{\partial z} \quad (46.35a)$$

$$= \frac{\partial b}{\partial z} \left[f + \left(\frac{\partial v}{\partial x} - \frac{\partial v}{\partial z} \frac{\partial b}{\partial x} \right) - \left(\frac{\partial u}{\partial y} - \frac{\partial u}{\partial z} \frac{\partial b}{\partial y} \right) \right] \quad (46.35b)$$

$$= \frac{\partial b}{\partial z} \left[f + \left(\frac{\partial v}{\partial x} \right)_b - \left(\frac{\partial u}{\partial y} \right)_b \right] \quad (46.35c)$$

$$= \frac{f + (\partial v / \partial x)_b - (\partial u / \partial y)_b}{\partial z / \partial b} \quad (46.35d)$$

$$= \frac{f + \tilde{\zeta}}{h} \quad (46.35e)$$

$$= Q. \quad (46.35f)$$

46.3 Isopycnal coordinate PV with irreversible processes

In Section 46.2, we considered the PV equation for an adiabatic, inviscid, hydrostatic, Boussinesq fluid using isopycnal vertical coordinates. We here extend to the case of friction in the momentum equation and diabatic heating in the buoyancy equation. We consider two ways to derive the governing equations. One is to convert the non-hydrostatic PV equation in Exercise 38.2 to isopycnal coordinates, after making the hydrostatic approximation. The second is to start from the equations of motion in isopycnal coordinates and derive the vorticity equation and then the PV equation. We make use of the vertically continuous equations, thus warranting our use of specific thickness, h , rather than layer thickness, h (see Section 46.1.3).

Note that much of this section is a specialization of the more general considerations of Section 45.3. Nonetheless, we here revisit some of the earlier derivations as a means to bolster our mathematical manipulation muscle.

46.3.1 Derivation method I

As derived earlier in this chapter, the equations of motion with diabatic heating and friction, written using isopycnal (or buoyancy) vertical coordinates, take the form

$$\left[\frac{\partial \mathbf{u}}{\partial t} \right]_b + (\mathbf{u} \cdot \nabla_b) \mathbf{u} + \dot{b} \frac{\partial \mathbf{u}}{\partial b} + \mathbf{f} \wedge \mathbf{u} = -\nabla_b M + \mathbf{F} \quad (46.36a)$$

$$\frac{\partial M}{\partial b} = -z \quad (46.36b)$$

$$\left[\frac{\partial h}{\partial t} \right]_b + \nabla_b \cdot (h \mathbf{u}) = -\frac{\partial (h \dot{b})}{\partial b} \quad (46.36c)$$

$$\frac{D b}{D t} = \dot{b}. \quad (46.36d)$$

Note that in this section choose to write the dia-surface transport operator in the form

$$w^{(b)} \partial_z = \dot{b} \partial_b. \quad (46.37)$$

We can make use of the material time derivative operator (46.6b) to write the material form of the equations

$$\frac{D\mathbf{u}}{Dt} + \mathbf{f} \wedge \mathbf{u} = -\nabla_b M + \mathbf{F} \quad (46.38a)$$

$$\frac{\partial M}{\partial b} = -z \quad (46.38b)$$

$$\frac{Dh}{Dt} + h \nabla_b \cdot \mathbf{u} = -h \frac{\partial \dot{b}}{\partial b}. \quad (46.38c)$$

Curl of the velocity equation

We start taking the curl, $\nabla_b \wedge$, of the velocity equation (46.36a), thus leading to the isopycnal vorticity equation

$$\left[\frac{\partial \tilde{\zeta}_a}{\partial t} \right]_b + (\mathbf{u} \cdot \nabla_b) \tilde{\zeta}_a + \dot{b} \left[\frac{\partial \tilde{\zeta}_a}{\partial b} \right] = -\tilde{\zeta}_a \nabla_b \cdot \mathbf{u} + \hat{z} \cdot \left[\frac{\partial \mathbf{u}}{\partial b} \wedge \nabla_b \dot{b} + \nabla_b \wedge \mathbf{F} \right]. \quad (46.39)$$

The left hand side of equation (46.39) is the material time derivative of absolute vorticity (see equation (46.6b)), so that

$$\frac{D\tilde{\zeta}_a}{Dt} = -\tilde{\zeta}_a \nabla_b \cdot \mathbf{u} + \hat{z} \cdot \left[\frac{\partial \mathbf{u}}{\partial b} \wedge \nabla_b \dot{b} + \nabla_b \wedge \mathbf{F} \right]. \quad (46.40)$$

Now make use of the thickness equation in the material form (46.38c) to eliminate the convergence $\nabla_b \cdot \mathbf{u}$ on the right hand side, thus leading to

$$\frac{D\tilde{\zeta}_a}{Dt} - \frac{\tilde{\zeta}_a}{h} \left[\frac{Dh}{Dt} - h \frac{\partial \dot{b}}{\partial b} \right] = \hat{z} \cdot \left[\frac{\partial \mathbf{u}}{\partial b} \wedge \nabla_b \dot{b} + \nabla_b \wedge \mathbf{F} \right]. \quad (46.41)$$

Introducing the isopycnal potential vorticity

$$Q = \frac{\tilde{\zeta}_a}{h} = \frac{\tilde{\zeta} + f}{h} \quad (46.42)$$

leads to

$$h \frac{DQ}{Dt} = \zeta_a \frac{\partial \dot{b}}{\partial b} + \hat{z} \cdot \left[\frac{\partial \mathbf{u}}{\partial b} \wedge \nabla_b \dot{b} + \nabla_b \wedge \mathbf{F} \right]. \quad (46.43)$$

Massaging the diabatic terms

The diabatic terms can be written

$$\zeta_a \frac{\partial \dot{b}}{\partial b} + \hat{z} \cdot \left[\frac{\partial \mathbf{u}}{\partial b} \wedge \nabla_b \dot{b} \right] = \zeta_a \frac{\partial \dot{b}}{\partial b} + \dot{b} \frac{\partial \tilde{\zeta}}{\partial b} - \hat{z} \cdot \left[\nabla_b \wedge \dot{b} \frac{\partial \mathbf{u}}{\partial b} \right] \quad (46.44a)$$

$$= \zeta_a \frac{\partial \dot{b}}{\partial b} + \dot{b} \frac{\partial \tilde{\zeta}_a}{\partial b} - \hat{z} \cdot \left[\nabla_b \wedge \dot{b} \frac{\partial \mathbf{u}}{\partial b} \right] \quad (46.44b)$$

$$= \frac{\partial(\tilde{\zeta}_a \dot{b})}{\partial b} - \hat{z} \cdot \left[\nabla_b \wedge \dot{b} \frac{\partial \mathbf{u}}{\partial b} \right] \quad (46.44c)$$

$$= \frac{\partial(\tilde{\zeta}_a \dot{b})}{\partial b} + \nabla_b \cdot \left[\hat{z} \wedge \dot{b} \frac{\partial \mathbf{u}}{\partial b} \right], \quad (46.44d)$$

where the second equality follows since the Coriolis parameter is independent of the buoyancy.

The PV equation

The PV equation takes the material form

$$\hbar \left[\frac{DQ}{Dt} \right] = \frac{\partial(\tilde{\zeta}_a \dot{b})}{\partial b} + \nabla_b \cdot \left[\hat{z} \wedge \dot{b} \frac{\partial \mathbf{u}}{\partial b} - \hat{z} \wedge \mathbf{F} \right]. \quad (46.45)$$

Expanding the material time derivative into its components (46.6b), and making use of the thickness equation (46.36c), leads to the Eulerian flux-form equation

$$\left[\frac{\partial(\hbar Q)}{\partial t} \right]_b + \nabla_b \cdot (\hbar Q \mathbf{u}) + \frac{\partial(\hbar Q \dot{b})}{\partial b} = \frac{\partial(\tilde{\zeta}_a \dot{b})}{\partial b} + \nabla_b \cdot \left[\hat{z} \wedge \dot{b} \frac{\partial \mathbf{u}}{\partial b} - \hat{z} \wedge \mathbf{F} \right]. \quad (46.46)$$

Since $\hbar Q = \tilde{\zeta}_a$, the ∂_b terms cancel, thus leaving the flux-form PV equation

$$\left[\frac{\partial(\hbar Q)}{\partial t} \right]_b = -\nabla_b \cdot \left[\hbar Q \mathbf{u} - \hat{z} \wedge \dot{b} \frac{\partial \mathbf{u}}{\partial b} + \hat{z} \wedge \mathbf{F} \right]. \quad (46.47)$$

46.3.2 Derivation method II

The alternative method to derive the PV equation in isopycnal coordinates is to start from the hydrostatic Boussinesq PV equation in geopotential vertical coordinates, and directly transform to isopycnal coordinates. For this purpose we start from the discussion in Section 38.5.2 to write the material evolution of PV for a hydrostatic and Boussinesq fluid

$$\frac{DQ}{Dt} = \nabla \cdot \left[(f \hat{z} + \boldsymbol{\omega}_{hy}) \dot{b} + b \nabla \wedge \mathbf{F} \right], \quad (46.48)$$

where

$$Q = \boldsymbol{\omega}_a^{hy} \cdot \nabla b = \boldsymbol{\omega}^{hy} \cdot \nabla b + f \frac{\partial b}{\partial z} \quad \text{and} \quad \boldsymbol{\omega}_{hy} = -\hat{x} \frac{\partial v}{\partial z} + \hat{y} \frac{\partial u}{\partial z} + \hat{z} \left[\frac{\partial v}{\partial x} - \frac{\partial u}{\partial y} \right]. \quad (46.49)$$

The simplest term in equation (46.48) to transform to isopycnal coordinates is the curl of the horizontal friction vector

$$\hat{z} \cdot (\nabla \wedge \mathbf{F}) = \hat{z} \cdot \nabla \wedge (F^x, F^y, 0), \quad (46.50)$$

which takes the form

$$\hat{z} \cdot (\nabla \wedge \mathbf{F}) = \hbar^{-1} \hat{z} \cdot (\nabla_b \wedge \mathbf{F}) = -\hbar^{-1} \nabla_b \cdot (\hat{z} \wedge \mathbf{F}). \quad (46.51)$$

The diabatic term requires some more work. Since the vorticity has zero divergence, the diabatic term can be written as

$$\nabla \cdot \left[(f \hat{z} + \boldsymbol{\omega}_{hy}) \dot{b} \right] = (f \hat{z} + \boldsymbol{\omega}_{hy}) \cdot \nabla \dot{b} \quad (46.52a)$$

$$= f \frac{\partial \dot{b}}{\partial z} - \frac{\partial v}{\partial z} \frac{\partial \dot{b}}{\partial x} + \frac{\partial u}{\partial z} \frac{\partial \dot{b}}{\partial y} + \left[\frac{\partial v}{\partial x} - \frac{\partial u}{\partial y} \right] \frac{\partial \dot{b}}{\partial z}. \quad (46.52b)$$

We now introduce horizontal derivatives on isopycnal surfaces according to the following relation (see Section 43.11.2)

$$\nabla_z = \nabla_b + \nabla_z b \left[\frac{\partial z}{\partial b} \right] \frac{\partial}{\partial z} \quad (46.53)$$

Doing so leads to

$$\begin{aligned}
\nabla \cdot [(f \hat{z} + \boldsymbol{\omega}_{hy}) \dot{b}] &= f \frac{\partial \dot{b}}{\partial z} - \frac{\partial v}{\partial z} \frac{\partial \dot{b}}{\partial x} + \frac{\partial u}{\partial z} \frac{\partial \dot{b}}{\partial y} + \left[\frac{\partial v}{\partial x} - \frac{\partial u}{\partial y} \right] \frac{\partial \dot{b}}{\partial z} \\
&= f \frac{\partial \dot{b}}{\partial z} - \frac{\partial v}{\partial z} \left(\left[\frac{\partial \dot{b}}{\partial x} \right]_b + \left[\frac{\partial b}{\partial x} \right]_z \frac{\partial z}{\partial b} \frac{\partial \dot{b}}{\partial z} \right) + \frac{\partial u}{\partial z} \left(\left[\frac{\partial \dot{b}}{\partial y} \right]_b + \left[\frac{\partial b}{\partial y} \right]_z \frac{\partial z}{\partial b} \frac{\partial \dot{b}}{\partial z} \right) \\
&\quad + \frac{\partial \dot{b}}{\partial z} \left(\left[\frac{\partial v}{\partial x} \right]_b + \left[\frac{\partial b}{\partial x} \right]_z \frac{\partial z}{\partial b} \frac{\partial v}{\partial z} \right) - \frac{\partial \dot{b}}{\partial z} \left(\left[\frac{\partial u}{\partial y} \right]_b + \left[\frac{\partial b}{\partial y} \right]_z \frac{\partial z}{\partial b} \frac{\partial u}{\partial z} \right) \\
&= f \frac{\partial \dot{b}}{\partial z} - \frac{\partial v}{\partial z} \left[\frac{\partial \dot{b}}{\partial x} \right]_b + \frac{\partial u}{\partial z} \left[\frac{\partial \dot{b}}{\partial y} \right]_b + \tilde{\zeta} \left[\frac{\partial \dot{b}}{\partial z} \right] \\
&= \frac{\partial b}{\partial z} \left(\tilde{\zeta}_a \left[\frac{\partial \dot{b}}{\partial b} \right] - \frac{\partial v}{\partial b} \left[\frac{\partial \dot{b}}{\partial x} \right]_b + \frac{\partial u}{\partial b} \left[\frac{\partial \dot{b}}{\partial y} \right]_b \right) \\
&= h^{-1} \left(\tilde{\zeta}_a \left[\frac{\partial \dot{b}}{\partial b} \right] + \hat{z} \cdot \left[\frac{\partial \mathbf{u}}{\partial b} \wedge \nabla_b \dot{b} \right] \right) \\
&= h^{-1} \left(\tilde{\zeta}_a \left[\frac{\partial \dot{b}}{\partial b} \right] + \dot{b} \left[\frac{\partial \tilde{\zeta}_a}{\partial b} \right] + \nabla_b \cdot \left[\hat{z} \wedge \dot{b} \frac{\partial \mathbf{u}}{\partial b} \right] \right) \\
&= h^{-1} \left(\frac{\partial(\dot{b} \tilde{\zeta}_a)}{\partial b} + \nabla_b \cdot \left[\hat{z} \wedge \dot{b} \frac{\partial \mathbf{u}}{\partial b} \right] \right). \tag{46.54}
\end{aligned}$$

To reach the penultimate step we noted that $\partial f / \partial b = 0$ so that we could form the derivative of the absolute vorticity. Bringing the pieces together leads to the PV equation (46.45) derived starting from the isopycnal version of the equations of motion.

46.3.3 Comments

The flux-form PV equation (46.47) manifests the impermeability theorem of Chapter 39 since the right hand side is the isopycnal convergence of a flux.



Shallow water thickness weighted averaging

There is a range of mathematical formalisms of use to frame the study of how linear waves, nonlinear waves, eddies, and fully developed turbulence interact with a mean flow. A distinctly geophysical element enters these studies through the primary role of vertical stratification arising from gravitation, with stratification particularly important for large scales flows where motions are approximately hydrostatic. A further specialization to the oceanographic context arises since there are few regions of the ocean where zonal averages apply, which contrasts to the atmospheric case.¹ The *thickness weighted averaging* (TWA) method has emerged as a particularly useful formalism for stratified flows, with particular use for studies of geostrophic eddies and their parameterization. In this chapter, we develop the TWA equations for the adiabatic stacked shallow water model.

As noted in the introduction to Part V of this book, the adiabatic stacked shallow water model exposes key facets of stratified geophysical flows without requiring the mathematics of generalized vertical coordinates developed in Chapters 43, 44, 45, and 46. The core simplification arises by assuming that horizontal fluid motion has no vertical dependence within a shallow water layer, which then means that vertical motion is a linear function of vertical position within each layer. That is, the shallow water fluid moves as extensible vertical columns (Section 31.2.7). It follows that horizontal pressure gradients do not need to be projected along the slope of the layer since they are vertically constant within a layer. In contrast, this projection is needed for a continuously stratified fluid described by generalized vertical coordinates, as illustrated in Figure 43.4. Hence, the shallow water equations for momentum, thickness, and tracers retain their use of Cartesian coordinates even though the layer interfaces undulate and are thus not generally horizontal. This mathematical feature of shallow water fluids aids in our pedagogical development of the TWA method.

[Young \(2012\)](#) offers an elegant application of the TWA to the continuously stratified Boussinesq hydrostatic fluid, with this paper the culmination of many years of prior work (see [Young \(2012\)](#) for citations). Penetrating the TWA approach for continuously stratified fluids requires an understanding of generalized vertical coordinates and the attendant tensor analysis such as that developed in Chapters 43, 44, 45, and 46. In the present chapter, we focus on the TWA equations for the adiabatic shallow water model. Doing so minimizes the mathematical requirements while exposing the key physical concepts. It also offers a useful baseline for those using stacked shallow water models for studies of adiabatic waves and geostrophic turbulence. Digesting the material in this chapter, and then coupling to skills in generalized vertical coordinate tensor analysis, prepares one for penetrating [Young \(2012\)](#). Afterward, the mathematically inclined reader may study the work of [Maddison and Marshall \(2013\)](#), who provide a more general mathematical framework for thickness weighted averaging.

¹The Southern Ocean is a notable ocean exception, as discussed in Sections 20.4 and 32.6.

READER'S GUIDE FOR THIS CHAPTER

In this chapter we assume a working knowledge of the shallow water model presented in Chapters 31, 32, and 35, with particular attention given to the pressure force as realized both as a contact force and a body force (see Chapter 20 and Section 32.4). The TWA equations offer a useful framework for eddy parameterizations, with parameterizations for the tracer equation discussed in Chapters 49 and 50. Parameterizations for both the tracer equation and the momentum equation remain a topic of ongoing research. These studies are of particular focus in physical oceanography given the computational expense of resolving quasi-geostrophic eddies in global climate simulations.

47.1	Loose threads	1050
47.2	The unaveraged thickness weighted equations	1051
47.3	Thickness transport by the bolus velocity	1051
47.3.1	Rectified effects	1052
47.3.2	An undulating fluid layer	1052
47.3.3	Stokes drift	1052
47.3.4	Linearized thickness perturbations	1054
47.3.5	Correlation between thickness and velocity	1054
47.3.6	Do we need the bolus velocity?	1055
47.4	Averaging operators	1055
47.4.1	Reynolds average	1055
47.4.2	Ensemble average	1055
47.4.3	Thickness weighted average	1056
47.4.4	Comments	1056
47.5	Thickness equation and tracer equation	1057
47.5.1	TWA thickness equation	1057
47.5.2	Tracer equation	1058
47.5.3	Vertical velocity	1058
47.6	Horizontal momentum equation	1058
47.6.1	Kinetic stress and Reynolds stress	1059
47.6.2	Thickness and pressure gradient correlation	1059
47.6.3	Zonal and meridional Eliassen-Palm fluxes	1061
47.6.4	Interfacial stresses from geostrophic eddies	1062
47.7	Vorticity and potential vorticity	1063
47.7.1	Derivation	1063
47.7.2	Specializing to non-divergent barotropic flow	1064

47.1 Loose threads

- Build more into the bolus discussion in Section 47.3 as per Section 4.5 of *McWilliams* (2006).
- More discussion of Taylor-Bretherton relation (47.71).
- Write the EP fluxes for quasi-geostrophic shallow water Rossby waves.
- Write the EP fluxes for 2d non-divergent Rossby waves.
- Formulate the TWA energy equations.
- Is there a gauge that offers good options for parameterization?

47.2 The unaveraged thickness weighted equations

The thickness weighted averaging formalism starts from flux-form evolution equations for extensive quantities (per unit horizontal area) rather than intensive quantities. In the context of the shallow water model, this approach means we focus on the thickness equation (31.70a), the thickness weighted tracer equation (31.70b), and the thickness weighted velocity equation (32.45) (also called the momentum equation)

$$\frac{\partial h_k}{\partial t} + \nabla \cdot (h_k \mathbf{u}_k) = 0 \quad (47.1a)$$

$$\frac{\partial(h_k C)}{\partial t} + \nabla \cdot (h_k \mathbf{u}_k C) = 0 \quad (47.1b)$$

$$\frac{\partial(h_k \mathbf{u}_k)}{\partial t} + \nabla \cdot [h_k \mathbf{u}_k \otimes \mathbf{u}_k] + f \hat{\mathbf{z}} \wedge (h_k \mathbf{u}_k) = -(h_k / \rho) \nabla p_k. \quad (47.1c)$$

The density, ρ , appearing in the momentum equation (47.1c) is the Boussinesq reference density, normally chosen as the density of the uppermost layer, $\rho = \rho_1$. The thickness and tracer equations do not couple to other layers, and as such we can drop the layer index, $k = 1, N$, when analyzing these equations. For the momentum equation, we will expose the interface indices, $k \pm 1/2$, when considering pressure form stresses.

For this chapter, it proves useful to move seamlessly between the thickness weighted pressure gradient body force and its equivalent contact force version studied in Section 32.4. The contact force version of the momentum equation reveals the pressure form stresses acting on the upper and lower interfaces of a shallow water layer. It also brings stresses (kinetic stresses and pressure stresses) together into the divergence of a momentum flux. As such, this formulation follows that of Cauchy as discussed in Section 18.1.3. The eddy correlation portion of the momentum flux is known as the *Eliassen-Palm* flux.²

When the dust settles, the TWA equations are isomorphic to the unaveraged equations (47.1a)-(47.1c), yet with the addition of momentum flux convergences to the right hand sides that arise from subgrid correlations (i.e., convergence of the Eliassen-Palm flux). The momentum eddy fluxes are connected to the potential vorticity fluxes, with the connection known as the *Taylor-Bretherton identity*. The isomorphism provides some motivation to favor the TWA approach since properties of the unaveraged equations are directly reflected in the TWA equations. It also provides a suitable framework for parameterizing the subgrid correlations within the context of flux-form conservation laws. Even so, any formalism for an eddy and mean decomposition is subjective since the mean flow and eddying fluctuations are defined by the analyst not by the physics. Hence, arguments concerning what is a preferable framework are subject to the needs of the analyst and have no physically objective foundation.

47.3 Thickness transport by the bolus velocity

Prior to diving into the formalism of thickness weighted averaging, we study the eddy-induced volume transport (more precisely, thickness transport) realized by linear waves within a layer of shallow water fluid. This discussion provides a specific example of the thickness transport by the *bolus velocity*, with further discussion offered in Sections 47.5.1 and 49.4.5. Much of our intuition for bolus transport is based on the following relatively simple example of Stokes drift.³

²See [Bühler \(2014b\)](#) for a historical perspective on the Eliassen-Palm flux, which was introduced by [Eliassen and Palm \(1960\)](#) in their study of stationary mountain waves.

³This example is based Section 2 of [Lee et al. \(1997\)](#).

Part of the motivation for TWA is that we do not need to compute the bolus velocity. Even so, understanding the basic physics of the bolus velocity renders useful insights into how eddies, even eddies as simple as linear waves, can provide a rectified transport of properties. We return to this point in Section 32.4.

47.3.1 Rectified effects

Rectification is the conversion of a fluctuating motion into motion in a particular direction. For example, the transformation of an alternating electrical current into a direct electrical current occurs through a rectifier. More generally, rectification arises from the breaking of a symmetry typically through a nonlinear mechanism. The primary example in fluid mechanics is the Stokes drift discussed in Section 30.3 as well as the current section. Stokes drift arises when linear waves have an amplitude that is a function of space, with this spatial dependence giving rise to net particle transport (the Stokes drift) in a preferred direction. Another example concerns the turbulent Stokes drift arising from nonlinear geostrophic waves and eddies in the ocean and atmosphere that lead to a net transport of buoyancy. In Section 32.6 we discuss the meridional transport of buoyancy by eddies in a channel, which is the canonical geophysical example of eddy induced transport.

47.3.2 An undulating fluid layer

Figure 47.1 shows a layer of constant density shallow water fluid within an adiabatic stacked shallow water model. Since the layers are immiscible, the total volume of fluid within this layer remains constant. In its unperturbed state with flat layer interfaces, the meridional velocity in the fluid layer is zero and the thickness is a constant, h_o . When perturbed, the thickness is written

$$h(y, t) = h_o + h'(y, t), \quad (47.2)$$

where we assume the perturbation only depends on (y, t) for simplicity. The layer thickness changes in time according to the convergence of the advective transport of thickness as found by the thickness equation (47.1a)

$$\frac{\partial h}{\partial t} = -\nabla \cdot (h \mathbf{u}), \quad (47.3)$$

where the convergence is computed within the layer and we drop the k layer index for brevity. As seen by Figure 47.1, undulations of the layer thickness at a point arise from the convergence of thickness advected to that point. Further assuming that there is no zonal dependence ($\partial_x = 0$) leads to the one-dimensional thickness equation

$$\frac{\partial h}{\partial t} = -\frac{\partial (h v)}{\partial y}. \quad (47.4)$$

47.3.3 Stokes drift

Consider a linear wave perturbation in the meridional velocity that propagates in the meridional direction

$$v'(y, t) = v_o \sin(k y - \omega t), \quad (47.5)$$

where k is a constant wave number and ω is a constant frequency. This longitudinal wave is depicted in Figure 47.1. We now follow the general formalism developed in Section 30.3 or equivalently in Section 49.2.4 to determine the Stokes drift associated with this wave.

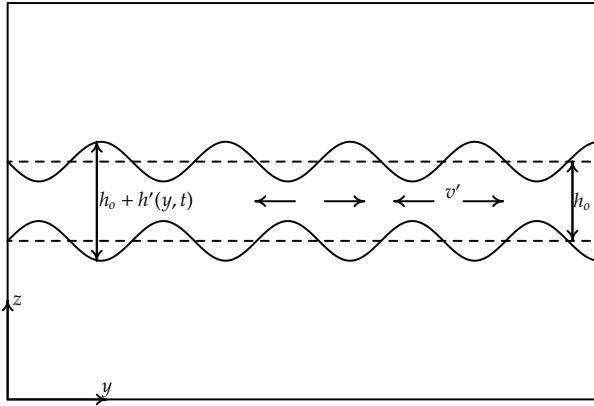


FIGURE 47.1: Shown here is a single layer of constant density fluid, with resting thickness $h = h_o$ and instantaneous thickness $h = h_o + h'(y, t)$. Associated with the undulations in thickness are fluctuations in the meridional velocity $v' = v_o \sin(ky - \omega t)$, depicted here by the alternating vectors within the layer. Vertical-meridional axes are shown in the lower left corner for orientation. We are not concerned with boundaries in the meridional direction.

We are only concerned with the meridional component of the velocity, so the fluid particle trajectory equation is given by

$$\frac{dY}{dt} = v_o \sin(kY - \omega t), \quad (47.6)$$

where $Y = Y(Y_o, t)$ is the meridional trajectory with initial position Y_o . Following equation (30.43) we can write the difference between the velocity following a fluid particle (the Lagrangian velocity for the moving fluid particle) from the velocity at the initial particle point (the Eulerian velocity at the initial point of the trajectory)

$$\frac{dY}{dt} - v(y, t) = v_o^2 k \cos(ky - \omega t) \int_0^t \sin(ky - \omega t') dt' \quad (47.7a)$$

$$= \frac{v_o^2 k}{\omega} (\cos^2(ky - \omega t) - \cos(ky - \omega t) \cos(ky)). \quad (47.7b)$$

Time averaging over a single wave period,

$$T = 2\pi/\omega \quad (47.8)$$

leads to the Stokes drift as per the general expression in equation (30.45)

$$V_{\text{Stokes}} = \frac{v_o^2 k}{2\omega}. \quad (47.9)$$

Introducing the phase speed for the wave $c = \omega/k$ allows us to write the Stokes drift as

$$V_{\text{Stokes}} = \frac{v_o^2}{2c}. \quad (47.10)$$

The Stokes drift becomes small when the phase speed is large, since the fluid particles have only a short time to feel each wave. In this case, there is only a small difference between the Eulerian and Lagrangian velocities. The converse holds for slow phase speeds, where Eulerian and Lagrangian velocities differ more. Note that if we were to consider a more careful asymptotic expansion, then the case of relatively slow phase speeds would require us to keep more terms in the expansion.

47.3.4 Linearized thickness perturbations

The velocity and thickness are written in terms of their rest state plus a perturbation due to the wave

$$h = h_o + h' \quad \text{and} \quad v = v', \quad (47.11)$$

where the velocity vanishes when the wave is absent. The thickness equation (47.4) thus takes the form

$$\frac{\partial h'}{\partial t} + h_o \frac{\partial v'}{\partial y} + v' \frac{\partial h'}{\partial y} = 0. \quad (47.12)$$

Linearizing this equation, and using the wave perturbation (47.5), leads to

$$\frac{\partial h'}{\partial t} + h_o v_o k \cos(ky - \omega t) = 0, \quad (47.13)$$

thus yielding the thickness perturbation

$$h' = h_o v'/c. \quad (47.14)$$

Hence, to leading order, the thickness perturbation is directly proportional to and in phase with the velocity perturbation.

47.3.5 Correlation between thickness and velocity

Over a single wave period $T = 2\pi/\omega$, the temporal correlation between the linear thickness perturbation and velocity perturbation is given by

$$\overline{h' v'} = \frac{1}{T} \int_0^T h' v' dt \quad (47.15a)$$

$$= \frac{h_o}{c T} \int_0^T v' v' dt \quad (47.15b)$$

$$= \frac{v_o^2 h_o}{c T} \int_0^T \sin^2(ky - \omega t) dt \quad (47.15c)$$

$$= \frac{v_o^2 h_o}{2c} \quad (47.15d)$$

$$= h_o V_{\text{Stokes}}, \quad (47.15e)$$

where we introduced the Stokes drift (47.10) to reach the final equality. A nonzero correlation $\overline{h' v'}$ means that the thickness has a nonzero tendency when averaged over a wave period.

The nonzero correlation in equation (47.15e) induces a thickness transport from the one-dimensional linear longitudinal waves. This transport arises from the Stokes drift induced by the waves; without Stokes drift there is no eddy thickness transport. This behavior exemplifies that for more general waves and eddies moving through fluid layers. For the general case, a nonzero bolus velocity (Section 47.5.1), determined by velocity-thickness correlations, induces an eddy thickness transport. We see that for the one-dimensional linear longitudinal wave example, the bolus velocity is the Stokes velocity, thus prompting certain authors to make the equality in general.

47.3.6 Do we need the bolus velocity?

Although the bolus transport is of fundamental importance for how we think about eddy induced Stokes transport from shallow water waves, more general fluctuations, such as those from turbulent geostrophic eddies, require a parameterization to determine the thickness transport. We considered such in Section 26.7 when studying geostrophic eddies in a zonally reentrant channel. As we see in the remainder of this chapter, the allure of TWA is that it dispenses with the need to parameterize the bolus velocity. Instead, the TWA equations absorb the bolus transport into the residual mean advection operator. Doing so moves the parameterization problem fully into the momentum equation.

47.4 Averaging operators

We denote the averaging or mean operator by an overbar

$$\text{average}(\Phi) = \overline{\Phi}, \quad (47.16)$$

where Φ is any field such as velocity, thickness, or tracer concentration. Deviations (also called fluctuations) from the mean are denoted by a prime so that the full field can be decomposed as

$$\Phi = \overline{\Phi} + \Phi'. \quad (47.17)$$

There are many averaging operators used in fluid mechanics, such as the wave average from Section 47.3, which is useful when the flow is dominantly linear waves; a long time average (formally an infinitely long time average), which is commonly used for climate studies; a space average, which is appropriate when the spatial sampling is coarse; general space filters or kernels, which are commonly used in large eddy simulations; and ensemble averages, which are generally assumed in traditional studies of turbulence.

47.4.1 Reynolds average

A *Reynolds average* is an operator that annihilates any fluctuating quantity, which then means that the average of an average is the identity operator

$$\overline{\Phi'} = 0 \iff \overline{\overline{\Phi}} = \overline{\Phi}, \quad (47.18)$$

which in turn means that

$$\overline{\Phi \Psi} = \overline{(\overline{\Phi} + \Phi') \Psi} = \overline{\Phi} \overline{\Psi}. \quad (47.19)$$

Reynolds averages are commonly used when deriving mean field equations, and yet these assumptions are not satisfied by many operators in practical use. Pursuing a formalism with averaging operators that are not Reynolds averages incurs extra technical issues that go beyond our aims, thus motivating us to retain the Reynolds average assumption (47.18) for our average operator.

47.4.2 Ensemble average

A further assumption we make is that the average operator commutes with space-time derivatives and integrals. This assumption does not strictly hold if the operator is a space and/or time average operator, at least not without a bit of work. However, this assumption holds for ensemble averages. An *ensemble average* is computed over an infinite number of realizations of the fluid flow, with approximations to this average afforded by finite sized ensembles.

Ensemble averages are typically assumed in traditional fluid turbulence studies. However, they are not always very practical nor are they the obvious choice when targeting a framework for parameterization. Even so, we prefer ensemble averages in this chapter to dispense with concerns about commutation with derivative and integral operators. We also make use of this averaging for our discussion of tracer kinematics in Chapter 49.

47.4.3 Thickness weighted average

The thickness weighted average of a field is defined as the ensemble average of the thickness weighted field, and then divided by the averaged thickness:

$$\widehat{\Phi} \equiv \frac{\overline{h\Phi}}{\overline{h}} \iff \overline{h}\widehat{\Phi} = \overline{h\Phi}, \quad (47.20)$$

with widehats adorning a thickness weighted average. Deviations from the thickness weighted average are denoted with two primes so that the unaveraged field is decomposed into its average plus fluctuation

$$\Phi = \widehat{\Phi} + \Phi''. \quad (47.21)$$

Since the overline average from Section 47.4.1 satisfies the Reynolds averaging assumption, so too does the thickness weighted average

$$\Phi = \widehat{\Phi} + \Phi'' \implies \widehat{\Phi''} = \frac{\overline{h\Phi''}}{\overline{h}} = 0. \quad (47.22)$$

We are thus able to derive the following related identities

$$\Phi\Psi = (\widehat{\Phi} + \Phi'')(\widehat{\Phi} + \Phi'') \implies \widehat{\Phi\Psi} = \widehat{\Phi}\widehat{\Psi} + \widehat{\Phi''}\widehat{\Psi''} \implies \overline{h\Phi\Psi} = \overline{h}\widehat{\Phi\Psi}. \quad (47.23)$$

We need one final result related to mixed averages and primes, which results from

$$\overline{h}\widehat{\Phi} = \overline{h\Phi} = \overline{h}\widehat{\Phi}, \quad (47.24)$$

so that⁴

$$\overline{h\Phi''} = \overline{h}(\overline{\Phi} - \widehat{\Phi}) = \overline{h\Phi} - \overline{h'}\overline{\Phi'} \neq 0. \quad (47.25)$$

Hence, the ensemble average of a fluctuation relative to thickness weighted mean is generally nonzero.

Derivative operators *do not* commute with the thickness weighted average, so that, for example,

$$\partial_x \widehat{u} \neq \widehat{\partial_x u}. \quad (47.26)$$

Hence, when deriving differential equations for thickness weighted fields, we first derive equations for the unaveraged thickness weighted quantities, and only thereafter do we apply the ensemble mean operator.

47.4.4 Comments

For the most part, we follow the notation of [Young \(2012\)](#). Nonetheless, we caution that notational clutter and variations on conventions can present a nontrivial barrier to penetrating the TWA literature. Indeed, for our purposes with the stacked shallow water model, there is one additional piece of notation concerning the discrete layer indices. Fortunately, much of the discrete layer notation can be streamlined by exposing just the half-integer indices for fields situated at layer interfaces, along with the layer density.

⁴Footnote #4 in [Young \(2012\)](#) is missing the $\overline{h\Phi}$ term appearing in equation (47.25).

47.5 Thickness equation and tracer equation

In this section we derive the TWA versions of the thickness equation (47.1a) and the tracer equation (47.1b). The derivations involve straightforward applications of the TWA averaging properties (47.22) and (47.23).

47.5.1 TWA thickness equation

Taking the ensemble average of the thickness equation (47.1a) renders

$$\partial_t \bar{h} + \nabla \cdot \bar{h} \bar{\mathbf{u}} = 0, \quad (47.27)$$

where we dropped the layer index, k , to reduce notation. Introducing the thickness weighted average according to equation (47.20) brings the thickness equation to the form

$$\partial_t \bar{h} + \nabla \cdot (\bar{h} \hat{\mathbf{u}}) = 0. \quad (47.28)$$

Consequently, the mean layer thickness, \bar{h} , evolves at a point in space according to the convergence of the thickness flux, $-\nabla \cdot (\bar{h} \hat{\mathbf{u}})$, with the flux determined by the thickness weighted velocity, $\hat{\mathbf{u}}$.

We find it useful to introduce the material time derivative operator defined with the thickness weighted velocity⁵

$$\frac{D^\sharp}{Dt} = \frac{\partial}{\partial t} + \hat{\mathbf{u}} \cdot \nabla = \frac{\partial}{\partial t} + \hat{u} \partial_x + \hat{v} \partial_y, \quad (47.29)$$

so that the Eulerian flux-form thickness equation (47.28) can be written in the material time derivative or advective form

$$\frac{D^\sharp \bar{h}}{Dt} = -\bar{h} \nabla \cdot \hat{\mathbf{u}}. \quad (47.30)$$

The isomorphism between the TWA thickness equation (47.28) with the unaveraged thickness equation (31.16) illustrates a distinct advantage of using the thickness weighted velocity, $\hat{\mathbf{u}}$. Even so, for some purposes it is useful to unpack the thickness weighted velocity into its two components

$$\hat{\mathbf{u}} = \bar{u} + \frac{\bar{h}' \mathbf{u}'}{\bar{h}} \equiv \bar{u} + \mathbf{u}^{\text{bolus}}, \quad (47.31)$$

where we defined the bolus velocity as

$$\mathbf{u}^{\text{bolus}} = \frac{\bar{h}' \mathbf{u}'}{\bar{h}}. \quad (47.32)$$

We discussed the bolus velocity in Section 47.3 and see it again in Section 49.4.5 when developing the ensemble mean tracer equation in isopycnal coordinates and considering its parameterizations. However, as per our discussion in Section 47.3.6, we do not need to know the bolus velocity if we write the averaged tracer and momentum equations in terms of the thickness weighted velocity.

⁵The D^\sharp/Dt notation is based on that used by [Young \(2012\)](#). The alternative \hat{D}/Dt is less suitable since $\bar{h}(\hat{D}/Dt) \neq (\bar{D}/Dt)\bar{h}$. In brief, an object adorned with a sharp symbol is consistent with thickness weighted averaging but is itself not the direct result of a thickness weighted average.

47.5.2 Tracer equation

Taking the ensemble average of the tracer concentration equation (47.1b) for a shallow water fluid layer renders

$$\partial_t(\bar{h} \bar{C}) + \nabla \cdot (\bar{h} \bar{C} \bar{\mathbf{u}}) = 0. \quad (47.33)$$

Making use of the thickness weighted averages from Section 47.4.3 allows us to write

$$\bar{h} \bar{C} = \bar{h} \hat{C} \quad \text{and} \quad \bar{h} \bar{C} \bar{\mathbf{u}} = \bar{h} (\hat{C} \hat{\mathbf{u}} + \widehat{C'' \mathbf{u}''}), \quad (47.34)$$

thus yielding the TWA tracer equation

$$\partial_t(\bar{h} \hat{C}) + \nabla \cdot (\bar{h} \hat{C} \hat{\mathbf{u}}) = -\nabla \cdot (\bar{h} \widehat{C'' \mathbf{u}''}). \quad (47.35)$$

The right hand side is the convergence of the thickness weighted eddy tracer flux. As seen in Section 49.5, the isopycnal form of the tracer equation is identical to that given here for a shallow water layer. In that discussion we present methods commonly used to parameterize the eddy flux convergence.

47.5.3 Vertical velocity

We generally have no need for the vertical velocity when working with the adiabatic stacked shallow water model. Nonetheless, it is interesting to define a vertical velocity component, w^\sharp , satisfying the continuity equation

$$\nabla_z \cdot \hat{\mathbf{u}} + \partial_z w^\sharp = 0. \quad (47.36)$$

As for the unaveraged vertical velocity component, $\partial_z w^\sharp$ is a linear function of z within the ensemble mean shallow water layers. Note that w^\sharp is not a thickness weighted velocity. Rather, it is the vertical velocity that is compatible, through the continuity equation, with the thickness weighted horizontal velocity. Note that a vertical velocity is needed for the continuously stratified Boussinesq fluid, and it is defined just as here for the shallow water.

47.6 Horizontal momentum equation

Taking the ensemble mean of the horizontal momentum equation (47.1c) renders

$$\partial_t(\bar{h} \bar{\mathbf{u}}) + \nabla \cdot [\bar{h} \bar{\mathbf{u}} \otimes \bar{\mathbf{u}}] + f \hat{z} \wedge (\bar{h} \bar{\mathbf{u}}) = -\bar{h} \nabla p / \rho, \quad (47.37)$$

where we dropped the layer interface label, k , for brevity. Again, we make use of the thickness weighted averages from Section 47.4.3 to write

$$\bar{h} \bar{\mathbf{u}} = \bar{h} \hat{\mathbf{u}} \quad (47.38a)$$

$$\bar{h} \bar{\mathbf{u}} \otimes \bar{\mathbf{u}} = \bar{h} (\hat{\mathbf{u}} \otimes \hat{\mathbf{u}} + \widehat{\mathbf{u}'' \otimes \mathbf{u}''}), \quad (47.38b)$$

so that equation (47.37) becomes

$$\partial_t(\bar{h} \hat{\mathbf{u}}) + \nabla \cdot [\bar{h} \hat{\mathbf{u}} \otimes \hat{\mathbf{u}}] + f \hat{z} \wedge (\bar{h} \hat{\mathbf{u}}) = -\nabla \cdot [\bar{h} \widehat{\mathbf{u}'' \otimes \mathbf{u}''}] - \bar{h} \nabla p / \rho. \quad (47.39)$$

The first term on the right hand side is similar to the eddy tracer flux convergence appearing in the TWA tracer equation (47.35). In contrast, the thickness weighted pressure gradient is fundamentally distinct from anything appearing in the tracer equation. Much in the remainder of this section is devoted to developing a physical and mathematical understanding of this term.

47.6.1 Kinetic stress and Reynolds stress

Following our discussion in Section 19.5, we introduce the shallow water kinetic stress tensor

$$\overline{\mathbb{T}^{\text{sw kinetic}}} = -\rho \mathbf{u} \otimes \mathbf{u}. \quad (47.40)$$

The kinetic stress arises from motion of the fluid, with its divergence, $\nabla \cdot (h \mathbb{T}^{\text{kinetic}})$, contributing to changes in the momentum of a shallow water fluid column. Decomposing the velocity into the TWA velocity and fluctuation leads to the ensemble mean of the thickness weighted kinetic stress

$$\overline{h \mathbb{T}^{\text{kinetic}}(\mathbf{u})} = -\rho \overline{h \mathbf{u} \otimes \mathbf{u}} \quad (47.41a)$$

$$= -\rho \overline{h} [\widehat{\mathbf{u}} \otimes \widehat{\mathbf{u}} + \widehat{\mathbf{u}''} \otimes \widehat{\mathbf{u}''}] \quad (47.41b)$$

$$= \overline{h} \mathbb{T}^{\text{kinetic}}(\widehat{\mathbf{u}}) + \overline{h} \mathbb{T}^{\text{Reynolds}}, \quad (47.41c)$$

where the eddy correlation is known as the *Reynolds stress*. The divergence of the thickness weighted Reynolds stress provides a rectified effect onto the mean flow due to nonzero eddy correlations.

47.6.2 Thickness and pressure gradient correlation

The ensemble mean of the thickness weighted pressure gradient can be written

$$\overline{h_k \nabla p_k} = \overline{h_k} \nabla \overline{p_k} + \overline{h'_k} \nabla \overline{p'_k}. \quad (47.42)$$

The eddy term is the correlation between layer thickness fluctuations and horizontal pressure gradient fluctuations. We here unpack this term by writing the pressure as a contact force rather than a body force. Doing so exposes the eddy interfacial form stresses that provide a vertical transfer of horizontal momentum. Note that some treatments simply say that $\overline{h'_k} \nabla \overline{p'_k}$ is the interfacial form stress. However, that identification is incorrect since, as we will see in this section, there is an additional piece arising from the gradient in the layer depth integrated pressure or, alternatively, the layer potential energy. To proceed we rely on the development given in Section 32.4.9, in which we exposed two equivalent expressions for the contact pressure. As part of the following derivation we make use of relations for pressure within a layer and at an interface

$$p_k(z) = p_{k-1/2} + g \rho_k (\eta_{k-1/2} - z) \quad (47.43a)$$

$$p_{k+1/2} - p_{k-1/2} = g \rho_k h_k = g \rho_k (\eta_{k-1/2} - \eta_{k+1/2}) \quad (47.43b)$$

$$p_{1/2} = p_a. \quad (47.43c)$$

Interfacial form stress plus gradient of layer depth integrated pressure

The first expression for thickness weighted pressure gradient is given by

$$-h_k \nabla p_k = -\nabla P_k + \mathbf{F}_k^{\text{form}}. \quad (47.44)$$

In this equation we introduced the vertically integrated pressure computed over layer-k

$$P_k = \int_{\eta_{k+1/2}}^{\eta_{k-1/2}} p_k(z) dz = h_k (g \rho_k h_k / 2 + p_{k-1/2}), \quad (47.45)$$

with its gradient providing a horizontal acceleration due to imbalances in the contact pressure acting along the vertical sides of an infinitesimal shallow water column. The second stress in equation (47.44) is the pressure form stress acting on the upper and lower layer interfaces

$$\mathbf{F}_k^{\text{form}} = p_{k-1/2} \nabla \eta_{k-1/2} - p_{k+1/2} \nabla \eta_{k+1/2} \equiv \delta_k (p_{k-1/2} \nabla \eta_{k-1/2}), \quad (47.46)$$

where

$$\delta_k \Phi_{k-1/2} = \Phi_{k-1/2} - \Phi_{k+1/2} = -(\Phi_{k+1/2} - \Phi_{k-1/2}) \quad (47.47)$$

is a difference operator acting on interface properties. The use of a backward difference operator is motivated since k increases down whereas \hat{z} points up. Additionally, we define the difference operator to only act on fields defined at the layer interface, with layer fields commuting with this operator so that, for example,

$$\delta_k(h_k \eta_{k-1/2}) = h_k \delta_k(\eta_{k-1/2}). \quad (47.48)$$

Dual interfacial pressure form stress plus gradient of layer potential energy

We focus on the following formulation that uses the dual form stress and potential energy since it lends itself to a straightforward interpretation for the case of geostrophic eddies. In this case we write the thickness weighted horizontal pressure gradient as

$$-h_k \nabla p_k = -\nabla \mathcal{P}_k + \mathbf{F}_k^{\text{dual}}. \quad (47.49)$$

In this equation we introduced the layer gravitational potential energy per area

$$\mathcal{P}_k = g \rho_k \int_{\eta_{k+1/2}}^{\eta_{k-1/2}} z dz = (g \rho_k / 2) (\eta_{k-1/2}^2 - \eta_{k+1/2}^2) = (g \rho_k / 2) \delta_k(\eta_{k-1/2}^2), \quad (47.50)$$

and the *dual pressure form stress* (see Section 32.4.8),

$$\mathbf{F}_k^{\text{dual}} = -\delta_k(\eta_{k-1/2} \nabla p_{k-1/2}) = \mathbf{F}_k^{\text{form}} - \nabla[\delta_k(\eta_{k-1/2} p_{k-1/2})]. \quad (47.51)$$

Since they differ by a gradient, the form stress and dual form stress have identical curls and so they contribute the same interfacial pressure torque as part of the layer vorticity evolution

$$-\nabla \wedge (h_k \nabla p_k) = \nabla \wedge \mathbf{F}_k^{\text{dual}} = \nabla \wedge \mathbf{F}_k^{\text{form}}. \quad (47.52)$$

The dual form stress is somewhat more convenient since, when the eddies are geostrophic, we can convert the eddy correlation into an eddy transport of buoyancy. We encountered that feature in Section 26.7 when studying the meridional transport from geostrophic eddies in a zonally reentrant channel.

Thickness and pressure gradient correlation

Making use of the potential energy and dual form stress as given by equation (47.49) allows us to write the ensemble mean thickness weighted pressure gradient as

$$-\overline{h_k \nabla p_k} = -\nabla \overline{\mathcal{P}_k} - \delta_k [\overline{\eta_{k-1/2} \nabla p_{k-1/2}}] \quad (47.53)$$

where minus the potential energy gradient is decomposed as

$$-\nabla \overline{\mathcal{P}_k} = -(g \rho_k / 2) \delta_k [\nabla(\overline{\eta_{k-1/2}})^2 + \nabla(\overline{\eta'_{k-1/2}})^2] \quad (47.54)$$

and the vertical convergence of the dual form stress is

$$-\delta_k [\overline{\eta_{k-1/2} \nabla p_{k-1/2}}] = -\delta_k [\overline{\eta_{k-1/2} \nabla p_{k-1/2}} + \overline{\eta'_{k-1/2} \nabla p'_{k-1/2}}]. \quad (47.55)$$

We are thus led to decompose the thickness weighted pressure gradient correlation as

$$-\overline{h'_k \nabla p'_k} = -(g \rho_k / 2) \nabla [\delta_k (\overline{\eta'_{k-1/2}})^2] - \delta_k [\overline{\eta'_{k-1/2} \nabla p'_{k-1/2}}]. \quad (47.56)$$

Again, the first term on the right hand side arises from the eddy potential energy and the second term from the dual eddy interfacial form stress. For orientation, in Figure 47.2 we illustrate the deviations of the interface positions relative to the ensemble mean.

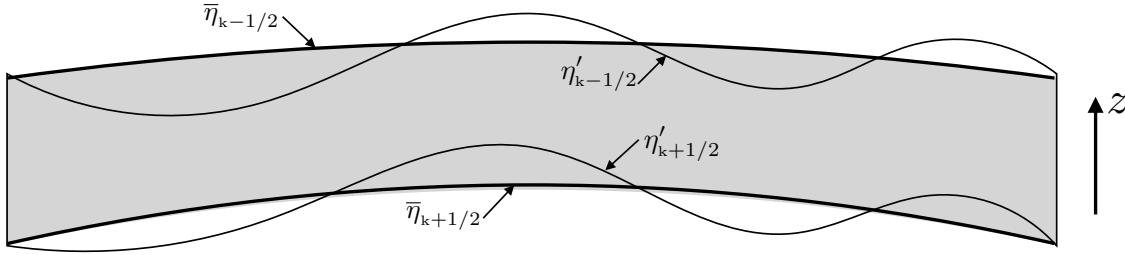


FIGURE 47.2: Schematic of the interface positions for a shallow water layer. The ensemble mean layer interfaces have vertical positions given by $z = \bar{\eta}_{k\pm 1/2}$, whereas the fluctuating interfaces are located at $z = \bar{\eta}_{k\pm 1/2} + \eta'_{k\pm 1/2}$. As depicted here, the ensemble mean interface positions are not generally horizontal.

47.6.3 Zonal and meridional Eliassen-Palm fluxes

Making use of the thickness and pressure gradient correlation in the form of equation (47.56) leads to the thickness weighted momentum equation

$$\begin{aligned} \partial_t(\bar{h}\hat{\mathbf{u}}) + \nabla \cdot (\bar{h}\hat{\mathbf{u}} \otimes \hat{\mathbf{u}}) + f\hat{\mathbf{z}} \wedge \bar{h}\hat{\mathbf{u}} + \bar{h}\nabla\bar{p}/\rho \\ = -\nabla \cdot (\bar{h}\hat{\mathbf{u}}'' \widehat{\otimes} \hat{\mathbf{u}}'') - (g\rho_k/2\rho)\nabla[\delta_k(\eta'_{k-1/2})^2] - \delta_k[\overline{\eta'_{k-1/2}\nabla p'_{k-1/2}}/\rho], \end{aligned} \quad (47.57)$$

where we only exposed the interface indices to reduce notational clutter. The subgrid scale correlations on the right hand side can be organized into the divergence of two tensors

$$\begin{aligned} \nabla \cdot (\bar{h}\hat{\mathbf{u}}'' \widehat{\otimes} \hat{\mathbf{u}}'') + (g\rho_k/2\rho)\nabla[\delta_k(\eta'_{k-1/2})^2] + \delta_k[\overline{\eta'_{k-1/2}\nabla p'_{k-1/2}}/\rho] \\ = \begin{bmatrix} \partial_x & \partial_y & \bar{h}^{-1}\delta_k \end{bmatrix} \begin{bmatrix} \bar{h}\hat{\mathbf{u}}''\hat{\mathbf{u}}'' & \bar{h}\hat{\mathbf{u}}''\hat{\mathbf{v}}'' & 0 \\ \bar{h}\hat{\mathbf{u}}''\hat{\mathbf{v}}'' & \bar{h}\hat{\mathbf{v}}''\hat{\mathbf{v}}'' & 0 \\ 0 & 0 & 0 \end{bmatrix} \end{aligned} \quad (47.58)$$

$$+ \begin{bmatrix} \partial_x & \partial_y & \bar{h}^{-1}\delta_k \end{bmatrix} \begin{bmatrix} (g\rho_k/2\rho)[\delta_k(\eta'_{k-1/2})^2] & 0 & 0 \\ 0 & (g\rho_k/2\rho)[\delta_k(\eta'_{k-1/2})^2] & 0 \\ (\bar{h}/\rho)\eta'_{k-1/2}\partial_x p'_{k-1/2} & (\bar{h}/\rho)\eta'_{k-1/2}\partial_y p'_{k-1/2} & 0 \end{bmatrix}. \quad (47.59)$$

The first tensor arises from Reynolds stresses and the second arises from eddy potential energy and eddy dual form stresses. When combined, the columns are the thickness weighted *Eliassen-Palm fluxes* for the zonal (column 1) and meridional (column 2) shallow water momentum equation

$$\mathbf{E}^{(uEP)} = \left[\bar{h}\hat{\mathbf{u}}''\hat{\mathbf{u}}'' + (g\rho_k/2\rho)[\delta_k(\eta'_{k-1/2})^2] \right] \hat{\mathbf{x}} + \bar{h}\hat{\mathbf{u}}''\hat{\mathbf{v}}'' \hat{\mathbf{y}} + \bar{h}(\overline{\eta'_{k-1/2}\partial_x p'_{k-1/2}}/\rho) \hat{\mathbf{z}} \quad (47.60a)$$

$$\mathbf{E}^{(vEP)} = \bar{h}\hat{\mathbf{u}}''\hat{\mathbf{v}}'' \hat{\mathbf{x}} + \left[\bar{h}\hat{\mathbf{v}}''\hat{\mathbf{v}}'' + (g\rho_k/2\rho)[\delta_k(\eta'_{k-1/2})^2] \right] \hat{\mathbf{y}} + \bar{h}(\overline{\eta'_{k-1/2}\partial_y p'_{k-1/2}}/\rho) \hat{\mathbf{z}}. \quad (47.60b)$$

We encountered the unaveraged version of these eddy stresses in Sections 32.3.3 and 32.4.9 when studying the shallow water momentum equation. [Maddison and Marshall \(2013\)](#) included the third column of zeros in equation (47.59) to emphasize that the Eliassen-Palm fluxes are the first and second columns to the *Eliassen-Palm flux tensor*. They illustrated the utility of this perspective by considering gauge transformations that result in non-zero elements in the third column.

The Eliassen-Palm fluxes are second order in eddy amplitude; i.e., they are quadratic in eddy fluctuations. Furthermore, they bring together the Reynolds stress, the eddy potential energy, and the dual eddy interfacial form stress. The convergence of the Eliassen-Palm fluxes provides an eddy

rectified acceleration on the thickness weighted velocity. To explicitly see this forcing, write the components to the mean field momentum equation (47.57) as⁶

$$\partial_t(\bar{h}\hat{u}) + \nabla \cdot (\bar{h}\hat{\mathbf{u}}\hat{u}) - f\bar{h}\hat{v} + \bar{h}\partial_x\bar{p}/\rho = -(\nabla_z + \hat{z}\bar{h}^{-1}\delta_k) \cdot \mathbf{E}^{(uEP)} \quad (47.61a)$$

$$\partial_t(\bar{h}\hat{v}) + \nabla \cdot (\bar{h}\hat{\mathbf{u}}\hat{v}) + f\bar{h}\hat{u} + \bar{h}\partial_y\bar{p}/\rho = -(\nabla_z + \hat{z}\bar{h}^{-1}\delta_k) \cdot \mathbf{E}^{(vEP)}. \quad (47.61b)$$

Equations (47.61a) and (47.61b) are isomorphic to the unaveraged horizontal momentum equation (47.1c), yet with the addition of the convergence of the Eliassen-Palm flux on the right hand side that encapsulate rectified effects from eddies. They can be written using the material time derivative (47.29)

$$\frac{D^\# \hat{u}}{Dt} - f\hat{v} + \partial_x\bar{p}/\rho = -\bar{h}^{-1}(\nabla_z + \hat{z}\bar{h}^{-1}\delta_k) \cdot \mathbf{E}^{(uEP)} \quad (47.62a)$$

$$\frac{D^\# \hat{v}}{Dt} + f\hat{u} + \partial_y\bar{p}/\rho = -\bar{h}^{-1}(\nabla_z + \hat{z}\bar{h}^{-1}\delta_k) \cdot \mathbf{E}^{(vEP)}. \quad (47.62b)$$

It is notable that these equations only make use of the thickness weighted velocity, $\hat{\mathbf{u}}$, as do the averaged thickness equation (47.28) and adveraged tracer equation (47.35). We advertised this point near the start of this chapter, noting that it greatly facilitates the practical use of the TWA equations. We further this correspondence in Section 47.7 by showing that the vorticity and potential vorticity equations also make use of $\hat{\mathbf{u}}$.

47.6.4 Interfacial stresses from geostrophic eddies

In Section 26.7 we studied the rectified effects from geostrophic eddies in a zonally re-entrant channel for a continuously stratified fluid. In that analysis we found that the zonal mean of isopycnal eddy form stresses are equivalent to the meridional eddy flux of buoyancy. We here consider similar questions within the context of the TWA shallow water fluid, here focusing on the interfacial transfer of momentum due to eddy dual form stresses as given by the vertical vectors

$$\mathbf{E}_{\text{form}}^{(uEP)} = \bar{h}(\overline{\eta'_{k-1/2} \partial_x p'_{k-1/2}}/\rho)\hat{z} \quad (47.63a)$$

$$\mathbf{E}_{\text{form}}^{(vEP)} = \bar{h}(\overline{\eta'_{k-1/2} \partial_y p'_{k-1/2}}/\rho)\hat{z}. \quad (47.63b)$$

As for the analysis in Section 26.7, we assume the fluctuations are geostrophic. In this manner we introduce an interface geostrophic velocity corresponding to the gradient of the interface pressure fluctuations⁷

$$\partial_x p'_{k-1/2} = f\rho v'_{k-1/2} \quad \text{and} \quad \partial_y p'_{k-1/2} = -f\rho u'_{k-1/2}, \quad (47.64)$$

in which case the dual form stress portion of the Eliassen-Palm fluxes take the form

$$\mathbf{E}_{\text{form}}^{(uEP)} = f\bar{h}(\overline{\eta'_{k-1/2} v'_{k-1/2}})\hat{z} \quad (47.65a)$$

$$\mathbf{E}_{\text{form}}^{(vEP)} = -f\bar{h}(\overline{\eta'_{k-1/2} u'_{k-1/2}})\hat{z}. \quad (47.65b)$$

For geostrophic eddies, the dual form stress terms take the form of an eddy transport of the area between $z = \bar{\eta}_{k-1/2}$ and $z = \eta'_{k-1/2}$ (see Figure 47.2). In a continuously stratified fluid, as in Section 26.7, we can relate the interface fluctuations, η' , to the buoyancy fluctuations, b' , in which case $\overline{\eta'_{k-1/2} u'_{k-1/2}}$ is proportional to the eddy buoyancy flux. Likewise, the potential energy term can be related to fluctuations in the squared buoyancy.

⁶Recall from equation (47.48) that the operator δ_k only acts on interface fields, so that $\bar{h}^{-1}\delta_k(\bar{h}) = \delta_k$. This convention thus allows us to unify the horizontal components to the Eliassen-Palm flux with the vertical component, as written in equations (47.60a) and (47.60b).

⁷Recall that ρ is the Boussinesq reference density, whereas ρ_k is the density for layer k.

47.7 Vorticity and potential vorticity

We follow the procedure from Chapter 35 to derive the vorticity and potential vorticity for the thickness weighted shallow water equations. In the process, we connect the eddy flux of potential vorticity to the Eliassen-Palm flux.

47.7.1 Derivation

Use the vector identities from Sections 35.1 and 35.2 to bring the material evolution equations (47.62a) and (47.62b) into their equivalent vector invariant forms

$$\partial_t \hat{u} - (f + \hat{\zeta}) \hat{v} = -\partial_x (\bar{p}/\rho + \hat{u} \cdot \hat{u}/2) - \bar{h}^{-1} (\nabla_z + \hat{z} \bar{h}^{-1} \delta_k) \cdot \mathbf{E}^{(uEP)} \quad (47.66a)$$

$$\partial_t \hat{v} + (f + \hat{\zeta}) \hat{u} = -\partial_y (\bar{p}/\rho + \hat{u} \cdot \hat{u}/2) - \bar{h}^{-1} (\nabla_z + \hat{z} \bar{h}^{-1} \delta_k) \cdot \mathbf{E}^{(vEP)}, \quad (47.66b)$$

where we introduced the relative vorticity of the thickness weighted horizontal velocity

$$\hat{\zeta} = \partial_x \hat{v} - \partial_y \hat{u}. \quad (47.67)$$

Taking ∂_x of the meridional equation (47.66b) and subtracting ∂_y of the zonal equation (47.66a) renders the evolution equation for absolute vorticity, $\hat{\zeta}_a = \hat{\zeta} + f$,

$$\frac{D^\sharp \hat{\zeta}_a}{Dt} + \hat{\zeta}_a \nabla \cdot \hat{u} = \partial_y [\bar{h}^{-1} (\nabla_z + \hat{z} \bar{h}^{-1} \delta_k) \cdot \mathbf{E}^{(uEP)}] - \partial_x [\bar{h}^{-1} (\nabla_z + \hat{z} \bar{h}^{-1} \delta_k) \cdot \mathbf{E}^{(vEP)}]. \quad (47.68)$$

Making use of the thickness equation (47.30) to replace $\nabla \cdot \hat{u}$ leads to the potential vorticity equation

$$\bar{h} \frac{D^\sharp \Pi^\sharp}{Dt} = -\nabla \cdot \mathbf{F}^\sharp \quad (47.69)$$

where

$$\Pi^\sharp = \frac{f + \partial_x \hat{v} - \partial_y \hat{u}}{\bar{h}} = \frac{f + \hat{\zeta}}{\bar{h}} \quad (47.70)$$

is the potential vorticity defined with the thickness weighted velocity and ensemble mean thickness. The corresponding eddy potential vorticity flux is a horizontal vector and given in terms of the divergence of the Eliassen-Palm fluxes

$$\mathbf{F}^\sharp = \hat{x} [\bar{h}^{-1} (\nabla_z + \hat{z} \bar{h}^{-1} \delta_k) \cdot \mathbf{E}^{(vEP)}] - \hat{y} [\bar{h}^{-1} (\nabla_z + \hat{z} \bar{h}^{-1} \delta_k) \cdot \mathbf{E}^{(uEP)}] + \hat{z} \wedge \nabla \Upsilon, \quad (47.71)$$

where Υ is an arbitrary gauge function.⁸ This equation connects the potential vorticity flux to the Eliassen-Palm fluxes and it is known as the *Taylor-Bretherton identity*. Remarkably, the potential vorticity flux also provides the eddy forcing to the thickness weighted velocity equation

$$\partial_t \hat{u} + (f + \hat{\zeta}) \hat{z} \wedge \hat{u} + \nabla (\bar{p}/\rho + \hat{u} \cdot \hat{u}/2) = -\hat{z} \wedge (\mathbf{F}^\sharp - \hat{z} \wedge \nabla \Upsilon), \quad (47.72)$$

which can also be written

$$\partial_t \hat{u} + \hat{z} \wedge (\bar{h} \hat{u} \Pi^\sharp + \mathbf{F}^\sharp - \hat{z} \wedge \nabla \Upsilon) + \nabla (\bar{p}/\rho + \hat{u} \cdot \hat{u}/2) = 0, \quad (47.73)$$

where $\bar{h} \hat{u} \Pi^\sharp + \mathbf{F}^\sharp - \hat{z} \wedge \nabla \Upsilon$ is the net (mean plus eddy plus gauge) potential vorticity flux.

⁸Equation (129) in [Young \(2012\)](#) should have a gauge function on its right hand side, which follows from his footnote #3. We provide an example of the need for a gauge function in Section 47.7.2.

47.7.2 Specializing to non-divergent barotropic flow

In Chapter 34 we studied the mechanics of a two dimensional fluid whose horizontal flow is non-divergent. As for the shallow water, the fluid moves as columns. However, since the horizontal flow is non-divergent, each column is rigid and so there is no stretching or squashing of columns. Correspondingly, there are no form stresses acting on these columns. We specialize the shallow water analysis in this section to rigid columnar motion as a means to verify that the Reynolds stresses appearing in the Eliassen-Palm flux formulation correspond to that arising from the non-divergent barotropic model.

For rigid fluid columns, the thickness weighted average reduces to just the ensemble mean since all layer thicknesses are fixed. Correspondingly, there are no form stresses acting at the layer interfaces since interfaces are horizontal. Hence, the Eliassen-Palm fluxes (47.60a) and (47.60b) reduce to just their Reynold stress contributions

$$h^{-1} \mathbf{E}^{(uEP)} = \overline{u' u'} \hat{\mathbf{x}} + \overline{u' v'} \hat{\mathbf{y}} \quad (47.74a)$$

$$h^{-1} \mathbf{E}^{(vEP)} = \overline{u' v'} \hat{\mathbf{x}} + \overline{v' v'} \hat{\mathbf{y}}. \quad (47.74b)$$

The corresponding eddy potential vorticity flux (47.71), absent the gauge term, is

$$\mathbf{F}^\sharp = \hat{\mathbf{x}} \nabla \cdot [\overline{u' v'} \hat{\mathbf{x}} + \overline{v' v'} \hat{\mathbf{y}}] - \hat{\mathbf{y}} \nabla \cdot [\overline{u' u'} \hat{\mathbf{x}} + \overline{u' v'} \hat{\mathbf{y}}] \quad (47.75a)$$

$$= \hat{\mathbf{x}} [\partial_x(\overline{u' v'}) + \partial_y(\overline{v' v'})] - \hat{\mathbf{y}} [\partial_x(\overline{u' u'}) + \partial_y(\overline{u' v'})]. \quad (47.75b)$$

Does the eddy potential vorticity flux (47.75b) agree, to within a gauge function, with the eddy flux resulting from a direct decomposition into eddy and mean within a two dimensional non-divergent model? To address this question, recall that the advective flux of potential vorticity for the two dimensional non-divergent flow is given by equation (34.38)

$$\mathbf{u} q = \mathbf{u} f + \nabla \cdot (\hat{\mathbf{z}} \wedge \mathcal{E}), \quad (47.76)$$

where \mathcal{E} is the trace-free anisotropic portion of the kinetic stress tensor

$$\mathcal{E} = \begin{bmatrix} -(u^2 - v^2)/2 & -uv \\ -uv & (u^2 - v^2)/2 \end{bmatrix}. \quad (47.77)$$

The mean of the potential vorticity flux is (47.76) is given by

$$\overline{\mathbf{u} q} = \overline{\mathbf{u} f} + \overline{\mathbf{u}' q'}, \quad (47.78)$$

where the flux computed from the mean fields is

$$\overline{\mathbf{u} q} = \overline{\mathbf{u}} (f + \bar{\zeta}), \quad (47.79)$$

whereas the eddy potential vorticity flux is

$$\overline{\mathbf{u}' q'} = \nabla \cdot [\hat{\mathbf{z}} \wedge \overline{\mathcal{E}(\mathbf{u}')}] \quad (47.80a)$$

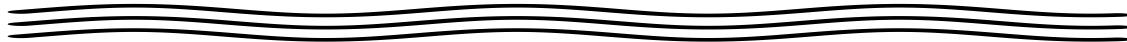
$$= \hat{\mathbf{x}} [\partial_x(\overline{u' v'}) + \partial_y(\overline{v' v'} - \overline{u' u'})/2] + \hat{\mathbf{y}} [\partial_x(\overline{v' v'} - \overline{u' u'})/2 - \partial_y(\overline{u' v'})] \quad (47.80b)$$

$$= \hat{\mathbf{x}} [\partial_x(\overline{u' v'}) + \partial_y(\overline{v' v'})] - \hat{\mathbf{y}} [\partial_x(\overline{u' u'}) + \partial_y(\overline{u' v'})] + (\hat{\mathbf{x}} \partial_y - \hat{\mathbf{y}} \partial_x) \overline{\mathbf{u}' \cdot \mathbf{u}'} / 2 \quad (47.80c)$$

$$= \mathbf{F}^\sharp - \hat{\mathbf{z}} \wedge \nabla(\overline{\mathbf{u}' \cdot \mathbf{u}'}) / 2. \quad (47.80d)$$

Hence, $\overline{\mathbf{u}' q'}$ agrees with \mathbf{F}^\sharp in equation (47.75b) to within a gauge function given by the eddy kinetic energy, so that their divergences are equal

$$\nabla \cdot \mathbf{F}^\sharp = \nabla \cdot \overline{\mathbf{u}' q'}. \quad (47.81)$$



Part IX

Mechanics of ocean scalars

Tracers and buoyancy are scalar fields that offer fundamental information about the mechanics of a fluid. For the ocean, tracers and buoyancy are critical to both its fluid mechanics and its role as part of the broader earth climate system. In this part of the book we study a suite of topics related to the physics and mathematics of tracers and buoyancy. This *scalar mechanics* offers a distinct view on the fluid flows and thus provides a useful complement to the mechanics of momentum, vorticity, and energy. Throughout this part of the book we encounter some of the canonical partial differential equations appearing in fluid mechanics. Mathematical facets of these equations are introduced in Chapter 3, which offers a synopsis of the linear partial differential equations of mathematical physics. By studying the chapters in this part of the book, we develop an appreciation of the many facets involved with studying ocean tracers and buoyancy. Here is a brief summary of these chapters.

- In Chapter 48 we dive deep into the maths and physics of advection and diffusion, which are the two processes that affect the evolution of ocean tracers.
- In Chapter 49 we introduce the notions of wave-mean flow interactions that give rise to eddy-induced advection (or skew diffusion) as well as diffusion. This chapter makes use of both geopotential coordinates as well as isopycnal coordinate equations from Chapter 46, thus providing an example of the dual roles these coordinates make in describing turbulent ocean flows.
- Chapter 50 offers a summary of the standard tracer parameterizations used for coarse resolution models of the ocean circulation. This chapter is quite useful to those wishing to get an overview of the methods used by many of the ocean models in use for climate studies.
- Chapter 51 then shifts gears a bit by focusing on ocean buoyancy and the study of global sea level. This topic is of growing concern for climate science as the ocean feels the effects from a warming atmosphere.
- Chapter 52 finishes the book with a long look at the mathematics and physics of water mass analysis.

Tracer advection and diffusion

In this chapter we study physical and mathematical aspects of advection and diffusion. These two processes affect matter concentrations through stirring and mixing, and as such we frame the presentation in terms of the tracer equation. Advection and diffusion also affect momentum and vorticity, though both are also affected by other processes such as pressure and Coriolis (for momentum; see Chapter 18) and stretching, twisting, and baroclinicity (for vorticity; see Chapter 36).

In the absence of diffusion, advection imparts a reversible stirring of fluid elements that increases the magnitude of tracer gradients. In contrast, diffusion provides an irreversible mixing of fluid elements that reduces the magnitude of tracer gradients. When acting together, advection is no longer a pure stirring and diffusion is no longer a pure mixing. [Eckart \(1948\)](#) articulated what has become the standard conceptual paradigm for stirring and mixing in geophysical fluids, with elements of that paradigm reflected in this chapter.

READER'S GUIDE FOR THIS CHAPTER

We make use of the tracer equation derived in Section 16.1, with results relevant for matter transport in both the atmosphere and ocean.^a The discussion of Green's functions in Section 48.10 is dependent on our study of Green's functions in Chapter 4, and this material supplements the [Haine et al. \(2021\)](#) review paper that summarizes many applications of Green's functions for passive ocean tracers.

^aResults for the oceanic Boussinesq fluid of Chapter 24 are found merely by setting the density factor, ρ , to a constant where it appears in the non-Boussinesq equations.

48.1	Loose threads	1068
48.2	Introduction	1068
48.3	Diffusion physics	1069
48.3.1	Diffusion of matter by random molecular motions	1069
48.3.2	Diffusion of matter by random turbulent motions	1070
48.3.3	Fick's law for matter diffusion	1070
48.3.4	Fourier's law for heat diffusion	1071
48.3.5	Newtonian frictional stress and momentum diffusion	1072
48.3.6	Further study	1072
48.4	Diffusion maths	1073
48.4.1	Sample diffusion tensors	1073
48.4.2	Evolution of tracer concentration powers	1073
48.4.3	Moments of tracer concentration	1074

48.4.4	Connecting tracer variance to the diffusion operator	1075
48.5	Advection physics	1077
48.5.1	The advection equation	1077
48.5.2	Eulerian time tendencies from advection	1078
48.5.3	Impermeability property of tracer isosurfaces	1078
48.6	Advection maths	1079
48.6.1	Material constancy of C^Γ	1079
48.6.2	Mass transport	1079
48.6.3	Advective tracer fluxes and skew tracer fluxes	1080
48.6.4	Skew diffusion	1081
48.6.5	A comment about skew fluxes and Lagrangian kinematics	1082
48.6.6	Further reading	1082
48.7	Advection and skewson	1082
48.7.1	Choosing a gauge	1083
48.7.2	Vertical gauge	1084
48.7.3	Boundary conditions	1084
48.8	Finite volume budgets with eddy velocities	1085
48.8.1	Advective formulation	1086
48.8.2	Skew flux formulation	1086
48.8.3	Domain with a tracer boundary	1087
48.8.4	Budget for a region with interior sides	1089
48.8.5	Budget for a stirred fluid in a region with interior sides	1091
48.9	Active tracers and dia-surface flow	1094
48.9.1	Adiabatic flow	1094
48.9.2	Diabatic processes generating dia-surface transport	1094
48.10	Green's function method for passive tracers	1096
48.10.1	Concerning time dependent domain boundaries	1096
48.10.2	Passive tracer boundary conditions	1096
48.10.3	Advection-diffusion initial-boundary value problem	1098
48.10.4	The Green's function and its adjoint	1098
48.10.5	Reciprocity relation	1099
48.10.6	Integral expression for the tracer concentration	1102
48.10.7	Properties of the solution	1103
48.10.8	Boundary propagator	1104
48.11	Exercises	1106

48.1 Loose threads

- Schematic for the boundary propagator.

48.2 Introduction

As derived in Section 16.1, the tracer equation takes on the general form

$$\rho \frac{DC}{Dt} = -\nabla \cdot \mathbf{J}(C), \quad (48.1)$$

where \mathbf{J} is a flux that embodies molecular diffusion as well as subgrid scale advection and subgrid scale diffusion (Chapter 50). Advective transport appears when transforming to an Eulerian or laboratory reference frame, in which case

$$\rho \frac{DC}{Dt} = \frac{\partial(\rho C)}{\partial t} + \nabla \cdot (\mathbf{v} \rho C), \quad (48.2)$$

with $\mathbf{v} \rho C$ the advective flux. In the absence of diffusion, advection renders a reversible stirring and stretching of fluid elements that generally increases the magnitude of concentration gradients. Advection does so while maintaining, for each fluid element, a fixed mass for all matter constituents and fixed specific entropy.¹ In contrast, diffusion affects an irreversible exchange, or mixing, of matter, thermodynamic, and mechanical properties between fluid elements. Correspondingly, diffusion reduces the magnitude of property gradients as it irreversibly exchanges properties between fluid elements.

48.3 Diffusion physics

The continuum approximation from Chapter 12 proposes that a macroscopic description of fluid motion does not require direct information about the motion of individual molecules. Nonetheless, random molecular motion and properties of the constituent molecules impact on fluid motion through the process of *molecular diffusion* of matter. Analogously, the random motion of fluid elements within a turbulent fluid give rise to *turbulent diffusive transport*.² In this section, we explore the basic physical nature of molecular and turbulent diffusion.

48.3.1 Diffusion of matter by random molecular motions

Consider a fluid comprised of a single matter constituent, such as a lake of pure H_2O . As discussed in Section 13.2, for a macroscopic description of this single-component fluid, a constant mass fluid element is identical to a constant mass material fluid parcel. That is, there is no mixing of matter since there is just a single matter component. Now place a dye tracer (Section 16.2) into a corner of the lake so that the lake is comprised of two material components (H_2O and dye). Even in the absence of ambient macroscopic fluid motion, the random motion of water and dye molecules produces an exchange of matter constituents between fluid elements. Consequently, the dye spreads outward from its initial position; i.e., it *diffuses* into the surrounding water.

We introduced the notion of matter exchange between fluid elements when discussing the tracer equation in Section 16.1. In the present context, matter exchange occurs through the random motion of molecules acting in the presence of a matter concentration gradient. Even though the continuum approximation has removed all explicit concern for details of molecular motion, we confront the underlying molecular nature of matter since molecular motions have a measurable impact on macroscopic fluid properties. This transport of matter by random molecular motions is known as *molecular diffusion*. A statistical description of molecular diffusion was first given by Einstein through his investigations of Brownian Motion ([Einstein, 1905](#)).

Diffusion of matter is a familiar process. For example, the odor from an open perfume bottle will spread throughout a room, even in the absence of macroscopic motion of air in the room. When the ambient macroscopic motion is zero, the spread of the perfume arises from random molecular motions whose properties depend on details of the molecules (e.g., their size, speed, inter-molecular forces). The time scale for molecular diffusion is generally much longer than the analogous *turbulent diffusion* that results if there is random motion in the macroscopic fluid, such as occurs by placing a fan next to the perfume bottle.

¹Recall from Chapter 21 that specific entropy remains materially constant on fluid elements in the absence of mixing.

²For our purposes, turbulence is characterized by a quasi-random motion of fluid elements.

48.3.2 Diffusion of matter by random turbulent motions

It is common for geophysical fluid systems to exhibit some form of turbulent motion. In these systems, the spread of matter by macroscopic turbulent motion is many times more efficient than the spread of matter from molecular motion. In such cases, we are justified in ignoring molecular diffusion since the efficiency of the turbulent diffusive transport is far greater than that from molecular diffusion.

Taylor (1921) described the statistical properties of turbulent diffusion, with many of his insights forming the basis for theories of how turbulent motion impacts on matter concentrations. In Taylor's theory, turbulent diffusion is not concerned with details of the molecular properties of the fluid. Rather, the properties of turbulent diffusion (e.g., the efficiency of the turbulent diffusion) depend just on the nature of the turbulent motion of fluid elements. In this way, turbulent diffusion as described by Taylor is a phenomena that sits fully in the realm of continuum mechanics, whereas molecular diffusion is a subject for the kinetic theory of gases. Correspondingly, each type of turbulent motion gives rise to a distinct form of turbulent diffusion. For example, in a geophysical context, turbulent diffusion associated with the breaking of internal gravity waves is distinct from turbulent diffusion by geostrophic eddies.

48.3.3 Fick's law for matter diffusion

Consider a fluid with a non-uniform matter concentration such as that drawn for a one-dimensional case in Figure 48.1. Random motion, due either to molecular motion or turbulent fluctuations, will transfer matter across an arbitrary point, line, or plane. Random motion preferentially moves matter from regions of high concentration to regions of low concentration, thus smoothing gradients. To a good approximation, the mass flux (mass per time per cross-sectional area) of matter is linearly proportional to the concentration gradient, and thus can be written in the form

$$\mathbf{J} = -\kappa \rho \nabla C. \quad (48.3)$$

In this equation, we introduced the positive proportionality factor $\kappa > 0$, known as the *kinematic diffusivity*, whereas the product $\kappa \rho$ is known as the *dynamic diffusivity*

$$\kappa \quad \text{kinematic diffusivity with SI units m s}^{-2} \quad (48.4)$$

$$\rho \kappa \quad \text{dynamic diffusivity with SI units kg m}^{-2} \text{ s}^{-2}. \quad (48.5)$$

The kinematic diffusivity has dimensions of squared length per time and it sets the efficiency or strength of the diffusion. The diffusive flux (48.3) is known as Fick's law of matter diffusion. It is the most common mathematical form used to represent the mixing of matter through diffusion. Note that the minus sign in the diffusive flux arises since the flux is directed down the concentration gradient.

The kinematic diffusivity has physical dimensions equal to the product of a length and a speed. For molecular diffusion, the kinematic diffusivity is proportional to the mean free path, L_{mfp} (see Section 12.3.3), and the root-mean-square molecular speed, v_{rms} (see Section 12.3.4). Each of these properties is a function of the molecules comprising the matter. For air, the mean free path is roughly 2×10^{-7} m and the RMS speed is 500 m s⁻¹, so that $L_{\text{mfp}} v_{\text{rms}} \approx 10^{-4} \text{ m}^2 \text{ s}^{-1}$. The precise value for the molecular diffusivity depends on the molecular properties of the matter diffusing through air. For turbulent diffusion, Prandtl suggested that we consider a characteristic length and velocity scale determined by properties of the turbulent flow. The turbulent length scale (also called the *mixing length*) is generally much larger than the molecular mean free path, whereas the

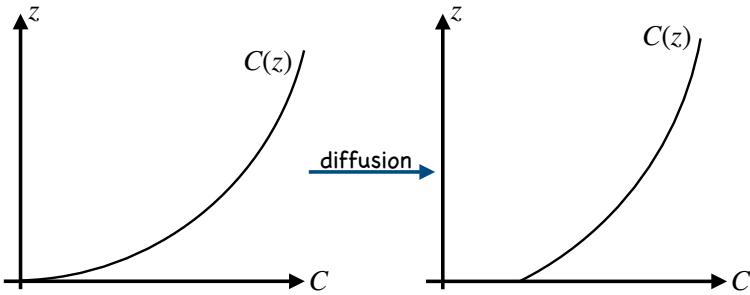


FIGURE 48.1: Shown here is a line graph illustrating the concentration, C , of a tracer drawn as a function of the space coordinate z , with the left panel showing the concentration at an earlier time than the right panel. Across any arbitrary point, transport of matter through random motions generally reduces the magnitude of the concentration gradient; i.e., the diffusive transport is down the concentration gradient. For example, where the concentration is relatively high, random motion mixes this high concentration with adjacent lower concentration, acting to lower the concentration in the originally high concentration region and raise the concentration in the originally low concentration region. In this particular example, $\partial C / \partial z > 0$, so that random fluid motions (either molecular or turbulent) lead to a diffusive flux directed in the $-\hat{z}$ direction; i.e., downward. This downward flux brings high concentration fluid into the lower/deeper regions and low concentration fluid into higher/shallower regions. The concentration is uniform in equilibrium, leading to a flat concentration profile.

turbulent velocity scale is much smaller than molecular speeds. Determination of these turbulent length and velocity scales is subject to large uncertainties and is the topic of much research.

In regions where the diffusive flux is not a constant, there will be a net transport of matter that leads to the reduction of the tracer concentration gradient. At a particular point in space, the concentration changes in time according to the convergence of the diffusive flux

$$\rho \frac{DC}{\partial t} = -\nabla \cdot \mathbf{J} = \nabla \cdot (\kappa \rho \nabla C). \quad (48.6)$$

That is, the concentration increases in regions where the diffusive flux, \mathbf{J} , converges, and decreases where the flux diverges. Expanding the divergence operator leads to

$$\rho \frac{DC}{\partial t} = \nabla(\kappa \rho) \cdot \nabla C + \rho \kappa \nabla^2 C. \quad (48.7)$$

The first term is nonzero in regions where the dynamic diffusivity, $\kappa \rho$, spatially varies. The second term is nonzero in regions where the curvature of the concentration is nonzero. Correspondingly, when the tracer concentration is uniform in space then both terms vanish, whereas the Laplacian term also vanishes when the concentration is linear in space.

48.3.4 Fourier's law for heat diffusion

In the same way that matter concentration gradients lead to diffusion by random motions, temperature gradients lead to diffusion of heat. The corresponding phenomenological relation is known as Fourier's law, with the diffusive flux given by

$$\mathbf{J} = -\gamma \rho \nabla T, \quad (48.8)$$

where $\gamma > 0$ is the temperature diffusivity. As for the matter diffusivity, the molecular thermal diffusivity can be expressed in terms of fundamental properties of the fluid, and it is different from the matter diffusivity. In general, matter diffuses by molecular processes slower than heat, so that the matter molecular diffusivity is smaller than the heat molecular diffusivity. In contrast, the turbulent thermal diffusivity is roughly the same as the matter diffusivity, since the turbulent diffusion of matter and heat are both mediated by the same turbulent fluctuations of fluid elements.

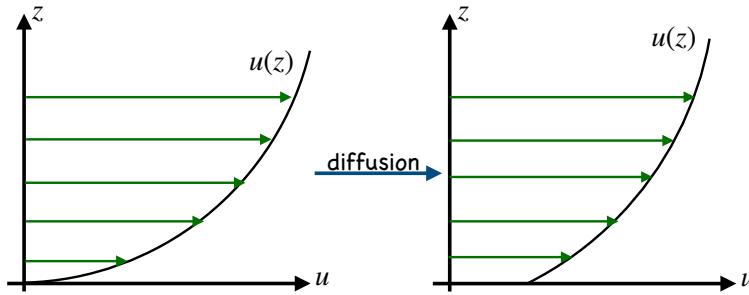


FIGURE 48.2: Shown here is a line graph illustrating the zonal velocity, u , as a function of the space coordinate z , with the left panel showing the velocity at an earlier time than the right panel. Across any arbitrary point, transport of momentum through random motions generally reduces the magnitude of the velocity gradient; i.e., the diffusive transport leads to a viscous stress that acts to reduce the velocity shear.

48.3.5 Newtonian frictional stress and momentum diffusion

In the same way that matter concentration and temperature gradients lead to diffusion by random motions, the momentum of fluid elements is exchanged through diffusion in the presence of viscosity. The corresponding phenomenological relation is known as Newton's law of viscous friction. As momentum is a vector, a general treatment of momentum transport through irreversible viscous processes involves a second order stress tensor and a fourth order viscosity tensor. For the specific case shown in Figure 48.2, with shear (i.e., nonzero velocity gradient) in a single direction, Newtonian frictional stress takes the form

$$\tau = \rho \mu \frac{\partial u}{\partial z}, \quad (48.9)$$

where $\mu > 0$ is the kinematic viscosity. Note the absence of a minus sign, in contrast to diffusive fluxes of scalars. The sign difference arises since it is the divergence of the stress tensor that leads to contact forces on the fluid, whereas it is the convergence of diffusive fluxes that leads to diffusion of matter and heat. We consider these general properties of the stress tensor when exploring the fluid dynamical equations in Chapter 18 and the nature of stress in Chapter 19.

For geophysical fluid mechanics, we are most generally interested in the molecular viscosity of water and air. Quite generally, the dynamic viscosity of water ($\rho \mu$) is about 10^2 times larger than that for air. But since the density of water is about 10^3 times larger than air, the kinematic viscosity of air is roughly 10 times greater than that of water.

The molecular kinematic viscosity can be expressed in terms of fundamental properties of the fluid, and it is different from the molecular matter diffusivity and molecular thermal diffusivity. For some turbulent processes, the turbulent viscosity is proportional to the scalar diffusivity. In general, the non-dimensional ratio of the viscosity to the diffusivity is known as the *Prandtl* number

$$\text{Pr} = \frac{\mu}{\kappa}. \quad (48.10)$$

Theories for the turbulent Prandtl number are largely empirical in nature, with first principles arguments elusive.

48.3.6 Further study

More thorough treatments of molecular diffusion for ideal gases can be found in books that describe the kinetic theory of gases, such as [Reif \(1965\)](#) and [Huang \(1987\)](#). The more terse treatment given in this section largely follows that from Section 1.5 of [Kundu et al. \(2016\)](#). A lucid treatment of Brownian motion in the context of turbulent diffusion is given by [Vallis \(2017\)](#).

48.4 Diffusion maths

We now explore various mathematical properties of the diffusion equation, here generalized to allow for distinct behavior in the different directions. Such distinctions are relevant especially in stratified fluids, where turbulent mixing across stratification surfaces is suppressed relative to turbulent mixing parallel to these surfaces (see Section 25.4). For this purpose, introduce the second order symmetric and positive definite diffusion tensor $\mathbb{K}_{mn} = \mathbb{K}_{nm}$. The resulting diffusive tracer flux takes the form

$$J_m = -\rho \mathbb{K}_{mn} \frac{\partial C}{\partial x^n}, \quad (48.11)$$

and the corresponding material diffusion equation is

$$\rho \frac{DC}{Dt} = -\nabla \cdot \mathbf{J} = \frac{\partial(\rho \mathbb{K}_{mn} \partial_n C)}{\partial x_m}. \quad (48.12)$$

48.4.1 Sample diffusion tensors

For the isotropic case of molecular diffusion considered in equation (48.3), the diffusion tensor takes on the form

$$\mathbb{K}_{mn} = \kappa \delta_{mn} \quad \text{isotropic diffusion.} \quad (48.13)$$

If we rotate the diffusive fluxes to be along surfaces of constant γ , then

$$\mathbb{K}_{mn} = \kappa (\delta_{mn} - \hat{\gamma}_m \hat{\gamma}_n) \quad \text{rotated diffusion,} \quad (48.14)$$

where

$$\hat{\gamma} = \frac{\nabla \gamma}{|\nabla \gamma|} \quad (48.15)$$

is the normal to the surface. The most common case in oceanography is to set γ equal to a measure of the buoyancy, in which case we have *neutral diffusion* (see Section 50.3).

48.4.2 Evolution of tracer concentration powers

For certain applications it is of interest to determine how diffusion acts on powers of the tracer concentration. For this purpose consider the material evolution of C^Γ , where $\Gamma \geq 1$

$$\rho \frac{DC^\Gamma}{Dt} = \Gamma C^{\Gamma-1} \rho \frac{DC}{Dt} \quad (48.16a)$$

$$= -\Gamma C^{\Gamma-1} \nabla \cdot \mathbf{J} \quad (48.16b)$$

$$= -\Gamma \nabla \cdot (C^{\Gamma-1} \mathbf{J}) + \Gamma (\Gamma - 1) C^{\Gamma-2} \nabla C \cdot \mathbf{J} \quad (48.16c)$$

$$= \partial_m \left[\rho \mathbb{K}_{mn} \frac{\partial C^\Gamma}{\partial x^n} \right] - \rho \Gamma (\Gamma - 1) C^{\Gamma-2} \frac{\partial C}{\partial x^n} \mathbb{K}_{mn} \frac{\partial C}{\partial x^m} \quad (48.16d)$$

$$= -\nabla \cdot \mathbf{J}(C^\Gamma) + \Gamma (\Gamma - 1) C^{\Gamma-2} \mathbf{J} \cdot \nabla C. \quad (48.16e)$$

The first term in equation (48.16e) is the convergence of the diffusive flux defined in terms of C^Γ , and it acts to diffuse C^Γ just like diffusion acts on C . The second term in equation (48.16e) is negative since the diffusion tensor is symmetric and positive-definite

$$\mathbf{J} \cdot \nabla C = -\rho \mathbb{K}_{mn} \frac{\partial C}{\partial x^m} \frac{\partial C}{\partial x^n} < 0. \quad (48.17)$$

That is, the diffusive flux corresponding to any positive definite and symmetric tensor is oriented down the tracer concentration gradient. Consequently, the second term in equation (48.16e) always acts to reduce the magnitude of C^Γ towards zero. We make use of this result in Section 48.4.3 when considering moments of tracer concentration.

48.4.3 Moments of tracer concentration

Consider the evolution of domain integrated tracer concentration and its powers, and focus on impacts just from diffusion. For that purpose, assume the boundaries are insulating so that $\mathbf{J} \cdot \hat{\mathbf{n}} = 0$ with $\hat{\mathbf{n}}$ the outward normal at the boundary. Also assume the total fluid mass in the domain remains fixed

$$m = \int \rho dV \iff \frac{dm}{dt} = 0. \quad (48.18)$$

We can thus treat the domain as material given that there is no exchange of mass or tracer across the boundaries.

Domain average tracer concentration

The domain averaged tracer concentration is given by

$$\bar{C} = \frac{\int C \rho dV}{m}, \quad (48.19)$$

and it follows that its time derivative vanishes since

$$m \frac{d\bar{C}}{dt} = \frac{d}{dt} \int C \rho dV = \int \frac{DC}{Dt} \rho dV = - \int \nabla \cdot \mathbf{J} dV = - \oint \mathbf{J} \cdot \hat{\mathbf{n}} d\mathcal{S} = 0, \quad (48.20)$$

where we set $\mathbf{J} \cdot \hat{\mathbf{n}} = 0$ to reach the last equality. Also note that we brought the time derivative inside the integral as a material derivative since the region is itself material, thus allowing us to make use of Reynold's transport theorem from Section 16.3.5. The result (48.20) follows since there is no change in the total mass of fluid nor is there any exchange of tracer across the boundaries. Hence, the domain averaged tracer concentration remains fixed in time.

Variance of tracer within the domain

The variance of the tracer concentration is defined by

$$\text{var}(C) \equiv \frac{\int (C - \bar{C})^2 \rho dV}{m} = \bar{C^2} - \bar{C}^2 \geq 0. \quad (48.21)$$

The tracer variance measures the deviation of the tracer concentration relative to the domain averaged concentration. Since the domain average remains fixed in time, the time change of the variance is given by

$$\frac{d[\text{var}(C)]}{dt} = \frac{d\bar{C^2}}{dt}. \quad (48.22)$$

Thus, it is common to refer to $\bar{C^2}$ as the tracer variance, though strictly speaking only their time derivatives are equal as per equation (48.22). Performing the time derivative, and again noting that the domain is material thus allowing us to use Reynolds transport theorem, renders

$$\frac{d\bar{C^2}}{dt} = \frac{d}{dt} \int C^2 \rho dV = 2 \int C \frac{DC}{Dt} \rho dV = -2 \int C \nabla \cdot \mathbf{J} dV = 2 \int \nabla C \cdot \mathbf{J} dV. \quad (48.23)$$

The final equality again made use of the assumed boundary condition $\mathbf{J} \cdot \hat{\mathbf{n}} = 0$. The time change in the tracer variance is thus determined by the integral of the projection of the tracer flux onto the tracer gradient. We already saw from equation (48.17) that diffusive fluxes are oriented down the tracer gradient. Consequently, diffusion of the tracer concentration results in a reduction in tracer variance

$$\frac{d[\text{var}(C)]}{dt} = \frac{d\overline{C^2}}{dt} \leq 0. \quad (48.24)$$

This result further supports our preconceived notions of diffusion as a process that removes differences (i.e., gradients) within the tracer field.

Evolution of arbitrary moments

Proceeding as before, and dropping boundary contributions, it is straightforward to show that the time derivative of an arbitrary tracer moment is given by

$$\frac{d\overline{C^\Gamma}}{dt} = \Gamma(\Gamma - 1) \int C^{\Gamma-2} \nabla C \cdot \mathbf{J} dV \leq 0. \quad (48.25)$$

For $\Gamma = 0$ we have an expression of mass conservation for the domain, whereas $\Gamma = 1$ is an expression of tracer conservation. The case of $\Gamma = 2$ yields the tracer variance result (48.24). The result for higher powers also holds. Hence, we conclude that the downgradient orientation of diffusive tracer fluxes acts to dissipate all powers of tracer concentration when integrated globally; i.e., all tracer moments are dissipated by diffusion.

48.4.4 Connecting tracer variance to the diffusion operator

We take a slight foray here into linear operator theory, in which we note that with natural boundary conditions (defined below) on the diffusive tracer flux, the diffusion operator is a linear self-adjoint operator.³ Consequently, the diffusion operator is related to a negative semi-definite functional (e.g., [Courant and Hilbert, 1953, 1962](#)). In the simplest case, the Laplacian operator $\nabla^2 C$ is equal to the functional derivative $\nabla^2 C = \delta \mathcal{F} / \delta C$, where

$$\mathcal{F} \equiv -\frac{1}{2} \int |\nabla C|^2 \rho d^3x \quad (48.26)$$

is the associated functional. In this subsection we prove this result for a general diffusion tensor K_{mn} acting on an arbitrary tracer concentration, C , so long as the diffusion tensor is not a function of the tracer concentration. As detailed by [Griffies et al. \(1998\)](#) and Chapter 16 of [Griffies \(2004\)](#), the connection between the diffusion operator and the functional derivative of the diffusion dissipation provides a useful framework for deriving numerical discretizations of the diffusion operator.

For this subsection it is useful to write

$$d^3x = dV \quad (48.27)$$

for the volume element. The reason will become apparent when reaching equation (48.35).

³We discussed the self-adjoint nature of the diffusion operator in Section 4.5.4 in our study of the Green's function for the diffusion equation.

Derivative of the diffusion dissipation functional

We introduce the *diffusion dissipation functional*

$$\mathcal{F} = \int \mathcal{L} d^3x \quad (48.28)$$

where the integrand is the negative semi-definite quadratic form

$$2\mathcal{L} = \mathbf{J} \cdot \nabla C = -\rho \partial_m C \mathbb{K}_{mn} \partial_n C \leq 0. \quad (48.29)$$

Our goal is to relate the diffusion operator, given by the convergence of the diffusion flux, $\mathcal{R} = -\nabla \cdot \mathbf{J}$, to the functional derivative of \mathcal{F} . To compute the functional derivative requires us to insert variations to the tracer field δC into the dissipation functional

$$\delta\mathcal{F} = \int \left[\delta C \frac{\delta\mathcal{L}}{\delta C} + \delta(\partial_m C) \frac{\delta\mathcal{L}}{\delta(\partial_m C)} \right] d^3x. \quad (48.30)$$

Note that functional variations are small perturbations to the form of the function, in which case

$$C \rightarrow C + \delta C \quad \text{with} \quad |\delta C| \ll |C|. \quad (48.31)$$

Integration by parts on the second term leads to

$$\delta\mathcal{F} = \int \left[\delta C \frac{\delta\mathcal{L}}{\delta C} + \partial_m \left(\delta C \frac{\delta\mathcal{L}}{\delta(\partial_m C)} \right) - \delta C \partial_m \left(\frac{\delta\mathcal{L}}{\delta(\partial_m C)} \right) \right] d^3x. \quad (48.32)$$

The middle term is a total derivative that integrates to a boundary contribution and the associated *natural boundary condition*

$$\hat{\mathbf{n}} \cdot \frac{\delta\mathcal{L}}{\delta \nabla C} = \hat{\mathbf{n}} \cdot \mathbf{J} = \text{boundary flux}, \quad (48.33)$$

with $\hat{\mathbf{n}}$ the boundary outward normal. This natural boundary condition is the familiar *Neumann boundary condition* from Chapter 4.

To focus on the connection between the diffusion operator and the diffusion dissipation functional, we ignore boundary fluxes so that the functional variation is given by

$$\delta\mathcal{F} = \int \delta C \left[\frac{\delta\mathcal{L}}{\delta C} - \partial_m \left(\frac{\delta\mathcal{L}}{\delta(\partial_m C)} \right) \right] d^3x. \quad (48.34)$$

Consequently, the functional derivative is given by

$$(d^3y)^{-1} \frac{\delta\mathcal{F}}{\delta C(\mathbf{y})} = \frac{\delta\mathcal{L}}{\delta C} - \partial_m \left[\frac{\delta\mathcal{L}}{\delta(\partial_m C)} \right], \quad (48.35)$$

where d^3y is the volume element at the field point \mathbf{y} . To reach the last step required the identity

$$\frac{\delta C(\mathbf{x})}{\delta C(\mathbf{y})} = d^3y \delta(\mathbf{x} - \mathbf{y}), \quad (48.36)$$

where $\delta(\mathbf{x} - \mathbf{y})$ is the Dirac delta function satisfying

$$\int \delta(\mathbf{x} - \mathbf{y}) d^3y = 1, \quad (48.37)$$

so long as the integration range includes the singular point $\mathbf{x} = \mathbf{y}$. Note that the delta function has dimensions of inverse volume, which necessitates the appearance of the volume factor, d^3y , on the right hand side of equation (48.36). The reader should be mindful that many treatments in mathematics texts ignore this extra volume factor. However, for physical applications it is important to maintain dimensional consistency, thus requiring us to retain the volume factor.

Connection to the diffusion operator

Reintroducing the specific form of the diffusion integrand $2\mathcal{L} = -\rho \partial_m C \mathbb{K}_{mn} \partial_n C$ leads to

$$(d^3y)^{-1} \frac{\delta \mathcal{F}}{\delta C(\mathbf{y})} = -\partial_m \left[\frac{\delta \mathcal{L}}{\delta (\partial_m C)} \right] = \partial_m (\rho \mathbb{K}_{mn} \partial_n C). \quad (48.38)$$

The second equality identifies the diffusion operator, thus revealing the connection between the dissipation functional, the diffusion fluxes, and the diffusion operator

$$(d^3y)^{-1} \frac{\delta \mathcal{F}}{\delta C(\mathbf{y})} = -\nabla \cdot \mathbf{J}(\mathbf{y}) = \mathcal{R}. \quad (48.39)$$

Why did we assume \mathbb{K}_{mn} is independent of C ?

There are many geophysical applications in which the diffusion tensor is a function of the tracer concentration, in which case the diffusion equation is no longer a linear differential equation. For example, the neutral diffusion of Section 50.3 makes use of a diffusion tensor that is a function of temperature and salinity gradients. In this case the functional derivative in terms of temperature or salinity appearing in equation (48.38) becomes

$$2 \frac{\delta \mathcal{L}}{\delta (\partial_m C)} = -2\rho \mathbb{K}_{mn} \partial_n C - \rho \partial_m C \partial_n C \frac{\delta \mathbb{K}_{mn}}{\delta (\partial_m C)}. \quad (48.40)$$

The specific form of the term $\delta \mathbb{K}_{mn}/\delta(\partial_m C)$ depends on details of the diffusion tensor. Hence, the general results derived above for the linear diffusion equation no longer hold for this nonlinear diffusion equation. We have more to say about nonlinear advection-diffusion in Section 48.9.

48.5 Advection physics

A *perfect* or *ideal* fluid is comprised of material fluid elements whose matter content and thermodynamic properties remain fixed. From the discussion of molecular diffusion in Section 48.3, we know that a perfect fluid can at most consist of a single matter constituent and uniform thermodynamic properties. The reason is that in the presence of multiple constituents with non-uniform concentrations, molecular motions irreversibly exchange matter among fluid elements. This matter exchange, or mixing, breaks the assumption of a perfect fluid. Nonetheless, we find many occasions to ignore molecular diffusion when focusing on macroscopic motions of the continuum fluid. Such is the case when considering the advection equation in the absence of any mixing.

48.5.1 The advection equation

In the absence of mixing or other irreversible processes, the matter content of a fluid element remains fixed as the element moves with the fluid. Since the total mass of the element is also constant, then the tracer concentration remains constant and thus satisfies the reversible (source-free) *advection equation*

$$\frac{DC}{Dt} = \frac{\partial C}{\partial t} + \mathbf{v} \cdot \nabla C = 0. \quad (48.41)$$

The first equality relates the material time derivative to the Eulerian time derivative plus advective transport (see Section 13.5), with \mathbf{v} the velocity of a fluid element. We can convert the *material*

form of the advection equation (48.41) into its flux-form by combining with the mass continuity equation (15.6)

$$\frac{\partial \rho}{\partial t} + \nabla \cdot (\rho \mathbf{v}) = 0, \quad (48.42)$$

which yields

$$\frac{\partial(\rho C)}{\partial t} + \nabla \cdot (\rho C \mathbf{v}) = 0. \quad (48.43)$$

Again, the material form of the advection equation is the trivial statement that tracer concentration remains constant on a fluid element in the absence of sources or mixing. Hence, a general solution to the advection equation is given by

$$C(\mathbf{x}, t) = C[\mathbf{X}(0)], \quad (48.44)$$

where $\mathbf{X}(0)$ is the initial position of a fluid element that is at the position \mathbf{x} at time t . If we know the trajectories for all fluid elements and their initial tracer concentration, we know the tracer concentration for all space and time. For those cases where trajectories are unknown, it is useful to make use of the Eulerian form of the advection equation in order to deduce the evolution of tracer.

48.5.2 Eulerian time tendencies from advection

At a point in the fluid, the advection equation (48.41) leads to the Eulerian time tendency for tracer concentration

$$\frac{\partial C}{\partial t} = -\mathbf{v} \cdot \nabla C. \quad (48.45)$$

Geometrically, the tendency arises from the projection of the fluid velocity onto the normal to concentration iso-surfaces. The concentration remains fixed in time (steady) at points where the velocity is parallel to concentration iso-surfaces. From the flux-form advection equation (48.43), the density-weighted tracer concentration (the tracer mass per volume) has an Eulerian time tendency given by the convergence of the advective flux

$$\frac{\partial(\rho C)}{\partial t} = -\nabla \cdot (\rho C \mathbf{v}). \quad (48.46)$$

The tendency vanishes at a point if there is no convergence of tracer mass towards the point.

48.5.3 Impermeability property of tracer isosurfaces

We now offer a geometric interpretation of the advection equation

$$\frac{\partial C}{\partial t} + \mathbf{v} \cdot \nabla C = 0, \quad (48.47)$$

following the discussion of dia-surface transport in Section 44.4. For this purpose, introduce the unit normal on a tracer isosurface

$$\hat{\mathbf{n}} = \frac{\nabla C}{|\nabla C|} \quad (48.48)$$

and the normal projection for the velocity of a point on that surface

$$\mathbf{v}^{(b)} \cdot \hat{\mathbf{n}} = -\frac{\partial_t C}{|\nabla C|}. \quad (48.49)$$

The advection equation (48.47) thus can be written as an impermeability condition for a tracer isosurface

$$\rho(\mathbf{v} - \mathbf{v}^{(b)}) \cdot \hat{\mathbf{n}} = 0 \quad \text{on } C \text{ isosurfaces.} \quad (48.50)$$

We encountered this condition in Section 15.4.2 when studying the kinematics of a moving material surface. Hence, in the absence of mixing, tracer isosurfaces are indeed material surfaces since they allow no fluid elements, moving with the fluid velocity \mathbf{v} , to cross them. This is an important kinematic result that will be extended in Section 48.8.5 to include effects from an eddy induced velocity.

48.6 Advection maths

We now explore various mathematical properties of the advection equation. For that purpose, recall the mass continuity equation (48.42) and flux-form tracer advection equation (48.46)

$$\frac{\partial \rho}{\partial t} + \nabla \cdot (\rho \mathbf{v}) = 0 \quad (48.51a)$$

$$\frac{\partial(\rho C)}{\partial t} + \nabla \cdot (\rho C \mathbf{v}) = 0. \quad (48.51b)$$

These equations are manifestly compatible in that the tracer equation (48.51b) reduces to the continuity equation (48.51a) if the tracer concentration is spatially uniform (see Section 16.1.4 for more discussion of compatibility).

48.6.1 Material constancy of C^Γ

A trivial consequence of the material constancy of tracer concentration is that C^Γ is also materially constant, for Γ an arbitrary number. We show this property mathematically by noting that the chain rule holds for a material time derivative, so that

$$\frac{DC^\Gamma}{Dt} = \Gamma C^{\Gamma-1} \frac{DC}{Dt} = 0. \quad (48.52)$$

Likewise, making use of the Eulerian form yields

$$\frac{\partial C^\Gamma}{\partial t} + \mathbf{v} \cdot \nabla C^\Gamma = \Gamma C^{\Gamma-1} \left[\frac{\partial C}{\partial t} + \mathbf{v} \cdot \nabla C \right] = 0. \quad (48.53)$$

We conclude that advection, in the absence of diffusion, serves to reversibly transport the tracer concentration without altering any of its powers. Correspondingly, all tracer moments are untouched by advection, which contrasts to the case of diffusion considered in Section 48.4.3.

48.6.2 Mass transport

The mass density time tendency

$$\frac{\partial \rho}{\partial t} = -\nabla \cdot (\mathbf{v} \rho) \quad (48.54)$$

remains unchanged if the advective mass flux, $\rho \mathbf{v}$ (dimensions of mass per time per area), is modified by the addition of a total curl

$$\rho \mathbf{v} \rightarrow \rho \mathbf{v}^\dagger = \rho \mathbf{v} + \nabla \wedge (\rho \Psi^*). \quad (48.55)$$

As in Section 17.5.1, the arbitrariness manifest in equation (48.55) is known as a *gauge symmetry*. The additional mass flux, $\nabla \wedge (\rho \Psi^*)$, leads to no accumulation of mass at a point since it has zero divergence. In the Boussinesq case with ρ set to a constant ρ_0 , the divergent-free velocity $\nabla \wedge \Psi^*$ leads to zero accumulation of volume at a point.

The non-divergent mass flux

$$\rho \mathbf{v}^* \equiv \nabla \wedge (\rho \Psi^*) \quad (48.56)$$

often arises when we decompose the mass flux into a mean and non-divergent eddy fluctuations. In that context, we make use of the following terminology:

$$\mathbf{v} = \text{Eulerian mean velocity} \quad (48.57a)$$

$$\rho \mathbf{v} = \text{Eulerian mean mass flux} \quad (48.57b)$$

$$\mathbf{v}^* = \text{eddy-induced velocity} \quad (48.57c)$$

$$\rho \Psi^* = \text{eddy-induced mass streamfunction} \quad (48.57d)$$

$$\rho \mathbf{v}^* = \nabla \wedge (\rho \Psi^*) = \text{eddy-induced mass flux} \quad (48.57e)$$

$$\mathbf{v}^\dagger = \mathbf{v} + \mathbf{v}^* = \text{residual mean velocity} \quad (48.57f)$$

$$\rho \mathbf{v}^\dagger = \rho(\mathbf{v} + \mathbf{v}^*) = \text{residual mean mass flux.} \quad (48.57g)$$

The name “residual mean” is motivated since the sum $\mathbf{v} + \mathbf{v}^*$ is often smaller than either term individually. That is, the eddy contribution often compensates for the mean, with sum of the mean and eddy representing a residual. We study particular forms of the eddy induced velocity in Chapter 50.

48.6.3 Advection tracer fluxes and skew tracer fluxes

Following from the previous discussion, we consider the advection equation with the advective tracer transport determined by the residual mean velocity

$$\frac{\partial(\rho C)}{\partial t} + \nabla \cdot (\rho C \mathbf{v}^\dagger) = 0. \quad (48.58)$$

Given the form (48.56) for the eddy mass flux $\rho \mathbf{v}^*$, we can write the advective tracer flux as

$$\rho C \mathbf{v}^\dagger = C(\rho \mathbf{v} + \rho \mathbf{v}^*) \quad (48.59a)$$

$$= C \rho \mathbf{v} + C \nabla \wedge (\rho \Psi^*) \quad (48.59b)$$

$$= C \rho \mathbf{v} + \nabla \wedge (C \rho \Psi^*) - \nabla C \wedge \rho \Psi^*. \quad (48.59c)$$

It is the divergence of the tracer flux that determines the time tendency, in which case the total curl plays no role

$$\nabla \cdot (\rho C \mathbf{v}^\dagger) = \nabla \cdot (\rho C \mathbf{v} + \rho C \mathbf{v}^*) \quad (48.60a)$$

$$= \nabla \cdot (\rho C \mathbf{v} - \nabla C \wedge \rho \Psi^*). \quad (48.60b)$$

That is, the divergence of the advective mass flux equals to the divergence of the skew tracer flux

$$\underbrace{\nabla \cdot (\rho C \mathbf{v}^*)}_{\text{advective flux divergence}} = \underbrace{\nabla \cdot (-\nabla C \wedge \rho \Psi^*)}_{\text{skew flux divergence}} \quad (48.61)$$

since the advective flux and skew flux differ by a rotational flux

$$\mathbf{J}^{\text{adv}} = \mathbf{J}^{\text{skew}} + \mathbf{J}^{\text{rot}} \quad (48.62)$$

where

$$\mathbf{J}^{\text{adv}} = C \rho \mathbf{v}^* \quad \text{and} \quad \mathbf{J}^{\text{skew}} = -\nabla C \wedge \rho \Psi^* \quad \text{and} \quad \mathbf{J}^{\text{rot}} = \nabla \wedge (\rho C \Psi^*). \quad (48.63)$$

Notably, the skew tracer flux is neither upgradient nor downgradient. Rather, it is oriented parallel to iso-surfaces of tracer concentration

$$\nabla C \cdot \mathbf{J}^{\text{skew}} = \nabla C \cdot (-\nabla C \wedge \rho \Psi^*) = 0. \quad (48.64)$$

This orientation serves as motivation for the name “skew.” Figure 48.3 provides a schematic of the skew tracer fluxes.

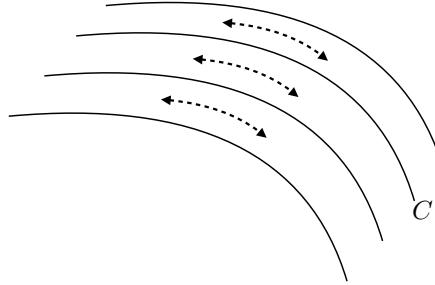


FIGURE 48.3: Skew fluxes (dashed lines with arrows) for a tracer C are oriented parallel to lines of constant tracer concentration (tracer isolines are the solid lines).

48.6.4 Skew diffusion

Introducing tensor labels brings the skew tracer flux into the form

$$J_m^{\text{skew}} = -(\nabla C \wedge \rho \Psi^*)_m \quad (48.65a)$$

$$= -\epsilon_{mnp} \frac{\partial C}{\partial x^n} \rho \Psi_p^* \quad (48.65b)$$

$$= -\rho \mathbb{A}_{mn} \frac{\partial C}{\partial x^n}, \quad (48.65c)$$

where we defined the anti-symmetric *skew diffusion* tensor

$$\mathbb{A}_{mn} = \epsilon_{mnp} \Psi_p^* = \begin{bmatrix} 0 & \Psi_3^* & -\Psi_2^* \\ -\Psi_3^* & 0 & \Psi_1^* \\ \Psi_2^* & -\Psi_1^* & 0 \end{bmatrix}. \quad (48.66)$$

We thus conclude that advection by a non-divergent mass flux is equivalent to skew-diffusion through the action of an anti-symmetric tensor.

Although leading to the same stirring operator, skew and advective fluxes possess rather complementary properties as listed here.

- **DERIVATIVE OPERATOR:** The skew flux is proportional to the vector streamfunction and the gradient of the tracer, whereas the advective flux is related to the curl of the streamfunction and the value of the tracer concentration. That is, the fluxes in effect swap the placement of the derivative operator. Correspondingly, the advective flux vanishes only if the velocity vanishes, whereas the skew flux vanishes when the tracer gradient vanishes (as for a diffusive flux).

- FLUX ORIENTATION: The orientation of the advective flux is determined by the velocity field, which is oriented according to trajectories of fluid particles. This orientation is the same regardless of the tracer. In contrast, a skew tracer flux is directed along lines of constant tracer; i.e., it is neither upgradient nor downgradient. Hence, orientation of the skew flux is directly tied to the tracer field, with each tracer yielding a generally distinct flux orientation. The very distinct orientations of the advective and skew fluxes is the origin of much confusion in regards to these fluxes. We explore many features of these geometric distinctions in Section 49.3 when studying eddy induced tracer fluxes.
- MATERIAL FLUX: Fluid elements carry a particular amount of trace matter so that an advective flux of a material tracer measures the passage of matter across an area per unit time (dimensions of mass per area per time), whereas a skew flux is not interpreted as the passage of matter across an area per time. This distinction is particularly important when interpreting boundary conditions discussed in Section 48.7.3.

In Section 48.7 we pursue the above points to further detail the dual relation between advection and skewson.

48.6.5 A comment about skew fluxes and Lagrangian kinematics

The advective tracer flux and skew tracer flux are very distinct vectors and we further explore the distinction in Sections 48.7 and 49.3. As detailed in each of those sections, it is a matter of convenience how one chooses to formulate the Eulerian tracer equation since the advective flux and skew flux lead to the same tracer evolution. Furthermore, the choice to formulate the tracer equation in terms of a skew flux in no way eliminates the Lagrangian perspective. That is, fluid particles, or fluid elements, still move through the fluid and transport tracer as part of this motion. The Lagrangian formulation of tracer evolution is naturally connected, via a transformation of reference frames, to the Eulerian advection equation. Even so, we do not somehow eliminate fluid particle motion when choosing to work with skew tracer fluxes. Instead, we merely choose to formulate the tracer equation in terms of the vector streamfunction rather than the velocity.

48.6.6 Further reading

The uses of residual-mean transport are many and varied in the ocean and atmospheric literature. [Vallis \(2017\)](#) offers a thorough and pedagogical treatment. Skew diffusion is treated in [Moffatt \(1983\)](#), in which he raises the connection to fluids with rotation and/or magnetic fields. [Middleton and Loder \(1989\)](#) applied these ideas to ocean gravity waves, tides, and Rossby waves. [Griffies \(1998\)](#) applied these ideas to the methods used for parameterizing tracer transport from ocean mesoscale eddies.

48.7 Advection and skewson

We introduced skew diffusion in Section 48.6.4 and will again encounter it in Chapterd 49 and 50. Following the terminology of Section 9.2 of [Griffies \(2004\)](#), we refer to *skewson* as any process that leads to tracer transport via skew fluxes, with skew diffusion a particular example. There are occasions where it is conceptually and operationally more convenient to use advective fluxes, such as when considering the transport of tracers by the flow field explicitly resolved by a numerical simulation. In contrast, skew fluxes are sometimes more convenient for certain subgrid scale eddy

parameterizations, such as the one discussed in Section 50.1. We here consider facets of advection and skewson for those interested in diving deeper into the details.

48.7.1 Choosing a gauge

Consider an arbitrary divergent-free mass transport

$$\nabla \cdot (\rho \mathbf{v}^*) = 0, \quad (48.67)$$

where the divergent-free constraint is satisfied by introducing a vector streamfunction

$$\rho \mathbf{v}^* = \nabla \wedge (\rho \Psi^*). \quad (48.68)$$

The streamfunction is arbitrary up to a gauge transformation

$$\rho \Psi' = \rho \Psi^* + \nabla(\rho \Lambda), \quad (48.69)$$

where Λ is a gauge function.

Changes to the skew flux under a gauge transformation

Although the velocity is invariant up to an arbitrary gauge function, the skew flux, $\mathbf{J}^{\text{skew}} = -\nabla C \wedge \rho \Psi^*$, changes. Nonetheless, the divergence of the skew flux is invariant, as we see by noting that

$$\nabla C \wedge [\rho \Psi^* + \nabla(\rho \Lambda)] = \nabla C \wedge (\rho \Psi^*) + \nabla \wedge [C \nabla(\rho \Lambda)]. \quad (48.70)$$

and since $\nabla \cdot \nabla \wedge [C \nabla(\rho \Lambda)] = 0$, the flux divergence, $\nabla \cdot \mathbf{J}^{\text{skew}}$, remains unchanged.

Coulomb gauge

We have some freedom in specifying the gauge function. One choice is to set $\Lambda = 0$. However, there are occasions in which it is useful to set the gauge function in a manner to cancel unwanted terms. A gauge commonly used in electrodynamics (e.g., Jackson (1975)) is the *Coulomb gauge*, in which

$$\nabla \cdot (\rho \Psi^*) = 0 \quad \text{Coulomb gauge.} \quad (48.71)$$

Making use of the curl identity (2.38c) leads to the Poisson equation for the vector potential

$$\nabla^2(\rho \Psi^*) = -\nabla \wedge (\rho \mathbf{v}^*). \quad (48.72)$$

This equation has a free-space Green's function given by the Coulomb-Ampere expression

$$\rho(\mathbf{x}, t) \Psi^*(\mathbf{x}, t) = \nabla \wedge \int \frac{\rho(\mathbf{x}', t) \mathbf{v}^*(\mathbf{x}', t)}{4\pi |\mathbf{x} - \mathbf{x}'|} dV', \quad (48.73)$$

where dV' is the volume element for integration over the test points, \mathbf{x}' . Although common in electrodynamics, we know of no geophysical fluid application making use of the Coulomb gauge.

48.7.2 Vertical gauge

As introduced in Section 17.5.2, a gauge commonly used for eddy parameterizations (Section 50.1) sets to zero one of the three components of the vector streamfunction. This gauge choice is available since there are only two independent functional degrees of freedom available from a divergence-free mass transport field. A common choice is the *vertical gauge* in which

$$\Psi_3^* = 0 \quad \text{vertical gauge.} \quad (48.74)$$

Let us see how we can generally make this gauge choice. Consider a vector streamfunction Φ that has all three components nonzero. Now consider the alternative streamfunction $\rho \Psi^* = \rho \Phi + \nabla(\rho \Lambda)$, with $\nabla(\rho \Lambda) = -\hat{z} \rho \Phi_3$. This choice in turn means that the third component of Ψ^* is zero.

To further specify the vertical gauge we invert the relations

$$\rho u^* = -\partial_z(\rho \Psi_2^*) \quad \text{and} \quad \rho v^* = \partial_z(\rho \Psi_1^*) \quad \text{and} \quad \rho w^* = \partial_x(\rho \Psi_2^*) - \partial_y(\rho \Psi_1^*) \quad (48.75)$$

to render the vector streamfunction

$$\rho \Psi^* = \hat{z} \wedge \int_{-H}^z \rho \mathbf{u}^* dz' = \hat{z} \wedge \underline{\mathbf{U}}^{(*\rho)} \quad (48.76)$$

where

$$\underline{\mathbf{U}}^{(*\rho)}(z) = \int_{-H}^z \rho \mathbf{u}^* dz' \quad (48.77)$$

is the horizontal mass transport associated with \mathbf{u}^* passing between the bottom and a depth $z \geq -H$. The anti-symmetric stirring tensor for the vertical gauge is given by

$$\rho \mathbb{A}_{mn} = \begin{pmatrix} 0 & 0 & \underline{U}^{(*\rho)} \\ 0 & 0 & \underline{V}^{(*\rho)} \\ -\underline{U}^{(*\rho)} & -\underline{V}^{(*\rho)} & 0 \end{pmatrix}, \quad (48.78)$$

and the corresponding skew, rotational, and advective fluxes are

$$\mathbf{J}^{\text{skew}} = -\underline{\mathbf{U}}^{(*\rho)} \partial_z C + \hat{z} \underline{\mathbf{U}}^{(*\rho)} \cdot \nabla_z C \quad (48.79a)$$

$$\mathbf{J}^{\text{rot}} = \partial_z(C \underline{\mathbf{U}}^{(*\rho)}) - \hat{z} \nabla_z \cdot (C \underline{\mathbf{U}}^{(*\rho)}) \quad (48.79b)$$

$$\mathbf{J}^{\text{adv}} = C (\partial_z \underline{\mathbf{U}}^{(*\rho)}) - \hat{z} C \nabla_z \cdot \underline{\mathbf{U}}^{(*\rho)}. \quad (48.79c)$$

Note that the identity $\mathbf{J}^{\text{adv}} = \mathbf{J}^{\text{skew}} + \mathbf{J}^{\text{rot}}$ is maintained by these expressions. The horizontal components to the skew flux vanish when the tracer is uniform in the vertical, and the vertical skew flux vanishes with a horizontally uniform tracer field. These properties manifest the skewed nature of the fluxes.

48.7.3 Boundary conditions

We assume that all external domain boundaries are material in regards to the velocity \mathbf{v}^* . Furthermore, even for moving domain boundaries, we assume that the suite of kinematic boundary conditions is based on the barycentric velocity \mathbf{v} (see Section 15.4), so that \mathbf{v}^* satisfies the no-normal flow condition even on moving boundaries

$$\hat{\mathbf{n}} \cdot \mathbf{v}^* = 0 \quad \text{external domain boundaries.} \quad (48.80)$$

As we discuss in Section 48.8.1, this boundary condition is required for the eddy-induced velocity to have zero impact on the total mass of an arbitrary tracer within the full fluid domain.

Correspondingly, the advective tracer flux also satisfies a no-normal boundary condition on all external boundaries

$$\hat{\mathbf{n}} \cdot \mathbf{J}^{\text{adv}} = \hat{\mathbf{n}} \cdot \mathbf{v}^* \rho C = 0. \quad (48.81)$$

The corresponding boundary condition for the skew flux is found by inserting the relation (48.62) into the advective flux boundary condition (48.81) to render

$$\hat{\mathbf{n}} \cdot \mathbf{J}^{\text{adv}} = \hat{\mathbf{n}} \cdot [\mathbf{J}^{\text{skew}} + \mathbf{J}^{\text{rot}}] = 0. \quad (48.82)$$

Hence, the skew flux generally has a non-zero normal component at the solid boundaries as determined by the rotational flux

$$\hat{\mathbf{n}} \cdot \mathbf{J}^{\text{skew}} = -\hat{\mathbf{n}} \cdot \mathbf{J}^{\text{rot}}. \quad (48.83)$$

Even so, there might be occasions in which $\hat{\mathbf{n}} \cdot \mathbf{J}^{\text{skew}} = 0$, which is ensured so long as

$$(-\nabla C \wedge \rho \Psi^*) \cdot \hat{\mathbf{n}} = -(\rho \Psi^* \wedge \hat{\mathbf{n}}) \cdot \nabla C = 0. \quad (48.84)$$

A sufficient condition is to have $\Psi^* \wedge \hat{\mathbf{n}} = 0$, in which case the vector streamfunction is parallel to the boundary normal. An alternative sufficient condition is to have the streamfunction vanish at the boundary. Further details for boundary conditions depend on physical properties of the velocity \mathbf{v}^* . We discuss one example in Section 50.1 as prescribed by the [Gent et al. \(1995\)](#) mesoscale eddy parameterization.

48.8 Finite volume budgets with eddy velocities

In this section we examine how an eddy induced velocity modifies the budgets for fluid mass and tracer mass in finite domains. For this purpose, write the local/differential mass and tracer budgets in the form

$$\frac{\partial \rho}{\partial t} + \nabla \cdot (\rho \mathbf{v}^\dagger) = 0 \quad (48.85a)$$

$$\frac{\partial(\rho C)}{\partial t} + \nabla \cdot (\rho \mathbf{v}^\dagger C + \mathbf{J}^{\text{diff}}) = 0, \quad (48.85b)$$

where (see Section 48.7)

$$\mathbf{v}^\dagger = \mathbf{v} + \mathbf{v}^* \quad \text{and} \quad \nabla \cdot (\rho \mathbf{v}^*) = 0, \quad (48.86)$$

and where \mathbf{J}^{diff} is a subgrid scale flux encompassing all processes, such as diffusion and boundary conditions, that are not represented by an eddy-induced advection. Given that $\nabla \cdot (\rho \mathbf{v}^*) = 0$, the mass budget (48.85a) can we written in the equivalent manners

$$\frac{\partial \rho}{\partial t} + \nabla \cdot (\rho \mathbf{v}^\dagger) = \frac{\partial \rho}{\partial t} + \nabla \cdot (\rho \mathbf{v}) = 0. \quad (48.87)$$

That is, the eddy-induced velocity does not lead to any local sources of fluid mass. This property will prove to be very important in the budget analyses of this section.

As shown in the following, the finite volume budgets for fluid mass and tracer mass also make use of the residual mean velocity, \mathbf{v}^\dagger . That result is not surprising, since the finite volume budgets are consistent with the differential budgets (48.85a) and (48.85b). Nonetheless, it is useful to expose the details as they appear in many budget analysis applications, such as the water mass and tracer mass analysis of Chapter 52. We furthermore explore how the budgets for tracer mass appear when formulated using advective fluxes versus skew fluxes. As we will see, the finite volume budgets are consistent across the variety of formulations only if the eddy velocity and eddy vector streamfunction satisfy specific boundary conditions as discussed in Section 48.7.3.

48.8.1 Advection formulation

Making use of the tracer equation (48.85b) in the Leibniz-Reynolds transport theorem (16.37) renders the finite volume tracer mass budget for an arbitrary domain, \mathcal{R}

$$\frac{d}{dt} \left[\int_{\mathcal{R}} \rho C dV \right] = - \oint_{\partial\mathcal{R}} \left[\rho C (\mathbf{v}^\dagger - \mathbf{v}^{(b)}) + \mathbf{J}^{\text{diff}} \right] \cdot \hat{\mathbf{n}} d\mathcal{S}, \quad (48.88)$$

where $\mathbf{v}^{(b)}$ is the velocity of a point on the domain boundary. Appearance of the residual mean velocity, \mathbf{v}^\dagger , in the finite volume budget (48.88) follows from its appearance in the local tracer budget (48.85b). We thus see that the eddy-induced velocity impacts on the tracer mass budget for an arbitrary domain. However, its impacts disappear when integrating over a closed or periodic fluid domain so long as

$$\mathbf{v}^* \cdot \hat{\mathbf{n}} = 0 \quad \text{on external fluid boundaries.} \quad (48.89)$$

We already encountered this boundary condition in Section 48.7.3. It holds on all boundaries, including those such as the ocean free surface that are time dependent and/or permeable. It is required if we assume the eddy-induced velocity does not modify the mass of any tracer in the full fluid domain. That assumption is generally made for eddy-induced velocities such as those associated with mesoscale and submesoscale eddies in the ocean (see Section 50.2).

Setting the tracer concentration to a constant in equation (48.88) leads to the fluid mass budget

$$\frac{d}{dt} \left[\int_{\mathcal{R}} \rho dV \right] = - \oint_{\partial\mathcal{R}} \rho (\mathbf{v}^\dagger - \mathbf{v}^{(b)}) \cdot \hat{\mathbf{n}} d\mathcal{S}, \quad (48.90)$$

where we set the diffusive tracer flux, \mathbf{J}^{diff} , to zero since there is no diffusion of fluid mass between fluid elements (Section 16.1). As for the differential form (48.85a), the mass budget for any domain is not changed by the eddy-induced velocity since

$$\nabla \cdot (\rho \mathbf{v}^*) = 0 \implies \oint_{\partial\mathcal{R}} \rho \mathbf{v}^* \cdot \hat{\mathbf{n}} d\mathcal{S} = 0, \quad (48.91)$$

so that the mass budget is given by

$$\frac{d}{dt} \left[\int_{\mathcal{R}} \rho dV \right] = - \oint_{\partial\mathcal{R}} \left[\rho (\mathbf{v}^\dagger - \mathbf{v}^{(b)}) \right] \cdot \hat{\mathbf{n}} d\mathcal{S} = - \oint_{\partial\mathcal{R}} \left[\rho (\mathbf{v} - \mathbf{v}^{(b)}) \right] \cdot \hat{\mathbf{n}} d\mathcal{S} \quad (48.92)$$

Hence, the eddy velocity contribution to the mass budget for any finite region vanishes, which is expected since it provides no net mass source to a region. Furthermore, one may choose to diagnose the right hand side of the mass budget in either the residual mean or Eulerian mean form. The choice is based on convenience, such as whether one has easier access to the residual mean velocity or the Eulerian mean velocity. Although the patterns of the fluxes across any particular boundary differs if $\mathbf{v}^* \neq 0$, the accumulation of mass within the region are identical for the two formulations.

48.8.2 Skew flux formulation

Now consider the complement perspective afforded by the skew flux formulation from Section 48.7. Here we decompose the advective tracer flux according to

$$C \rho \mathbf{v}^\dagger = C \rho \mathbf{v} - \nabla C \wedge \rho \Psi^* + \nabla \wedge (C \rho \Psi^*) = C \rho \mathbf{v} + \mathbf{J}^{\text{skew}} + \nabla \wedge (C \rho \Psi^*), \quad (48.93)$$

where we introduced the skew tracer flux arising from the eddy-induced streamfunction

$$\mathbf{J}^{\text{skew}} = -\nabla C \wedge (\rho \Psi^*). \quad (48.94)$$

The differential budget for tracer is thus given by

$$\frac{\partial(\rho C)}{\partial t} + \nabla \cdot [\rho C \mathbf{v} + \mathbf{J}^{\text{skew}} + \mathbf{J}^{\text{diff}}] = 0, \quad (48.95)$$

where the rotational term, $\nabla \wedge (C \rho \Psi^*)$, has zero divergence and thus does not affect the tracer budget. The corresponding finite volume tracer mass budget is

$$\frac{d}{dt} \left[\int_{\mathcal{R}} \rho C dV \right] = - \oint_{\partial \mathcal{R}} [\rho C (\mathbf{v} - \mathbf{v}^{(b)}) - \nabla C \wedge (\rho \Psi^*) + \mathbf{J}^{\text{diff}}] \cdot \hat{\mathbf{n}} d\mathcal{S}. \quad (48.96)$$

In this form, the contribution from the eddies is inside the skew flux rather than in the residual mean advective flux. Setting C to a constant reveals the mass budget as in the second form of equation (48.92).

48.8.3 Domain with a tracer boundary

We now apply the previous general budget discussion to a specific domain that anticipates the more complete budget analysis provided in Section 52.11. Here, we consider the fluid mass and tracer mass within an ocean region with at least one of its bounds determined by an isosurface of constant tracer concentration, as in Figure 48.4.

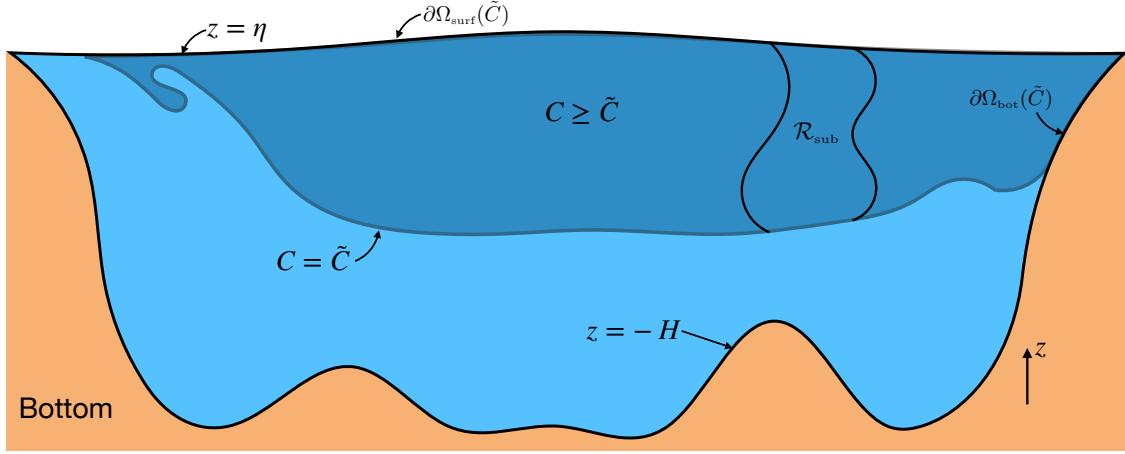


FIGURE 48.4: An ocean region with one of its boundaries set by a surface of constant tracer, $C = \tilde{C}$. Note that the region need not be monotonically stratified in the tracer concentration, nor does it need to be simply connected. The region is bounded at the top by $\partial\Omega_{\text{surf}}(\tilde{C})$; at the bottom by the $C = \tilde{C}$ isosurface as well as the solid-earth bottom $\partial\Omega_{\text{bot}}(\tilde{C})$. The region can generally be multiply connected. A subregion, \mathcal{R}_{sub} , is also considered where its sides extend from the free surface to the tracer isosurface, and they are assumed to be fully within the fluid domain. We develop the tracer and fluid mass budgets for region \mathcal{R}_{sub} in Section 48.8.4, whereas the budget in the full region $C \geq \tilde{C}$ is considered in Section 48.8.3.

Advection formulation

The tracer mass budget written using the advective formulation (48.88) takes the form

$$\frac{d}{dt} \left[\int_{\mathcal{R}} \rho C dV \right] = \int_{\partial\Omega_{\text{surf}}(\tilde{C})} Q_m C dA - \tilde{C} \int_{C=\tilde{C}} \rho (\mathbf{v}^\dagger - \mathbf{v}^{(b)}) \cdot \hat{\mathbf{n}} d\mathcal{S} - \oint_{\partial\mathcal{R}} \mathbf{J}^{\text{diff}} \cdot \hat{\mathbf{n}} d\mathcal{S}. \quad (48.97)$$

For the first right hand side term we made use of the surface kinematic boundary condition (15.73c), with $Q_m dA$ the mass per time crossing the surface interface and where dA is the horizontal projection of the interface area element $d\mathcal{S}$. We also made use of the exterior boundary condition (48.89) for the eddy-induced velocity. For the second term we pulled the tracer concentration outside of the boundary integral over the $C = \tilde{C}$ interface since the concentration is fixed at \tilde{C} on this interface.

The mass budget for this region, also formulated using advective fluxes, is given by

$$\frac{d}{dt} \left[\int_{\mathcal{R}} \rho dV \right] = \int_{\partial\Omega_{\text{surf}}(\tilde{C})} Q_m dA - \int_{C=\tilde{C}} \rho (\mathbf{v}^\dagger - \mathbf{v}^{(b)}) \cdot \hat{\mathbf{n}} d\mathcal{S}. \quad (48.98)$$

Combining this budget with the tracer mass budget allows us to write

$$\frac{d}{dt} [M_C - \tilde{C} M] = \int_{\partial\Omega_{\text{surf}}(\tilde{C})} Q_m (C - \tilde{C}) dA - \oint_{\partial\mathcal{R}} \mathbf{J}^{\text{diff}} \cdot \hat{\mathbf{n}} d\mathcal{S}, \quad (48.99)$$

where we introduced the shorthand for the tracer mass and fluid mass in the region

$$M_C = \int_{\mathcal{R}} C \rho dV \quad \text{and} \quad M = \int_{\mathcal{R}} \rho dV. \quad (48.100)$$

In Section 52.11.2 we motivate the name *internal tracer mass* for the quantity $M_C - \tilde{C} M$.

Skew flux formulation

The tracer mass budget formulated using skew tracer fluxes is generally given by equation (48.96). It takes on the following form for the domain in Figure 48.4

$$\frac{d}{dt} \left[\int_{\mathcal{R}} \rho C dV \right] = \int_{\partial\Omega_{\text{surf}}(\tilde{C})} Q_m C dA - \tilde{C} \int_{C=\tilde{C}} \rho (\mathbf{v} - \mathbf{v}^{(b)}) \cdot \hat{\mathbf{n}} d\mathcal{S} - \oint_{\partial\mathcal{R}} [-\nabla C \wedge (\rho \Psi^*) + \mathbf{J}^{\text{diff}}] \cdot \hat{\mathbf{n}} d\mathcal{S}, \quad (48.101)$$

and the corresponding budget for the fluid mass is

$$\frac{d}{dt} \left[\int_{\mathcal{R}} \rho dV \right] = \int_{\partial\Omega_{\text{surf}}(\tilde{C})} Q_m dA - \int_{C=\tilde{C}} \rho (\mathbf{v} - \mathbf{v}^{(b)}) \cdot \hat{\mathbf{n}} d\mathcal{S}. \quad (48.102)$$

As for the advective formulation, we combine the fluid mass budget equation (48.102) with the tracer mass equation (48.101) to render a budget equation for the internal mass content

$$\frac{d}{dt} [M_C - \tilde{C} M] = \int_{\partial\Omega_{\text{surf}}(\tilde{C})} Q_m (C - \tilde{C}) dA - \oint_{\partial\mathcal{R}} \mathbf{J}^{\text{diff}} \cdot \hat{\mathbf{n}} d\mathcal{S}, \quad (48.103)$$

which is the same as for the advective formulation given by equation (48.99).

Proving the budgets based on the two formulations are equivalent

The two tracer budgets, (48.97) and (48.101), must be the same since they measure changes to the tracer mass within the same region. Likewise, the two mass budgets, (48.98) and (48.102), must be the same, as are the two internal tracer mass budgets (48.99) and (48.103). We here expose the manipulations required to verify these equalities.

To prove the $C = \tilde{C}$ terms in the tracer budget equations (48.97) and (48.101) are the same, consider the identity (48.91) applied to the region under consideration

$$0 = \oint_{\partial\mathcal{R}} \rho \mathbf{v}^* \cdot \hat{\mathbf{n}} d\mathcal{S} = \int_{\partial\Omega_{\text{surf}}(\tilde{C})} \rho \mathbf{v}^* \cdot \hat{\mathbf{n}} d\mathcal{S} + \int_{\partial\Omega_{\text{bot}}(\tilde{C})} \rho \mathbf{v}^* \cdot \hat{\mathbf{n}} d\mathcal{S} + \int_{C=\tilde{C}} \rho \mathbf{v}^* \cdot \hat{\mathbf{n}} d\mathcal{S}. \quad (48.104)$$

The surface and bottom boundary terms vanish due to the external boundary condition (48.89); namely, $\mathbf{v}^* \cdot \hat{\mathbf{n}} = 0$ for each point along an external fluid boundary. We are thus led to conclude that

$$\int_{C=\tilde{C}} \rho \mathbf{v}^* \cdot \hat{\mathbf{n}} d\mathcal{S} = 0. \quad (48.105)$$

This boundary integral means that there is no net accumulation of mass in the region due to action of the eddy velocity. Notably, we generally have $\mathbf{v}^* \cdot \hat{\mathbf{n}} \neq 0$ at any particular point on the $C = \tilde{C}$ surface. Given the boundary integral (48.105) we are thus led to conclude

$$\int_{C=\tilde{C}} \rho (\mathbf{v}^\dagger - \mathbf{v}^{(b)}) \cdot \hat{\mathbf{n}} d\mathcal{S} = \int_{C=\tilde{C}} \rho (\mathbf{v} - \mathbf{v}^{(b)}) \cdot \hat{\mathbf{n}} d\mathcal{S}. \quad (48.106)$$

This identity proves that the two mass budgets (48.98) and (48.102) are indeed measuring changes to the same fluid mass, even though one computes the domain boundary fluxes based on the residual mean velocity, \mathbf{v}^\dagger , whereas the other uses the Eulerian mean, \mathbf{v} .

Next we need to show that the skew flux term vanishes when integrated around the domain boundary. For the $C = \tilde{C}$ boundary we have

$$\int_{C=\tilde{C}} [\nabla C \wedge (\rho \Psi^*)] \cdot \hat{\mathbf{n}} d\mathcal{S} = 0, \quad (48.107)$$

which follows since $\hat{\mathbf{n}}$ is parallel to ∇C along this boundary. For the external boundaries, equality of the tracer mass budgets (48.97) and (48.101) is satisfied for an arbitrary tracer concentration if one of the boundary conditions discussed in Section (48.7.3) is maintained; i.e., if Ψ^* vanishes on an external boundary or if it is parallel to the boundary normal direction ($\hat{\mathbf{n}} \wedge \Psi^* = 0$). Maintenance of either of these two boundary conditions allows us to conclude that the two budgets (48.97) and (48.101) are indeed identical.

48.8.4 Budget for a region with interior sides

Consider the subregion \mathcal{R}_{sub} shown in Figure 48.4. This region is bounded above by the free surface and below by the tracer isosurface, $C = \tilde{C}$, just like the region \mathcal{R} encountered in Section 48.8.3. Additionally, region \mathcal{R}_{sub} is bounded along its sides by surfaces assumed to be within the fluid interior. For much of this discussion we allow the sides to have an arbitrary shape and to move. Towards the end of this section we specialize to the case of static sides, such as might be relevant for a vertical section through the fluid and/or a numerical model grid cell.

Fluid mass budget

The fluid mass budget for the region \mathcal{R}_{sub} can be formulated using either the residual mean velocity or the Eulerian mean velocity

$$\frac{d}{dt} \left[\int_{\mathcal{R}_{\text{sub}}} \rho dV \right] = \int_{\partial\Omega_{\text{surf}}(\tilde{C})} Q_m dA - \int_{C=\tilde{C}} \rho (\mathbf{v}^\dagger - \mathbf{v}^{(b)}) \cdot \hat{\mathbf{n}} d\mathcal{S} - \int_{\text{sides}} \rho (\mathbf{v}^\dagger - \mathbf{v}^{(b)}) \cdot \hat{\mathbf{n}} d\mathcal{S} \quad (48.108a)$$

$$\frac{d}{dt} \left[\int_{\mathcal{R}_{\text{sub}}} \rho dV \right] = \int_{\partial\Omega_{\text{surf}}(\tilde{C})} Q_m dA - \int_{C=\tilde{C}} \rho (\mathbf{v} - \mathbf{v}^{(b)}) \cdot \hat{\mathbf{n}} d\mathcal{S} - \int_{\text{sides}} \rho (\mathbf{v} - \mathbf{v}^{(b)}) \cdot \hat{\mathbf{n}} d\mathcal{S}. \quad (48.108b)$$

The two budgets are identical since the eddy velocity satisfies $\oint \rho \mathbf{v}^* \cdot \hat{\mathbf{n}} d\mathcal{S} = 0$ for any domain, as well as $\mathbf{v}^* \cdot \hat{\mathbf{n}} = 0$ along any external domain boundary. Hence, as already noted in Section 48.8.1,

the eddy velocity contribution to the right hand side of equation (48.108a) vanishes; it provides no net mass source to any region. We next show the same equality holds for the tracer mass budgets, with that equality shown with a bit more effort.

Tracer mass budget

The advective flux formulation of the tracer mass budget is given by

$$\frac{d}{dt} \left[\int_{\mathcal{R}_{\text{sub}}} \rho C dV \right] = \int_{\partial\Omega_{\text{surf}}(\tilde{C})} Q_m C dA - \tilde{C} \int_{C=\tilde{C}} \rho (\mathbf{v}^\dagger - \mathbf{v}^{(b)}) \cdot \hat{\mathbf{n}} d\mathcal{S} - \oint_{\partial\mathcal{R}_{\text{sub}}} \mathbf{J}^{\text{diff}} \cdot \hat{\mathbf{n}} d\mathcal{S} - \int_{\text{sides}} C \rho (\mathbf{v}^\dagger - \mathbf{v}^{(b)}) \cdot \hat{\mathbf{n}} d\mathcal{S}, \quad (48.109)$$

and the corresponding skew flux formulation is

$$\frac{d}{dt} \left[\int_{\mathcal{R}_{\text{sub}}} \rho C dV \right] = \int_{\partial\Omega_{\text{surf}}(\tilde{C})} Q_m C dA - \tilde{C} \int_{C=\tilde{C}} \rho (\mathbf{v} - \mathbf{v}^{(b)}) \cdot \hat{\mathbf{n}} d\mathcal{S} - \oint_{\partial\mathcal{R}_{\text{sub}}} \mathbf{J}^{\text{diff}} \cdot \hat{\mathbf{n}} d\mathcal{S} - \int_{\text{sides}} C \rho (\mathbf{v} - \mathbf{v}^{(b)}) \cdot \hat{\mathbf{n}} d\mathcal{S} - \int_{\text{sides}} [-\nabla C \wedge (\rho \Psi^*)] \cdot \hat{\mathbf{n}} d\mathcal{S}. \quad (48.110)$$

As for the discussion in Section 48.8.3, we introduce the internal tracer mass and make use of the fluid mass budgets (48.108a) and (48.108b) to write the advective form of the internal mass budget

$$\frac{d}{dt} [M_C - \tilde{C} M] = \int_{\partial\Omega_{\text{surf}}(\tilde{C})} Q_m (C - \tilde{C}) dA - \oint_{\partial\mathcal{R}} \mathbf{J}^{\text{diff}} \cdot \hat{\mathbf{n}} d\mathcal{S} - \int_{\text{sides}} (C - \tilde{C}) \rho (\mathbf{v}^\dagger - \mathbf{v}^{(b)}) \cdot \hat{\mathbf{n}} d\mathcal{S}, \quad (48.111)$$

and the corresponding skew flux form of the same budget

$$\frac{d}{dt} [M_C - \tilde{C} M] = \int_{\partial\Omega_{\text{surf}}(\tilde{C})} Q_m (C - \tilde{C}) dA - \oint_{\partial\mathcal{R}} \mathbf{J}^{\text{diff}} \cdot \hat{\mathbf{n}} d\mathcal{S} - \int_{\text{sides}} (C - \tilde{C}) \rho (\mathbf{v} - \mathbf{v}^{(b)}) \cdot \hat{\mathbf{n}} d\mathcal{S} - \int_{\text{sides}} [-\nabla C \wedge (\rho \Psi^*)] \cdot \hat{\mathbf{n}} d\mathcal{S}. \quad (48.112)$$

We now examine the right hand side of the budgets (48.111) and (48.112) to show they are indeed measuring the same tracer mass budget. For that purpose, consider the skew flux integral in equation (48.112) and note that the integrand vanishes on both the top of the domain, at $z = \eta$, and bottom at $C = \tilde{C}$, thus allowing us to write

$$-\int_{\text{sides}} [-\nabla C \wedge (\rho \Psi^*)] \cdot \hat{\mathbf{n}} d\mathcal{S} = -\oint_{\partial\mathcal{R}_{\text{sub}}} [-\nabla C \wedge (\rho \Psi^*)] \cdot \hat{\mathbf{n}} d\mathcal{S}, \quad (48.113)$$

where the right hand side is an integral around the full domain boundaries. Now reintroduce the eddy induced velocity and rotational flux to have

$$-\oint_{\partial\mathcal{R}_{\text{sub}}} [-\nabla C \wedge (\rho \Psi^*)] \cdot \hat{\mathbf{n}} d\mathcal{S} = -\oint_{\partial\mathcal{R}_{\text{sub}}} [C \rho \mathbf{v}^* - \nabla \wedge (C \rho \Psi^*)] \cdot \hat{\mathbf{n}} d\mathcal{S}. \quad (48.114)$$

The rotational flux has zero divergence, so that Gauss's divergence theorem means that the rotational flux vanishes when integrated along the domain boundaries

$$\oint_{\partial\mathcal{R}_{\text{sub}}} \nabla \wedge (C \rho \Psi^*) \cdot \hat{\mathbf{n}} d\mathcal{S} = 0. \quad (48.115)$$

The eddy advection term in equation (48.114) vanishes on the top boundary at $z = \eta$ due to the boundary condition $\mathbf{v}^* \cdot \hat{\mathbf{n}} = 0$, thus yielding

$$\oint_{\partial\mathcal{R}_{\text{sub}}} C \rho \mathbf{v}^* \cdot \hat{\mathbf{n}} d\mathcal{S} = \int_{\text{sides}} C \rho \mathbf{v}^* \cdot \hat{\mathbf{n}} d\mathcal{S} + \int_{C=\tilde{C}} C \rho \mathbf{v}^* \cdot \hat{\mathbf{n}} d\mathcal{S} \quad (48.116a)$$

$$= \int_{\text{sides}} C \rho \mathbf{v}^* \cdot \hat{\mathbf{n}} d\mathcal{S} + \tilde{C} \int_{C=\tilde{C}} \rho \mathbf{v}^* \cdot \hat{\mathbf{n}} d\mathcal{S}. \quad (48.116b)$$

Again make use of the property $\nabla \cdot (\rho \mathbf{v}^*) = 0$ and $\hat{\mathbf{n}} \cdot \mathbf{v}^* = 0$ at $z = \eta$ to write

$$0 = \oint_{\partial\mathcal{R}_{\text{sub}}} \rho \mathbf{v}^* \cdot \hat{\mathbf{n}} d\mathcal{S} = \int_{\text{sides}} \rho \mathbf{v}^* \cdot \hat{\mathbf{n}} d\mathcal{S} + \oint_{C=\tilde{C}} \rho \mathbf{v}^* \cdot \hat{\mathbf{n}} d\mathcal{S}, \quad (48.117)$$

which gives us

$$-\int_{\text{sides}} [-\nabla C \wedge (\rho \Psi^*)] \cdot \hat{\mathbf{n}} d\mathcal{S} = -\oint_{\partial\mathcal{R}_{\text{sub}}} C \rho \mathbf{v}^* \cdot \hat{\mathbf{n}} d\mathcal{S} = -\int_{\text{sides}} (C - \tilde{C}) \rho \mathbf{v}^* \cdot \hat{\mathbf{n}} d\mathcal{S}. \quad (48.118)$$

Making use of this result in the skew flux formulated budget equation (48.112) brings it to the advective flux form found in equation (48.111).

We are thus led to conclude that the right hand side to equation (48.112) does indeed equal to the right hand side of equation (48.111). So although the formulation of the boundary flux contributions is rather distinct between the advective flux and skew flux formulations, the resulting tracer mass budget is the same. The choice for how to formulate the budget is thus a matter of convenience.

48.8.5 Budget for a stirred fluid in a region with interior sides

Although contained within the formalism developed in Section 48.8.4, it is revealing to specialize those budgets to the case of zero mixing, in which $\mathbf{J}^{\text{diff}} = 0$ and there is zero boundary mass flux, $Q_m = 0$. In this case the fluid is reversibly stirred. Examining the finite region budgets for this pure stirring case allows us to further reveal the complementary treatments available from advection versus skewson.

Summary of the differential stirring formalism

As explored in this chapter, an Eulerian description of tracer stirring can arise from either advection or skewson. In the presence of an eddy induced velocity we here consider two representations of tracer stirring, with the first being advection by the residual mean velocity, \mathbf{v}^\dagger

$$\rho \frac{D^\dagger C}{Dt} = \frac{\partial(\rho C)}{\partial t} + \nabla \cdot (\rho \mathbf{v}^\dagger C) = 0. \quad (48.119)$$

This formulation makes it clear that surfaces of constant C are material as defined by the residual mean velocity rather than by the Eulerian mean, \mathbf{v} . That is, tracer isosurfaces satisfy the residual mean impermeability condition

$$\rho (\mathbf{v}^\dagger - \mathbf{v}^{(b)}) \cdot \hat{\mathbf{n}} = 0 \quad \text{on } C \text{ isosurfaces,} \quad (48.120)$$

with

$$\hat{\mathbf{n}} = \frac{\nabla C}{|\nabla C|} \quad \text{and} \quad \mathbf{v}^{(b)} \cdot \hat{\mathbf{n}} = -\frac{\partial_t C}{|\nabla C|}. \quad (48.121)$$

The impermeability condition (48.120) offers a geometric interpretation of the tracer equation (48.119) following from the discussion of dia-surface transport in Section 44.4. Correspondingly, Lagrangian fluid particles moving with the residual mean velocity do not cross tracer isosurfaces even if those isosurfaces move. Furthermore, we observe that the eddy induced velocity has a nonzero projection across tracer isosurfaces

$$(\mathbf{v}^\dagger - \mathbf{v}^{(b)}) \cdot \hat{\mathbf{n}} = 0 \implies (\mathbf{v} - \mathbf{v}^{(b)}) \cdot \hat{\mathbf{n}} = -\mathbf{v}^* \cdot \hat{\mathbf{n}} \quad \text{on } C \text{ isosurfaces.} \quad (48.122)$$

This property of the eddy induced velocity was emphasized by [McDougall and McIntosh \(2001\)](#). It reveals that in the absence of mixing, eddy motion crossing tracer isosurfaces is exactly balanced by Eulerian plus boundary motion, thus leaving a net zero residual mean transfer of matter across the surface. Equation (48.122) is a key kinematic property used for interpreting features of the finite volume budgets detailed below.

Our second means to represent tracer stirring makes use of advection by the Eulerian mean velocity plus skewson by the eddy induced streamfunction

$$\rho \frac{DC}{Dt} + \nabla \cdot [-\nabla C \wedge (\rho \Psi^*)] = \frac{\partial(\rho C)}{\partial t} + \nabla \cdot [\rho \mathbf{v} C - \nabla C \wedge (\rho \Psi^*)] = 0. \quad (48.123)$$

In terms of the eddy streamfunction, $\rho \Psi^*$, the impermeability condition (48.120) takes on the form

$$[\rho \mathbf{v} + \nabla \wedge (\rho \Psi^*) - \rho \mathbf{v}^{(b)}] \cdot \hat{\mathbf{n}} = 0 \quad \text{on } C \text{ isosurfaces.} \quad (48.124)$$

Budgets via residual mean advection

The mass budget formulated in terms of residual mean advection, and the corresponding residual mean advective flux formulation of the tracer mass budget, are given by

$$\frac{d}{dt} \left[\int_{\mathcal{R}_{\text{sub}}} \rho dV \right] = - \int_{C=\tilde{C}} \rho (\mathbf{v}^\dagger - \mathbf{v}^{(b)}) \cdot \hat{\mathbf{n}} dS - \int_{\text{sides}} \rho (\mathbf{v}^\dagger - \mathbf{v}^{(b)}) \cdot \hat{\mathbf{n}} dS \quad (48.125a)$$

$$\frac{d}{dt} \left[\int_{\mathcal{R}_{\text{sub}}} \rho C dV \right] = - \tilde{C} \int_{C=\tilde{C}} \rho (\mathbf{v}^\dagger - \mathbf{v}^{(b)}) \cdot \hat{\mathbf{n}} dS - \int_{\text{sides}} C \rho (\mathbf{v}^\dagger - \mathbf{v}^{(b)}) \cdot \hat{\mathbf{n}} dS. \quad (48.125b)$$

The residual mean impermeability condition (48.120) for the $C = \tilde{C}$ surface renders a simplification to the fluid mass and tracer mass budgets

$$\frac{d}{dt} \left[\int_{\mathcal{R}_{\text{sub}}} \rho dV \right] = - \int_{\text{sides}} \rho (\mathbf{v}^\dagger - \mathbf{v}^{(b)}) \cdot \hat{\mathbf{n}} dS \quad (48.126a)$$

$$\frac{d}{dt} \left[\int_{\mathcal{R}_{\text{sub}}} \rho C dV \right] = - \int_{\text{sides}} C \rho (\mathbf{v}^\dagger - \mathbf{v}^{(b)}) \cdot \hat{\mathbf{n}} dS. \quad (48.126b)$$

Hence, in the residual mean formulation, the only fluxes that affect changes to the mass budgets are those that cross the side faces of the region.

Budgets via Eulerian mean advection plus eddy skewson

The mass budget formulated in terms of Eulerian mean advection, and the corresponding tracer mass budget using eddy skewson, are given by

$$\frac{d}{dt} \left[\int_{\mathcal{R}_{\text{sub}}} \rho dV \right] = - \int_{C=\tilde{C}} \rho (\mathbf{v} - \mathbf{v}^{(b)}) \cdot \hat{\mathbf{n}} d\mathcal{S} - \int_{\text{sides}} \rho (\mathbf{v} - \mathbf{v}^{(b)}) \cdot \hat{\mathbf{n}} d\mathcal{S} \quad (48.127a)$$

$$\begin{aligned} \frac{d}{dt} \left[\int_{\mathcal{R}_{\text{sub}}} \rho C dV \right] &= -\tilde{C} \int_{C=\tilde{C}} \rho (\mathbf{v} - \mathbf{v}^{(b)}) \cdot \hat{\mathbf{n}} d\mathcal{S} - \int_{\text{sides}} C \rho (\mathbf{v} - \mathbf{v}^{(b)}) \cdot \hat{\mathbf{n}} d\mathcal{S} \\ &\quad - \int_{\text{sides}} [-\nabla C \wedge (\rho \Psi^*)] \cdot \hat{\mathbf{n}} d\mathcal{S}. \end{aligned} \quad (48.127b)$$

We already saw in Section 48.8.4 how to bring the right hand side terms into the form realized by the residual mean advective approach. So there is no question concerning the equivalence of the advective and skew flux formulations for the tracer mass budget. Nonetheless, what is here clearly emphasized is that the skew flux approach requires us to account for Eulerian advective transport across the $C = \tilde{C}$ isosurface, whereas for the advective flux approach the only flux in equation (48.126b) is that crossing the region side boundaries. Even so, as stated earlier, an Eulerian mean transport of tracer across the $C = \tilde{C}$ isosurface *does not* correspond to material transport across this surface. The reason is that material transport is determined by the residual mean velocity, \mathbf{v}^\dagger , as per the residual mean impermeability conditions 48.122 and 48.124. So even though there is a contribution to the skew flux formulated budget from Eulerian transport across the $C = \tilde{C}$ material surface, there remains zero net material crossing that surface.

Zero Eulerian mean advection and static side walls

One further specialization serves to clearly emphasize the complementary nature of the advective and skew flux approaches. Here, we assume the sides of the region are static and the Eulerian mean velocity vanishes. With a zero Eulerian velocity, the residual mean impermeability condition (48.122) means that on the $C = \tilde{C}$ isosurface, the normal component of the eddy velocity is balanced by the boundary velocity as per the impermeability condition (48.128):

$$(\mathbf{v}^* - \mathbf{v}^{(b)}) \cdot \hat{\mathbf{n}} = 0 \quad \text{on } C \text{ isosurfaces and with } \mathbf{v} = 0. \quad (48.128)$$

When formulated using the residual mean advection, the fluid mass budget (48.126a) and tracer mass budget (48.126b) reduce in this case to

$$\frac{d}{dt} \left[\int_{\mathcal{R}_{\text{sub}}} \rho dV \right] = - \int_{\text{sides}} \rho \mathbf{v}^* \cdot \hat{\mathbf{n}} d\mathcal{S} \quad (48.129a)$$

$$\frac{d}{dt} \left[\int_{\mathcal{R}_{\text{sub}}} \rho C dV \right] = - \int_{\text{sides}} C \rho \mathbf{v}^* \cdot \hat{\mathbf{n}} d\mathcal{S}, \quad (48.129b)$$

so that these budgets are only affected by eddy advection across the side boundaries. The corresponding mass budget written in terms of Eulerian mean advection, (48.127a), and tracer mass budget written in terms of skew fluxes, (48.127b), are given by

$$\frac{d}{dt} \left[\int_{\mathcal{R}_{\text{sub}}} \rho dV \right] = + \int_{C=\tilde{C}} \rho \mathbf{v}^{(b)} \cdot \hat{\mathbf{n}} d\mathcal{S} \quad (48.130a)$$

$$\frac{d}{dt} \left[\int_{\mathcal{R}_{\text{sub}}} \rho C dV \right] = +\tilde{C} \int_{C=\tilde{C}} \rho \mathbf{v}^{(b)} \cdot \hat{\mathbf{n}} d\mathcal{S} - \int_{\text{sides}} [-\nabla C \wedge (\rho \Psi^*)] \cdot \hat{\mathbf{n}} d\mathcal{S}. \quad (48.130b)$$

For the mass budget, (48.130a), there are no contributions to the side walls since they are static and the Eulerian advection is assumed to vanish. The only contribution comes from the eddy term acting on the $C = \tilde{C}$ isosurface where $\mathbf{v}^{(b)} \cdot \hat{\mathbf{n}} = \mathbf{v}^* \cdot \hat{\mathbf{n}}$. For the tracer mass budget, (48.130b), we also have the eddy contribution on the $C = \tilde{C}$ isosurface, plus skew fluxes that penetrate the side walls.

The right hand sides to the fluid mass budgets (48.129a) and (48.130a), and tracer mass budgets (48.129b) and (48.130b), are remarkably distinct. Even so, they both measure the same budgets. Furthermore, in both cases the $C = \tilde{C}$ boundary is a material boundary as defined by the residual mean velocity.

48.9 Active tracers and dia-surface flow

An *active tracer* is a tracer that impacts the dynamics, with potential temperature and salinity the canonical examples. Active tracers directly impact the buoyancy, which in turn affects pressure and velocity. Hence, the advection-diffusion equation for active tracers is nonlinear since the velocity field is dependent on active tracers. We here write the advection-diffusion equation in terms of the residual mean velocity using potential temperature as an example active tracer

$$\rho \frac{D^\dagger \theta}{Dt} = \rho \left[\frac{\partial}{\partial t} + \mathbf{v}^\dagger \cdot \nabla \right] \theta = -\nabla \cdot \mathbf{J}^{\text{diff}}(\theta). \quad (48.131)$$

Further nonlinearities arise when the subgrid scale diffusion tensor is itself a function of the buoyancy, as discussed at the end of Section 48.4.4, and/or when the parameterized eddy-induced velocity is a function of the buoyancy, as discussed in Section 50.2.

48.9.1 Adiabatic flow

In an adiabatic fluid, potential temperature is materially invariant. When advected by the residual mean velocity this material invariance takes on the form

$$\frac{D^\dagger \theta}{Dt} = \frac{\partial \theta}{\partial t} + \mathbf{v}^\dagger \cdot \nabla \theta = 0 \quad \text{adiabatic.} \quad (48.132)$$

Furthermore, following the kinematics of Section 15.4.2, the adiabatic residual mean flow field does not penetrate surfaces of constant potential temperature (θ -isosurfaces are impermeable) since

$$\mathbf{v}^\dagger \cdot \nabla \theta = -\frac{\partial \theta}{\partial t}. \quad (48.133)$$

In this case we say that residual mean advection reversibly stirs the potential temperature field. This property of the residual mean velocity was also considered in the discussion of pure stirring in Section 48.8.4.

48.9.2 Diabatic processes generating dia-surface transport

Potential temperature is no longer materially invariant in the presence of diabatic processes such as mixing typically parameterized by diffusion. Correspondingly, the residual mean velocity picks up a component, w^{dia} , that crosses the moving potential temperature surface thus making it permeable. Hence, advective transport in the presence of mixing is no longer reversible. We follow the kinematics from Section 44.4 to render the expression (44.26) for w^{dia} , which we here write as

$$w^{\text{dia}} \equiv \hat{\mathbf{n}} \cdot (\mathbf{v}^\dagger - \mathbf{v}^{(\theta)}) = \frac{1}{|\nabla \theta|} \frac{D^\dagger \theta}{Dt} \quad (48.134)$$

where

$$\hat{\mathbf{n}} = \frac{\nabla\theta}{|\nabla\theta|} \quad \text{and} \quad \mathbf{v}^{(\theta)} \cdot \nabla\theta = -\frac{\partial\theta}{\partial t}. \quad (48.135)$$

Rearrangement of equation (48.134) renders the kinematic identity

$$\frac{D^\dagger\theta}{Dt} = \frac{\partial\theta}{\partial t} + \mathbf{v}^\dagger \cdot \nabla\theta = w^{\text{dia}} |\nabla\theta|. \quad (48.136)$$

With nonzero w^{dia} , we no longer have residual mean advection preserving properties along fluid element trajectories. We offer examples to illustrate the physics underlying this identity.

Diffusion with no fluid motion

Diffusion is the canonical example of a diabatic process (Section 48.3)

$$\rho \frac{D^\dagger\theta}{Dt} = \nabla \cdot (\kappa \rho \nabla\theta), \quad (48.137)$$

with $\kappa > 0$ the scalar kinematic diffusivity. Diffusion in turn drives a diabatic transport velocity

$$\rho w^{\text{dia}} = \frac{\nabla \cdot (\kappa \rho \nabla\theta)}{|\nabla\theta|}. \quad (48.138)$$

Consider a horizontally homogeneous potential temperature field. If buoyancy is alone determined by potential temperature then there is no fluid motion since buoyancy surfaces are flat (and we assume the eddy-induced motion is also zero). Yet in the presence of vertical diffusion and vertical stratification there is a diabatic transport since

$$\rho w^{\text{dia}} = \frac{\partial_z(\kappa \rho \partial_z\theta)}{|\partial_z\theta|} \neq 0. \quad (48.139)$$

In the absence of fluid flow, the dia-surface transport is determined solely by movement of the potential temperature surfaces. Correspondingly, potential temperature evolution is determined only by vertical diffusion since with $\mathbf{v}^\dagger = 0$ we have

$$\frac{\partial\theta}{\partial t} = \rho^{-1} \frac{\partial}{\partial z} \left[\kappa \rho \frac{\partial\theta}{\partial z} \right] = w^{\text{dia}} \left| \frac{\partial\theta}{\partial z} \right|. \quad (48.140)$$

Steady state advective-diabatic balance

A steady state potential temperature field in the presence of diabatic processes is realized when there is an exact balance between advective transport and dia-surface transport enabled by diffusion

$$\rho \mathbf{v}^\dagger \cdot \nabla\theta = \rho w^{\text{dia}} |\nabla\theta| = \nabla \cdot (\kappa \rho \nabla\theta). \quad \text{steady state.} \quad (48.141)$$

That is, maintaining static θ -surfaces ($\partial_t\theta = 0$) requires the residual mean advective transport to cross θ surfaces (left hand side) by an amount that exactly balances diabatic processes such as diffusion (right hand side). This example illustrates that in the presence of mixing, advection is no longer an adiabatic stirring process. Namely, in the steady state, advection equals diffusion.

48.10 Green's function method for passive tracers

Passive tracers (Section 16.2) have no impact on the flow field so that their patterns serve to “trace” the effects from advective and diffusive transport without affecting the transport velocity or diffusion tensor. Correspondingly, the passive tracer advection-diffusion equation is a linear partial differential equation since the tracers do not impact on the flow, the fluid density, or the diffusion tensor. With some qualifications identified below, we are able to make use of the Green's function methods from Chapter 4 when studying solutions to the passive tracer equation. In this section we develop the Green's function method for use with the advection-diffusion equation.

48.10.1 Concerning time dependent domain boundaries

The ocean free surface is time dependent, so that the domain, \mathcal{R} , is itself time dependent. Mathematically, this time dependence means that time and space integration do not commute. In particular, space integration is generally written in the form

$$\int_{\mathcal{R}} dV = \iint \left[\int_{\eta_b}^{\eta} dz \right] dx dy, \quad (48.142)$$

where $z = \eta_b(x, y)$ is the static bottom and $z = \eta(x, y, t)$ is the time dependent free surface. We must, in turn, first compute the space integration over the full domain and then do the time integration when forming reciprocity and the Green's function solution to the tracer equation.

The free surface undulations make the vertical extent of the domain time dependent. Additionally, in an ocean with sloping sidewalls, the horizontal domain boundaries are also time dependent as the ocean moves up and down the sloping sides. However, allowing for the horizontal domain extent to fluctuate does not introduce any fundamentally new kinematics in our analysis. The reason is that when integrating to the lateral domain boundaries, all terms vanish since the water depth vanishes at the edge of the sloping beaches. We saw this kinematic result when integrating the angular momentum budget in a channel with sloping sidewalls in Section 20.4.

Therefore, we assume the horizontal extent of the domain to be static in order to slightly ease the analysis. We do so by imagining a few meter high vertical seawall placed around the ocean domain edges, and by assuming a minimum depth so that there is nonzero water everywhere in the domain. These assumptions are common in ocean modeling, except in models allowing for wetting and drying of land/ocean cells. So in conclusion, we limit our analysis to time dependence of the vertical extent of the domain, with the horizontal extent static. Such limitation can be removed without much difficulty but doing so adds nothing new fundamentally.

48.10.2 Passive tracer boundary conditions

In this section, we are concerned with the evolution of a smooth passive tracer concentration, C , which is the dimensionless ratio of the tracer mass to seawater mass. Boundary conditions play a key role in the evolution. We here discuss the boundary conditions placed on the passive tracer along the ocean bottom, at $z = \eta_b(x, y)$, and the free surface, at $z = \eta(x, y, t)$.

Ocean bottom

At the static ocean bottom we generically consider a no-flux condition for the diffusive flux

$$\mathbf{J} \cdot \hat{\mathbf{n}} = -\rho \mathbb{K} \cdot \nabla C \cdot \hat{\mathbf{n}} = 0 \quad \text{at } z = \eta_b. \quad (48.143)$$

Along with the kinematic no-normal flow condition, $\mathbf{v} \cdot \hat{\mathbf{n}} = 0$, the no-flux condition means that there is zero tracer flux through the bottom.

Ocean free surface

At the ocean free surface we use results from Section 15.4.3, which developed the boundary conditions for mass flux \mathcal{Q}_m (mass per time per area) across a permeable free surface, as well as Section 16.5, which developed the analogous boundary conditions for tracers. In particular, equation (16.71) provides an expression for the net mass flux of tracer crossing the free surface, \mathcal{Q}_C , written as the sum of an advective flux plus a non-advective flux

$$\mathcal{Q}_C = C \mathcal{Q}_m - \mathbf{J} \cdot \hat{\mathbf{n}} = \underbrace{C \mathcal{Q}_m}_{\text{advective}} + \underbrace{\rho \mathbb{K} \cdot \nabla C \cdot \hat{\mathbf{n}}}_{\text{diffusive}} \quad \text{for } \mathbf{x} \in \partial \mathcal{R}_{\text{surface}}, \quad (48.144)$$

where C is the concentration at $z = \eta$, and we assumed the non-advective flux is given by a diffusive flux. We consider the following prescribed boundary conditions.

- ROBIN CONDITIONS: Prescribing the boundary tracer mass flux, \mathcal{Q}_C , leads to a Robin or mixed boundary condition

$$\mathcal{Q}_C = \text{prescribed} = C \mathcal{Q}_m + \rho \mathbb{K} \cdot \nabla C \cdot \hat{\mathbf{n}} \quad \text{for } \mathbf{x} \in \partial \mathcal{R}_{\text{surface}}. \quad (48.145)$$

This boundary condition is relevant for enthalpy and salt, with full discussion given in Section 51.3. However, the Robin condition is rarely used for passive tracers along the ocean surface and so it will not be considered further.

- NEUMANN CONDITIONS: Prescribing the diffusive flux leads to the Neumann boundary condition

$$\rho \mathbb{K} \cdot \nabla C \cdot \hat{\mathbf{n}} = \text{prescribed} \equiv \Sigma(\mathbf{x}, t) \quad \text{for } \mathbf{x} \in \partial \mathcal{R}_{\text{surface}}. \quad (48.146)$$

This surface ocean boundary condition is also rarely used for passive tracers, though we will examine it within the following.

- DIRICHLET CONDITIONS: Prescribing the value of the tracer concentration at the boundary leads to the Dirichlet boundary condition

$$C = \text{prescribed} \equiv \sigma(\mathbf{x}, t) \quad \text{for } \mathbf{x} \in \partial \mathcal{R}_{\text{surface}}. \quad (48.147)$$

This boundary condition is the most commonly used condition for passive tracers, and thus will be our favored choice in the following. Note that both the Neumann and Dirichlet conditions generally involve a net transport of tracer, $\mathcal{Q}_C \neq 0$, across the ocean boundary.

As we will see in Section 48.10.5, the Neumann boundary condition in the presence of a surface mass flux is problematic due to the associated non-closed reciprocity relation satisfied by the Green's function and its adjoint. The absence of a suitable reciprocity relation makes it difficult to use the Green's function method, since one would need to solve for both the Green's function and its adjoint. In contrast, the Dirichlet condition allows for a simple reciprocity relation, identical to that for the diffusion equation (Section 4.7.4), thus making the Green's function method with a Dirichlet condition suitable even in the presence of a surface mass flux.

48.10.3 Advection-diffusion initial-boundary value problem

As a problem in mathematical physics, we study the initial-boundary value problem for a smooth passive tracer concentration, C , which is the dimensionless ratio of the tracer mass to seawater mass. The tracer is affected by advection and diffusion on a spatial domain, \mathcal{R} , in the presence of a tracer source, $\rho \Lambda$ (with dimensions of tracer mass per volume per time), with initial data available for the density and tracer concentration at time $t = t_{\text{init}}$. The initial-boundary value problem in the presence of Neumann or Dirichlet boundary conditions is given by

$$\frac{\partial(\rho C)}{\partial t} + \nabla \cdot (\rho \mathbf{v}^\dagger C - \rho \mathbb{K} \cdot \nabla C) = \rho \Lambda \quad \mathbf{x} \in \mathcal{R}, t \geq t_{\text{init}} \quad (48.148a)$$

$$\rho C = \rho I \quad \mathbf{x} \in \mathcal{R}, t = t_{\text{init}} \quad (48.148b)$$

$$\hat{\mathbf{n}} \cdot \rho \mathbb{K} \cdot \nabla C = \rho \Sigma \quad \text{or} \quad \rho C = \rho \sigma \quad \mathbf{x} \in \partial \mathcal{R}, t \geq t_{\text{init}} \quad (48.148c)$$

The prescribed initial condition data for the tracer concentration at time $t = t_{\text{init}}$ is given by $I(\mathbf{x})$, and the initial density is also prescribed at this time, $\rho(\mathbf{x}, t_{\text{init}})$. We consider two options for the boundary condition at $\mathbf{x} \in \partial \mathcal{R}$: (i) the Neumann boundary condition with a prescribed flux, $\hat{\mathbf{n}} \cdot \rho \mathbb{K} \cdot \nabla C = \rho \Sigma$, or the (ii) Dirichlet boundary condition with a prescribed value, $\rho C = \rho \sigma$. Furthermore, we assume the flow field, $\mathbf{v}(\mathbf{x}, t)$, the eddy-induced velocity, $\mathbf{v}^*(\mathbf{x}, t)$, the seawater density, $\rho(\mathbf{x}, t)$, and the diffusivity tensor, $\mathbb{K}(\mathbf{x}, t)$, are known functions of space-time that are determined by solving for the dynamics, kinematics, thermodynamics, and active tracers. Finally, we assume the tracer concentration source, $\Lambda(\mathbf{x}, t)$, does not itself depend on the tracer concentration, C , thus ensuring linearity of the partial differential equation (48.148a).

48.10.4 The Green's function and its adjoint

Green's function problem

The Green's function corresponding to the passive tracer advection-diffusion equations (48.148a)-(48.148c) satisfies the following causal boundary value problem

$$\frac{\partial[\rho G(\mathbf{x}, t|\mathbf{x}_0, t_0)]}{\partial t} + \nabla_{\mathbf{x}} \cdot [\rho \mathbf{v}^\dagger G(\mathbf{x}, t|\mathbf{x}_0, t_0) - \rho \mathbb{K} \cdot \nabla_{\mathbf{x}} G(\mathbf{x}, t|\mathbf{x}_0, t_0)] = \delta(\mathbf{x} - \mathbf{x}_0) \delta(t - t_0) \quad (48.149a)$$

$$G(\mathbf{x}, t < t_0|\mathbf{x}_0, t_0) = 0 \quad (48.149b)$$

$$\hat{\mathbf{n}}_{\mathbf{x}} \cdot \mathbb{K} \cdot \nabla_{\mathbf{x}} G(\mathbf{x} \in \partial \mathcal{R}, t|\mathbf{x}_0, t_0) = 0 \quad \text{or} \quad G(\mathbf{x} \in \partial \mathcal{R}, t|\mathbf{x}_0, t_0) = 0. \quad (48.149c)$$

The space-time point (\mathbf{x}_0, t_0) is where the Dirac delta source is located, which is located within the spatial domain, \mathcal{R} , and occurs after the initial time

$$\mathbf{x}_0 \in \mathcal{R} \quad \text{and} \quad t_0 \geq t_{\text{init}}. \quad (48.150)$$

The Green's function satisfies homogeneous boundary conditions corresponding to those satisfied by the passive tracer concentration in equation (48.148c). Finally, since the Dirac delta source, $\delta(\mathbf{x} - \mathbf{x}_0) \delta(t - t_0)$, has dimensions of inverse volume times inverse time, the Green's function has dimensions of inverse mass. We physically interpret the Green's function as the tracer concentration resulting from an impulsive tracer concentration source, divided by the mass of tracer injected by the source.

Adjoint Green's function problem

The adjoint Green's function, \tilde{G} , satisfies the adjoint problem

$$-\frac{\partial[\rho \tilde{G}(\mathbf{x}, t|\mathbf{x}_0, t_0)]}{\partial t} + \nabla_{\mathbf{x}} \cdot [-\rho \mathbf{v}^\dagger \tilde{G}(\mathbf{x}, t|\mathbf{x}_0, t_0) - \rho \mathbb{K} \cdot \nabla_{\mathbf{x}} \tilde{G}(\mathbf{x}, t|\mathbf{x}_0, t_0)] = \delta(\mathbf{x} - \mathbf{x}_0) \delta(t - t_0) \quad (48.151a)$$

$$\tilde{G}(\mathbf{x}, t > t_0|\mathbf{x}_0, t_0) = 0 \quad (48.151b)$$

$$\hat{\mathbf{n}}_{\mathbf{x}} \cdot \mathbb{K} \cdot \nabla_{\mathbf{x}} \tilde{G}(\mathbf{x} \in \partial\mathcal{R}, t|\mathbf{x}_0, t_0) = 0 \text{ or } \tilde{G}(\mathbf{x} \in \partial\mathcal{R}, t|\mathbf{x}_0, t_0) = 0. \quad (48.151c)$$

Note the sign change on both the time derivative, as for the diffusion equation in Section 4.7.3, as well as the advection term. The sign change on advection is expected since with time running backwards, so too does the velocity of a fluid particle. Hence, the adjoint advection-diffusion equation is a backwards in time advection-diffusion equation.

48.10.5 Reciprocity relation

We here derive the reciprocity relation satisfied by the Green's function, G , and its adjoint, \tilde{G} . The derivation follows that in Section 4.5.4 for Poisson's equation and Section 4.7.4 for the diffusion equation. A new feature here arises from the advection operator, and another arises from allowing the domain boundary to be time dependent as occurs at the ocean free surface. We will see that the reciprocity relation for Neumann boundary conditions (48.149c) and (48.151c) does not “close” when there is mass transport across the ocean free surface (see equation (48.166) below). In contrast, the reciprocity relation closes with Dirichlet boundary conditions, taking the same form as for the diffusion equation in equation (4.129). We have more to say on this distinct behavior after its derivation.

Notation and setup

We consider the Green's function partial differential equation (48.149a) with a Dirac delta source $\delta(\mathbf{x} - \mathbf{x}_1) \delta(t - t_1)$, along with the adjoint Green's function equation (48.151a) with a Dirac delta source, $\delta(\mathbf{x} - \mathbf{x}_2) \delta(t - t_2)$, where both sources are within the spatial domain and both occur later than the initial time:

$$\mathbf{x}_1, \mathbf{x}_2 \in \mathcal{R} \quad \text{and} \quad t_{\text{init}} < t_1, t_2. \quad (48.152)$$

We also follow the approach in Section 4.7.4 for the diffusion equation by introducing the arbitrarily large time, T , such that $-T < t_1, t_2 < T$. As for the diffusion problem, T will drop out from the final expression for the tracer concentration. Additionally, to help ease notational clutter, we make use of the following shorthand where convenient

$$G(\mathbf{x}, t|\mathbf{x}_1, t_1) = G(1) \quad \text{and} \quad \tilde{G}(\mathbf{x}, t|\mathbf{x}_2, t_2) = \tilde{G}(2). \quad (48.153)$$

Cross-multiplication

Multiply the Green's function equation by $\tilde{G}(2)$ and the adjoint equation by $G(1)$ to find

$$\tilde{G}(2) \left(\partial_t[\rho G(1)] + \nabla_{\mathbf{x}} \cdot [\rho \mathbf{v}^\dagger G(1) - \rho \mathbb{K} \cdot \nabla_{\mathbf{x}} G(1)] \right) = \tilde{G}(2) \delta(\mathbf{x} - \mathbf{x}_1) \delta(t - t_1) \quad (48.154a)$$

$$G(1) \left(-\partial_t[\rho \tilde{G}(2)] + \nabla_{\mathbf{x}} \cdot [-\rho \mathbf{v}^\dagger \tilde{G}(2) - \rho \mathbb{K} \cdot \nabla_{\mathbf{x}} \tilde{G}(2)] \right) = G(1) \delta(\mathbf{x} - \mathbf{x}_2) \delta(t - t_2). \quad (48.154b)$$

In the following, we work from the left hand side of equation (48.154a) and bring the differential operators from $G(1)$ onto $\tilde{G}(2)$. The result of this movement will be equation (48.154b) plus some

extra terms whose form depends on the causality condition and boundary conditions. Integration over space and time will then render the reciprocity relation.

Self-adjointness of the generalized Laplacian operator

The generalized Laplacian operator term on the left hand side of equation (48.154a) can be written

$$\begin{aligned} & -\tilde{G}(2) \nabla_{\mathbf{x}} \cdot [\rho \mathbb{K} \cdot \nabla_{\mathbf{x}} G(1)] \\ & = \nabla_{\mathbf{x}} \cdot [-\tilde{G}(2) \rho \mathbb{K} \cdot \nabla_{\mathbf{x}} G(1) + G(1) \rho \mathbb{K} \cdot \nabla_{\mathbf{x}} \tilde{G}(2)] - G(1) \nabla_{\mathbf{x}} \cdot [\rho \mathbb{K} \cdot \nabla_{\mathbf{x}} \tilde{G}(2)]. \end{aligned} \quad (48.155)$$

A spatial integration of this equation over the region \mathcal{R} , and use of the homogeneous boundary conditions (either Neumann or Dirichlet) (48.148c) and (48.151c), eliminates the divergence term to reveal

$$\begin{aligned} & \int_{\mathcal{R}} \tilde{G}(\mathbf{x}, t | \mathbf{x}_2, t_2) \nabla_{\mathbf{x}} \cdot [\rho \mathbb{K} \cdot \nabla_{\mathbf{x}} G(\mathbf{x}, t | \mathbf{x}_1, t_1)] dV \\ & = \int_{\mathcal{R}} G(\mathbf{x}, t | \mathbf{x}_1, t_1) \nabla_{\mathbf{x}} \cdot [\rho \mathbb{K} \cdot \nabla_{\mathbf{x}} \tilde{G}(\mathbf{x}, t | \mathbf{x}_2, t_2)] dV. \end{aligned} \quad (48.156)$$

This equality proves that the generalized Laplacian operator with a symmetric diffusion tensor is self-adjoint, which is a result already encountered in Section 4.7.4 when discussing the diffusion equation. This result holds for either Neumann or Dirichlet boundary conditions.

Time derivative plus advection

Next write the time derivative and advection portion of equation (48.154a) as

$$\begin{aligned} & \tilde{G}(2) \left(\partial_t [\rho G(1)] + \nabla_{\mathbf{x}} \cdot [\rho \mathbf{v}^\dagger G(1)] \right) \\ & = \partial_t [\tilde{G}(2) \rho G(1)] + \nabla_{\mathbf{x}} \cdot [\tilde{G}(2) \rho \mathbf{v}^\dagger G(1)] - G(1) \rho \left[\partial_t \tilde{G}(2) + \mathbf{v}^\dagger \cdot \nabla_{\mathbf{x}} \tilde{G}(2) \right] \\ & = \partial_t [\tilde{G}(2) \rho G(1)] + \nabla_{\mathbf{x}} \cdot [\tilde{G}(2) \rho \mathbf{v}^\dagger G(1)] - G(1) \left(\partial_t (\rho \tilde{G}(2)) + \nabla \cdot [\rho \mathbf{v}^\dagger \tilde{G}(2)] \right), \end{aligned} \quad (48.157)$$

where we used the mass continuity equation (15.6) for the final equality. Rearrangement thus leads to

$$\begin{aligned} & \tilde{G}(2) \left(\partial_t [\rho G(1)] + \nabla_{\mathbf{x}} \cdot [\rho \mathbf{v}^\dagger G(1)] \right) - G(1) \left(-\partial_t [\rho \tilde{G}(2)] - \nabla \cdot [\rho \mathbf{v}^\dagger \tilde{G}(2)] \right) \\ & = \partial_t [\tilde{G}(2) \rho G(1)] + \nabla_{\mathbf{x}} \cdot [\tilde{G}(2) \rho \mathbf{v}^\dagger G(1)]. \end{aligned} \quad (48.158)$$

Space integration

Now integrate equations (48.154a) and (48.154b) over the spatial domain, \mathcal{R} , subtract these two equations, and make use of the results (48.156) and (48.158) to reveal

$$\begin{aligned} & \tilde{G}(\mathbf{x}_1, t | \mathbf{x}_2, t_2) \delta(t - t_1) - G(\mathbf{x}_2, t | \mathbf{x}_1, t_1) \delta(t - t_2) \\ & = \int_{\mathcal{R}} \left[\partial_t [\tilde{G}(2) \rho G(1)] + \nabla_{\mathbf{x}} \cdot [\tilde{G}(2) \rho \mathbf{v}^\dagger G(1)] \right] dV. \end{aligned} \quad (48.159)$$

The divergence term on the right hand side takes the form

$$\int_{\mathcal{R}} \nabla_{\mathbf{x}} \cdot [\tilde{G}(2) \rho \mathbf{v}^\dagger G(1)] dV = \oint_{\partial\mathcal{R}} \tilde{G}(2) G(1) \rho \mathbf{v}^\dagger \cdot \hat{\mathbf{n}}_{\mathbf{x}} dS \quad \text{divergence theorem} \quad (48.160a)$$

$$= \oint_{\partial\mathcal{R}} \tilde{G}(2) G(1) \rho \mathbf{v} \cdot \hat{\mathbf{n}}_{\mathbf{x}} dS \quad \hat{\mathbf{n}} \cdot \mathbf{v}^* = 0 \quad (48.160b)$$

$$= \int_{z=\eta} \tilde{G}(2) G(1) \rho \mathbf{v} \cdot \hat{\mathbf{n}}_{\mathbf{x}} dS \quad \mathbf{v} \cdot \hat{\mathbf{n}} = 0 \text{ for } z = \eta_b \quad (48.160c)$$

$$= \int_{z=\eta} \tilde{G}(2) G(1) \rho \mathbf{v} \cdot \nabla(z - \eta) dA \quad \text{equation (15.76)} \quad (48.160d)$$

$$= \int_{z=\eta} \tilde{G}(2) G(1) \rho (w - \mathbf{u} \cdot \nabla \eta) dA. \quad (48.160e)$$

The time derivative term takes the form

$$\int_{\mathcal{R}} \partial_t [\tilde{G}(2) G(1) \rho] dV = \frac{\partial}{\partial t} \left[\int_{\mathcal{R}} \tilde{G}(2) G(1) \rho dV \right] - \int_{z=\eta} [\tilde{G}(2) G(1) \rho \partial_t \eta] dA, \quad (48.161)$$

where we made use of Leibniz's rule to bring the time derivative across the integral sign and made note of the time dependent free surface, $z = \eta(x, y, t)$. Combining equations (48.161) and (48.160e) leads to

$$\tilde{G}(\mathbf{x}_1, t | \mathbf{x}_2, t_2) \delta(t - t_1) - G(\mathbf{x}_2, t | \mathbf{x}_1, t_1) \delta(t - t_2) \quad (48.162a)$$

$$= \frac{\partial}{\partial t} \left[\int_{\mathcal{R}} \tilde{G}(2) G(1) \rho dV \right] + \int_{z=\eta} \tilde{G}(2) G(1) [\rho (w - \mathbf{u} \cdot \nabla \eta - \partial_t \eta)] dA \quad (48.162b)$$

$$= \frac{\partial}{\partial t} \left[\int_{\mathcal{R}} \tilde{G}(2) G(1) \rho dV \right] - \int_{z=\eta} \tilde{G}(2) G(1) Q_m dA, \quad (48.162c)$$

where the final equality follows from the surface ocean kinematic boundary condition (15.79), with Q_m the mass per time per horizontal area crossing the ocean surface.

Time integration

We are now ready to integrate equation (48.162c) over time, with its left hand side leading to

$$\begin{aligned} & \int_{-T}^T \left[\tilde{G}(\mathbf{x}_1, t | \mathbf{x}_2, t_2) \delta(t - t_1) - G(\mathbf{x}_2, t | \mathbf{x}_1, t_1) \delta(t - t_2) \right] dt \\ &= \tilde{G}(\mathbf{x}_1, t_1 | \mathbf{x}_2, t_2) - G(\mathbf{x}_2, t_2 | \mathbf{x}_1, t_1), \end{aligned} \quad (48.163)$$

which used the sifting property (4.28). There are two terms that appear when time integrating the time derivative from equation (48.162c), with each term vanishing due to the causality conditions (48.149b) and (48.151b)

$$\int_{\mathcal{R}} \tilde{G}(\mathbf{x}, t = T | \mathbf{x}_2, t_2) G(\mathbf{x}, t = T | \mathbf{x}_1, t_1) \rho dV = 0 \iff \tilde{G}(\mathbf{x}, t = T | \mathbf{x}_2, t_2) = 0 \quad (48.164)$$

$$\int_{\mathcal{R}} \tilde{G}(\mathbf{x}, t = t_{\text{init}} | \mathbf{x}_2, t_2) G(\mathbf{x}, t = t_{\text{init}} | \mathbf{x}_1, t_1) \rho dV = 0 \iff G(\mathbf{x}, t = t_{\text{init}} | \mathbf{x}_1, t_1) = 0. \quad (48.165)$$

We are thus left with

$$\tilde{G}(\mathbf{x}_1, t_1 | \mathbf{x}_2, t_2) - G(\mathbf{x}_2, t_2 | \mathbf{x}_1, t_1) = - \int_{t_{\text{init}}}^{t_1} \left[\int_{z=\eta} \tilde{G}(\mathbf{x}, t | \mathbf{x}_2, t_2) G(\mathbf{x}_1, t_1 | \mathbf{x}, t) Q_m dA \right] dt, \quad (48.166)$$

which we refer to as a *non-closed reciprocity relation* between G and \tilde{G} . Note that the time limits on the integral follow from causality on the Green's function and its adjoint.

Closed form reciprocity in special cases

There are two cases in which the relation (48.166) leads to a closed reciprocity relation:

- Zero mass flux across surface: $Q_m = 0$.
- Homogeneous Dirichlet boundary conditions: $\tilde{G}(\mathbf{x} \in \partial\mathcal{R}, t | \mathbf{x}_0, t_0) = G(\mathbf{x} \in \partial\mathcal{R}, t | \mathbf{x}_0, t_0) = 0$.

In either case we are led to

$$\tilde{G}(\mathbf{x}_1, t_1 | \mathbf{x}_2, t_2) = G(\mathbf{x}_2, t_2 | \mathbf{x}_1, t_1), \quad (48.167)$$

which is the same reciprocity (4.129) satisfied for the diffusion equation Green's functions.

The more nuanced reciprocity for the advection-diffusion equation arises from the advective mass flux at the ocean free surface boundary. The mass flux couples the ocean with its surrounding media (e.g., the atmosphere, rivers, or cryosphere), and in so doing precludes a general closed reciprocity relation. However, the Dirichlet boundary condition closes the surface boundary through the homogeneous Green's function boundary conditions. Most applications of Green's function methods for passive ocean tracers make use of Dirichlet boundary conditions, in which case we are afforded a closed reciprocity relation even with a free surface open to mass transport.

48.10.6 Integral expression for the tracer concentration

We are now ready to express the passive tracer concentration, C , as a suite of integrals involving the Green's function and the known boundary and initial conditions as well as the known source function. The process for deriving this expression is identical to that used in Section 48.10.5 for reciprocity, with the following steps offered for completeness.

Derivation setup

The initial-boundary value problem for the passive tracer is given by

$$\frac{\partial(\rho C)}{\partial t} + \nabla \cdot (\rho \mathbf{v}^\dagger C - \rho \mathbb{K} \cdot \nabla C) = \rho \Lambda \quad \mathbf{x} \in \mathcal{R}, t \geq t_{\text{init}} \quad (48.168a)$$

$$\rho C = \rho I \quad \mathbf{x} \in \mathcal{R}, t = t_{\text{init}} \quad (48.168b)$$

$$\rho C = \rho \sigma \quad \mathbf{x} \in \partial\mathcal{R}, t \geq t_{\text{init}} \quad (48.168c)$$

where we only consider the Dirichlet boundary condition to ensure a closed reciprocity relation in the presence of surface mass fluxes (Section 48.10.5). The corresponding adjoint Green's function satisfies

$$-\frac{\partial[\rho \tilde{G}(\mathbf{x}, t | \mathbf{x}_0, t_0)]}{\partial t} + \nabla_{\mathbf{x}} \cdot [-\rho \mathbf{v}^\dagger \tilde{G}(\mathbf{x}, t | \mathbf{x}_0, t_0) - \rho \mathbb{K} \cdot \nabla_{\mathbf{x}} \tilde{G}(\mathbf{x}, t | \mathbf{x}_0, t_0)] = \delta(\mathbf{x} - \mathbf{x}_0) \delta(t - t_0) \quad (48.169a)$$

$$\tilde{G}(\mathbf{x}, t > t_0 | \mathbf{x}_0, t_0) = 0 \quad (48.169b)$$

$$\tilde{G}(\mathbf{x} \in \partial\mathcal{R}, t | \mathbf{x}_0, t_0) = 0, \quad (48.169c)$$

with the reciprocity condition (48.167) holding since we chose Dirichlet boundary conditions. Multiplying the adjoint Green's function equation (48.169a) by $C(\mathbf{x}, t)$ and performing manipulations

just like those for reciprocity leads to

$$\begin{aligned} -\partial_t(\rho C \tilde{G}) + \nabla_{\mathbf{x}} \cdot [\tilde{G} \rho \mathbb{K} \cdot \nabla C - C \rho \mathbb{K}(\mathbf{x}, t) \cdot \nabla_{\mathbf{x}} \tilde{G} - C \rho \mathbf{v}^\dagger \tilde{G}] + \tilde{G} \rho \Lambda \\ = C(\mathbf{x}, t) \delta(\mathbf{x} - \mathbf{x}_0) \delta(t - t_0). \quad (48.170) \end{aligned}$$

With the homogeneous Dirichlet conditions satisfied by \tilde{G} on the spatial boundaries, a space and time integration over (\mathbf{x}, t) leads to

$$\begin{aligned} C(\mathbf{x}_0, t_0) &= \int_{\mathcal{R}} \tilde{G}(\mathbf{x}, t_{\text{init}} | \mathbf{x}_0, t_0) \rho(\mathbf{x}, t_{\text{init}}) I(\mathbf{x}) dV + \int_{t_{\text{init}}}^{t_0} \left[\int_{\mathcal{R}} \tilde{G}(\mathbf{x}, t | \mathbf{x}_0, t_0) \rho(\mathbf{x}, t) \Lambda(\mathbf{x}, t) dV \right] dt \\ &\quad - \int_{t_{\text{init}}}^{t_0} \left[\oint_{\partial \mathcal{R}} \sigma(\mathbf{x}, t) \rho(\mathbf{x}, t) \mathbb{K}(\mathbf{x}, t) \cdot \nabla_{\mathbf{x}} \tilde{G}(\mathbf{x}, t | \mathbf{x}_0, t_0) \cdot \hat{\mathbf{n}}_{\mathbf{x}} dS \right] dt. \quad (48.171) \end{aligned}$$

Use of the reciprocity relation (48.167) allows us to write this equation in terms of the Green's function rather than the adjoint Green's function

$$\begin{aligned} C(\mathbf{x}_0, t_0) &= \int_{\mathcal{R}} G(\mathbf{x}_0, t_0 | \mathbf{x}, t_{\text{init}}) \rho(\mathbf{x}, t_{\text{init}}) I(\mathbf{x}) dV + \int_{t_{\text{init}}}^{t_0} \left[\int_{\mathcal{R}} G(\mathbf{x}_0, t_0 | \mathbf{x}, t) \rho(\mathbf{x}, t) \Lambda(\mathbf{x}, t) dV \right] dt \\ &\quad - \int_{t_{\text{init}}}^{t_0} \left[\oint_{\partial \mathcal{R}} \sigma(\mathbf{x}, t) \rho(\mathbf{x}, t) \mathbb{K}(\mathbf{x}, t) \cdot \nabla_{\mathbf{x}} G(\mathbf{x}_0, t_0 | \mathbf{x}, t) \cdot \hat{\mathbf{n}}_{\mathbf{x}} dS \right] dt. \quad (48.172) \end{aligned}$$

Finally, swapping labels $(\mathbf{x}_0, t_0) \leftrightarrow (\mathbf{x}, t)$ renders

$$\begin{aligned} C(\mathbf{x}, t) &= \int_{\mathcal{R}} G(\mathbf{x}, t | \mathbf{x}_0, t_{\text{init}}) \rho(\mathbf{x}_0, t_{\text{init}}) I(\mathbf{x}_0) dV_0 + \int_{t_{\text{init}}}^t \left[\int_{\mathcal{R}} G(\mathbf{x}, t | \mathbf{x}_0, t_0) \rho(\mathbf{x}_0, t_0) \Lambda(\mathbf{x}_0, t_0) dV_0 \right] dt_0 \\ &\quad - \int_{t_{\text{init}}}^t \left[\oint_{\partial \mathcal{R}} \sigma(\mathbf{x}_0, t_0) \rho(\mathbf{x}_0, t_0) \mathbb{K}(\mathbf{x}_0, t_0) \cdot \nabla_{\mathbf{x}_0} G(\mathbf{x}, t | \mathbf{x}_0, t_0) \cdot \hat{\mathbf{n}}_{\mathbf{x}_0} dS_0 \right] dt_0. \quad (48.173) \end{aligned}$$

This solution manifests causality since the concentration at time t is a function only of processes occurring from t_{init} up to time t .

48.10.7 Properties of the solution

The integral solution (48.173) is of the same form as equation (4.141) for the diffusion equation. Properties of this solution, and corresponding properties of the Green's function, follow from those satisfied by the diffusion equation as detailed in Section 4.7.6. We here summarize these properties for completeness.

The role of advection and diffusion at boundaries

Explicit contributions from the advective flux are absent from the solution (48.173). Namely, there are no advective flux contributions at the surface boundary due to the homogeneous Dirichlet boundary conditions imposed on the Green's function. For the ocean bottom, material and rigid no-flux conditions mean that $\mathbf{v} \cdot \hat{\mathbf{n}} = 0$ at the bottom. The presence of advection arises only through its effect on the Green's function, which is affected by both advection and diffusion.

Furthermore, notice how in the absence of diffusion (i.e., $\mathbb{K} = 0$) the Dirichlet boundary data is unable to penetrate into the ocean interior since the surface boundary integral vanishes from equation (48.173). In effect, the surface boundary becomes a material surface when there is no diffusion. That is, diffusive mixing is needed for boundary data to move into the interior. This role for diffusion was also identified when studying the surface flux condition for salt and freshwater in Section 51.3.

Initial conditions

When sampling the tracer concentration at the initial time, $t \rightarrow t_{\text{init}}$, all the time integrals vanish from the solution (48.173), thus leaving

$$\lim_{t \rightarrow t_{\text{init}}} C(\mathbf{x}, t) = \lim_{t \rightarrow t_{\text{init}}} \int_{\mathcal{R}} G(\mathbf{x}, t | \mathbf{x}_0, t_{\text{init}}) \rho(\mathbf{x}_0, t_{\text{init}}) I(\mathbf{x}_0) dV_0. \quad (48.174)$$

Self-consistency implies that the Green's function satisfies

$$\lim_{t \rightarrow t_{\text{init}}} G(\mathbf{x}, t | \mathbf{x}_0, t_{\text{init}}) \rho(\mathbf{x}_0, t_{\text{init}}) = \delta(\mathbf{x} - \mathbf{x}_0) \quad \text{with } \mathbf{x}, \mathbf{x}_0 \in \mathcal{R}, \quad (48.175)$$

so that

$$\lim_{t \rightarrow t_{\text{init}}} \int_{\mathcal{R}} \rho(\mathbf{x}, t_{\text{init}}) G(\mathbf{x}, t | \mathbf{x}_0, t_{\text{init}}) I(\mathbf{x}_0) dV_0 = \int_{\mathcal{R}} \delta(\mathbf{x} - \mathbf{x}_0) I(\mathbf{x}_0) dV_0 = I(\mathbf{x}). \quad (48.176)$$

Dirichlet boundary conditions

Evaluating the Dirichlet solution (48.173) on a spatial boundary, $\mathbf{x} \in \partial\mathcal{R}$, eliminates both the volume integrals given that the Green's function satisfies homogeneous Dirichlet boundary conditions. The tracer concentration (48.173) thus takes the form

$$C(\mathbf{x}, t) = - \int_{t_{\text{init}}}^t \left[\oint_{\partial\mathcal{R}} \sigma(\mathbf{x}_0, t_0) \rho(\mathbf{x}_0, t_0) \mathbb{K}(\mathbf{x}_0, t_0) \cdot \nabla_{\mathbf{x}_0} G(\mathbf{x}, t | \mathbf{x}_0, t_0) \cdot \hat{\mathbf{n}}_{\mathbf{x}_0} d\mathcal{S}_0 \right] dt_0 \quad \text{with } \mathbf{x} \in \partial\mathcal{R}. \quad (48.177)$$

Self-consistency with the Dirichlet boundary condition (48.148c) implies that the Green's function, when both spatial points are evaluated on the boundary, satisfies

$$\rho(\mathbf{x}_0, t_0) \mathbb{K}(\mathbf{x}_0, t_0) \cdot \nabla_{\mathbf{x}_0} G(\mathbf{x}, t | \mathbf{x}_0, t_0) \cdot \hat{\mathbf{n}}_{\mathbf{x}_0} = -\delta(t - t_0) \delta^{(2)}(\mathbf{x} - \mathbf{x}_0) \quad \text{with } \mathbf{x}, \mathbf{x}_0 \in \partial\mathcal{R}, \quad (48.178)$$

so that

$$C(\mathbf{x}, t) = \int_{t_{\text{init}}}^t \left[\oint_{\partial\mathcal{R}} \sigma(\mathbf{x}_0, t_0) \delta(t - t_0) \delta^{(2)}(\mathbf{x} - \mathbf{x}_0) d\mathcal{S}_0 \right] dt_0 = \sigma(\mathbf{x}, t) \quad \text{with } \mathbf{x} \in \partial\mathcal{R}. \quad (48.179)$$

48.10.8 Boundary propagator

Defining the boundary propagator

As for the diffusion equation in Section 4.7.7, we here introduce the boundary propagator for the advection-diffusion equation with Dirichlet boundary conditions. For this purpose, consider the special case of a passive tracer with zero interior source and with zero initial condition, thus satisfying the initial-boundary value problem

$$\frac{\partial(\rho C)}{\partial t} + \nabla \cdot (\rho \mathbf{v}^\dagger C - \rho \mathbb{K} \cdot \nabla C) = 0 \quad \mathbf{x} \in \mathcal{R}, t \geq t_{\text{init}} \quad (48.180a)$$

$$\rho C = 0 \quad \mathbf{x} \in \mathcal{R}, t = t_{\text{init}} \quad (48.180b)$$

$$\rho C = \rho \sigma \quad \mathbf{x} \in \partial\mathcal{R}, t \geq t_{\text{init}}, \quad (48.180c)$$

which leads to the simplification of the Green's function solution (48.173)

$$C(\mathbf{x}, t) = - \int_{t_{\text{init}}}^t \left[\oint_{\partial\mathcal{R}} \sigma(\mathbf{x}_0, t_0) \rho(\mathbf{x}_0, t_0) \mathbb{K}(\mathbf{x}_0, t_0) \cdot \nabla_{\mathbf{x}_0} G(\mathbf{x}, t | \mathbf{x}_0, t_0) \cdot \hat{\mathbf{n}}_{\mathbf{x}_0} d\mathcal{S}_0 \right] dt_0. \quad (48.181)$$

The tracer concentration at a point in space-time is determined by the history of the advection and diffusion that transfers boundary information to this point. To manifest this cause-effect relation, it is useful to define the boundary propagator just as for the diffusion equation

$$G^{\text{bp}}(\mathbf{x}, t | \mathbf{x}_0, t_0) \equiv -\rho(\mathbf{x}_0, t_0) \mathbb{K}(\mathbf{x}_0, t_0) \cdot \nabla_{\mathbf{x}_0} G(\mathbf{x}, t | \mathbf{x}_0, t_0) \cdot \hat{\mathbf{n}}_{\mathbf{x}_0} \quad \text{with } \mathbf{x}_0 \in \partial\mathcal{R}, \quad (48.182)$$

with G^{bp} having dimensions $L^{-2} T^{-1}$. The boundary propagator thus brings the tracer concentration (48.181) into the rather tidy form

$$C(\mathbf{x}, t) = \int_{t_{\text{init}}}^t \left[\oint_{\partial\mathcal{R}} \sigma(\mathbf{x}_0, t_0) G^{\text{bp}}(\mathbf{x}, t | \mathbf{x}_0, t_0) d\mathcal{S}_0 \right] dt_0. \quad (48.183)$$

Inhomogeneous Dirichlet at the surface and homogeneous Neumann at the bottom

In applications of passive tracers to study ocean circulation, it is common to apply inhomogeneous Dirichlet boundary conditions just at the ocean surface, and homogeneous Neumann boundary conditions (no-flux) at the ocean bottom

$$\frac{\partial(\rho C)}{\partial t} + \nabla \cdot (\rho \mathbf{v}^\dagger C - \rho \mathbb{K} \cdot \nabla C) = 0 \quad \mathbf{x} \in \mathcal{R}, t \geq t_{\text{init}} \quad (48.184a)$$

$$C = 0 \quad \mathbf{x} \in \mathcal{R}, t = t_{\text{init}} \quad (48.184b)$$

$$\rho C = \rho \sigma \quad \mathbf{x} \in \partial\mathcal{R}_{\text{surface}}, t \geq t_{\text{init}} \quad (48.184c)$$

$$\hat{\mathbf{n}}_{\mathbf{x}} \cdot \mathbb{K} \cdot \nabla_{\mathbf{x}} C = 0 \quad \mathbf{x} \in \partial\mathcal{R}_{\text{bottom}}, t \geq t_{\text{init}}. \quad (48.184d)$$

Note that since $\hat{\mathbf{n}} \cdot \mathbf{v} = 0$ at the solid earth ocean bottom, kinematics imposes no advective flux through the bottom, $\hat{\mathbf{n}} \cdot \mathbf{v} C = 0$. Since the bottom boundary conditions are homogeneous, the solution (48.181) also holds for the initial-boundary value problem (48.184a)-(48.184d). The key distinction, however, is that the Green's function now satisfies the following boundary value problem

$$\frac{\partial[\rho G(\mathbf{x}, t | \mathbf{x}_0, t_0)]}{\partial t} + \nabla_{\mathbf{x}} \cdot [\rho \mathbf{v}^\dagger G(\mathbf{x}, t | \mathbf{x}_0, t_0) - \rho \mathbb{K} \cdot \nabla_{\mathbf{x}} G(\mathbf{x}, t | \mathbf{x}_0, t_0)] = \delta(\mathbf{x} - \mathbf{x}_0) \delta(t - t_0) \quad (48.185a)$$

$$G(\mathbf{x}, t < t_0 | \mathbf{x}_0, t_0) = 0 \quad (48.185b)$$

$$G(\mathbf{x}, t | \mathbf{x}_0, t_0) = 0 \quad \mathbf{x} \in \partial\mathcal{R}_{\text{surface}} \quad (48.185c)$$

$$\hat{\mathbf{n}}_{\mathbf{x}} \cdot \mathbb{K} \cdot \nabla_{\mathbf{x}} G(\mathbf{x}, t | \mathbf{x}_0, t_0) = 0 \quad \mathbf{x} \in \partial\mathcal{R}_{\text{bottom}}. \quad (48.185d)$$

Boundary value problem for the boundary propagator

Following the more detailed presentation in Section 4.7.7 for the diffusion equation, we are led to the following boundary value problem satisfied by the boundary propagator

$$\frac{\partial[\rho G^{\text{bp}}(\mathbf{x}, t | \mathbf{x}_0, t_0)]}{\partial t} + \nabla_{\mathbf{x}} \cdot [\rho \mathbf{v}^\dagger G^{\text{bp}}(\mathbf{x}, t | \mathbf{x}_0, t_0) - \rho \mathbb{K} \cdot \nabla_{\mathbf{x}} G^{\text{bp}}(\mathbf{x}, t | \mathbf{x}_0, t_0)] = 0, \quad \mathbf{x} \in \mathcal{R} \quad (48.186a)$$

$$G^{\text{bp}}(\mathbf{x}, t | \mathbf{x}_0, t_0) = 0, \quad \mathbf{x} \notin \partial\mathcal{R}, t \leq t_0 \quad (48.186b)$$

$$G^{\text{bp}}(\mathbf{x}, t | \mathbf{x}_0, t_0) = \delta(t - t_0) \delta^{(2)}(\mathbf{x} - \mathbf{x}_0), \quad \mathbf{x}, \mathbf{x}_0 \in \partial\mathcal{R}. \quad (48.186c)$$

The boundary propagator acts as the mediator between boundary data, σ , and interior points, with the transfer of information realized through both advection and diffusion. A focus on the boundary propagator rather than the Green's function allows us to dispense with the need to compute the normal gradient of the Green's function at the boundary, with that calculation rather awkward in practice. Also recall our discussion in Section 4.8.5, where we argued that the boundary propagator can be considered the impulse response function for spatially distributed sources. Here, the mediation of the Dirac boundary sources is performed by advection plus diffusion, whereas in Section 4.8 we only considered linear damping and diffusion.

Normalization of the boundary propagator

As seen in Sections 48.3 and 48.4, diffusion acts to smooth all structure in the tracer field. Hence, if the boundary data is a uniform constant, $\sigma = \sigma_{\text{const}}$, then given sufficient time the tracer concentration will equal to this constant, $C = \sigma_{\text{const}}$. This steady state result is independent of details for the velocity field and for the diffusivity tensor, with details of advection and the diffusivity acting only to modify the time scale for the equilibration. Assuming we wait long enough, or equivalently that the initial condition occurs infinitely far in the past, then the tracer concentration solution (48.183) leads to the normalization of the boundary propagator

$$\lim_{t_{\text{init}} \rightarrow -\infty} \int_{t_{\text{init}}}^t \left[\oint_{\partial\mathcal{R}} G^{\text{bp}}(\mathbf{x}, t | \mathbf{x}_0, t_0) d\mathcal{S}_0 \right] dt_0 = 1 \quad \text{for } \mathbf{x} \in \mathcal{R}. \quad (48.187)$$

This normalization holds for all field points, \mathbf{x} , within the region. Even though this condition was derived by assuming the special case of constant boundary data, it holds in general since the Green's function, and by extension the boundary propagator, are independent of the boundary data prescribed for the tracer concentration.

48.11 Exercises

EXERCISE 48.1: VERTICAL DIFFUSION OF TEMPERATURE IN THE OCEAN ([Vallis, 2017](#))

There is a natural time scale associated with diffusive transport. This time scale can be found from scaling the diffusion equation, which reveals that it takes the form

$$\tau_{\text{diffusion}} = \frac{\Delta^2}{\kappa} \quad (48.188)$$

where Δ is the length scale and κ is the kinematic diffusivity (dimensions of squared length per time). We now make use of this time scale to consider the diffusion of temperature in the ocean, with diffusion due solely to molecular processes.

Using the observed value of molecular diffusivity of temperature in water (look it up), estimate the time for a temperature anomaly to mix from the top of the ocean to the bottom, assuming vertical diffusion through the molecular diffusivity is the only means for mixing. This time scale follows from the one-dimensional diffusion equation and is determined by the diffusivity and the depth of the ocean. Comment on whether you think the real ocean has reached equilibrium after the last ice age (which ended about 12Kyr ago).

EXERCISE 48.2: ANALYTICAL SOLUTION TO ONE-DIMENSIONAL DIFFUSION EQUATION

Consider a one-dimensional diffusion equation

$$\frac{\partial C}{\partial t} = \kappa \frac{\partial^2 C}{\partial z^2}, \quad (48.189)$$

where C is a tracer concentration (e.g., temperature or salinity), κ is a constant kinematic diffusivity, and z is the vertical coordinate. Assume the domain has fixed boundaries at $z = 0$ and $z = H$.

- (a) Assume there is a zero flux of tracer at the two boundaries. Mathematically express this no-flux boundary condition.
- (b) Assume that the initial tracer concentration is confined to an area near the center of the domain. Use dimensional analysis to estimate the time scale for the concentration to homogenize throughout the domain.
- (c) Consider the initial-boundary value problem

$$\frac{\partial C}{\partial t} = \kappa \frac{\partial^2 C}{\partial z^2}, \quad (48.190a)$$

$$\text{no-flux boundary condition from part (b)} \quad (48.190b)$$

$$C(z, t = 0) = C_0 \cos(Kz), \quad (48.190c)$$

where C_0 is a constant. What values for the wave-number, K , satisfy the no-flux boundary condition?

- (d) Solve the diffusion equation analytically for the given initial condition. Hint: consult your favorite partial differential equation book to learn how to solve this linear 1+1 dimensional diffusion equation.
- (e) Explain how the analytical answer you obtained is consistent with the dimensional analysis answer from part (b).

EXERCISE 48.3: DISSIPATIVE PROPERTIES OF DIFFUSION

This exercise explores the dissipative property of diffusion when acting on a tracer extrema.

- (a) ONE-DIMENSIONAL DIFFUSION

Consider the diffusion equation in one spatial dimension, and assume a Boussinesq fluid in which case the density factors are all constant and so can be dropped

$$\frac{\partial C}{\partial t} = \frac{\partial}{\partial z} \left[\kappa \frac{\partial C}{\partial z} \right] \quad (48.191a)$$

$$= \frac{\partial \kappa}{\partial z} \frac{\partial C}{\partial z} + \kappa \frac{\partial^2 C}{\partial z^2}, \quad (48.191b)$$

where $\kappa(z, t)$ is an *eddy diffusivity* (also *turbulent diffusivity*). The eddy diffusivity is assumed to be a function of (z, t) , with the spatial dependence determined by the flow. Show that a tracer extrema, C^* , evolves under diffusion according to

$$\frac{\partial C^*}{\partial t} = \kappa \frac{\partial^2 C^*}{\partial z^2}. \quad (48.192)$$

So what does diffusion do to a local maxima (e.g., a local hot region) in the tracer field? What about a minima (e.g., a local cold region)?

(b) THREE-DIMENSIONAL DIFFUSION

Generalize the above one dimensional result to three dimensions, whereby the diffusivity κ becomes a symmetric positive-definite diffusion *tensor*, in which case

$$\frac{\partial C}{\partial t} = \frac{\partial}{\partial x^m} \left[\kappa^{mn} \frac{\partial C}{\partial x^n} \right]. \quad (48.193)$$

Now consider an extrema in the tracer field, which is defined by

$$\frac{\partial C^*}{\partial x^n} = 0 \quad \forall n = 1, 2, 3. \quad (48.194)$$

Prove that three dimensional diffusion acts to *dissipate* an extrema. Hint: recall some linear algebra properties of a symmetric positive-definite matrix. In particular, note that a symmetric positive-definite matrix has positive eigenvalues.

EXERCISE 48.4: DIFFUSION INCREASES INFORMATION ENTROPY OF A TRACER CONCENTRATION
 Diffusion is an irreversible process. Here we illustrate this property by considering the *information entropy* associated with a non-negative tracer concentration⁴

$$\mathcal{S}_C \equiv - \int (C \ln C) \rho \, dV. \quad (48.195)$$

Show that

$$\frac{d\mathcal{S}_C}{dt} \geq 0 \quad (48.196)$$

over a material region with $C > 0$ and diffusion downgradient, $\mathbf{J} \cdot \nabla C < 0$. That is, diffusion always increases the information entropy. Hint: follow the discussion of tracer moments in Section 48.4.3.

EXERCISE 48.5: ONE-DIMENSIONAL ADVECTION

Consider the advection equation in one space dimension without boundaries

$$\frac{\partial C}{\partial t} + u \frac{\partial C}{\partial x} = 0 \quad (48.197a)$$

$$C(x, z, t=0) = C_0 \cos(k x) \quad (48.197b)$$

$$u(z, t) = \alpha z \cos(\omega t). \quad (48.197c)$$

The specified zonal velocity is non-divergent, oscillatory in time, and vertically sheared

$$\frac{\partial u}{\partial z} = \alpha \cos(\omega t), \quad (48.198)$$

with ω the radial frequency of the temporal oscillations. What is the tracer concentration at times $t > 0$? Hint: make use of the exact solution given by equation (48.44).

EXERCISE 48.6: SKEW FLUX FOR OCEAN MESOSCALE EDDIES

Consider a middle-latitude mesoscale ocean eddy respecting geostrophic balance (see Section 26.4) on an f -plane (constant Coriolis parameter) and incompressibility. In this case, the horizontal eddy-induced velocity at the ocean surface is non-divergent

$$\mathbf{u}^* = \nabla \wedge \hat{\mathbf{z}} \psi. \quad (48.199)$$

⁴Information entropy is used in statistical physics as a measure of the order/disorder of a probability distribution. We here apply these notions to measure the information entropy of a tracer concentration.

In this equation, the geostrophic streamfunction is given by

$$\psi = -\hat{z} \frac{g \eta}{f}, \quad (48.200)$$

with f the Coriolis parameter, g the gravitational acceleration, and η the sea level undulation associated with the eddy. Since the fluid is incompressible, the mass transport equals to the volume transport times a constant reference density, ρ_0 .

- (a) Determine the skew diffusion tensor (48.66).
- (b) Determine the skew tracer flux (48.79a).

EXERCISE 48.7: INTEGRATION BETWEEN TWO CLOSED TRACER CONTOURS

This exercise introduces some ideas of use for determining processes affecting the transport of matter across a tracer contour. Note that in general, the tracer concentration is a function of time. However, the present suite of questions concerns the instantaneous geometry of the tracer field, so that time dependence is not considered.

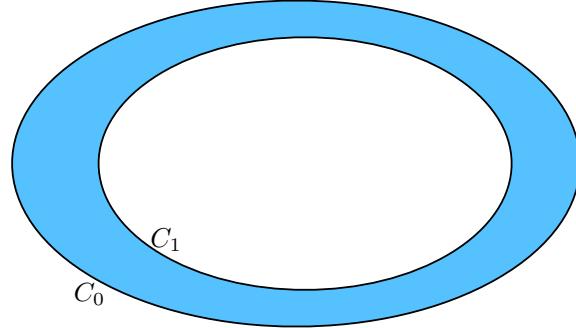


FIGURE 48.5: Illustrating the area contained between two closed tracer contours, $C_0 \leq C(x, y, t) \leq C_1$. Exercise 48.7 develops some mathematical expressions for integration within this area, with the resulting expressions of use for the analyses of tracer transport.

- (a) Consider a closed two-dimensional region bounded by two contours of tracer concentration, $C_0 \leq C(x, y, t) \leq C_1$, such as shown in Figure 48.5. Derive the following expression for the area enclosed by the two contours

$$\mathcal{A} = \int_{C_0}^{C_1} dC \oint \frac{dl}{|\nabla C|}. \quad (48.201)$$

In this expression, dl is the line element for a path taken in a counter-clockwise direction along a contour of constant C . We also assume the tracer concentration is not uniform in the region of interest so that $\nabla C \neq 0$.

- (b) As a corollary, show that for

$$\mathcal{A}(C) = \int_{C_0}^C dC' \oint \frac{dl}{|\nabla C'|} \quad (48.202)$$

we have the identity

$$\frac{\partial \mathcal{A}(C)}{\partial C} = \oint \frac{dl}{|\nabla C|}. \quad (48.203)$$

In words, this result means that the area between two tracer contours has a partial derivative, with respect to the tracer contour, equal to the line integral on the right hand side. The area per C is smaller in regions where the tracer gradient is larger; i.e., there is less area “concentration” in regions of strong tracer gradient.

- (c) Use the above two results to prove the following form of the Fundamental Theorem of Calculus

$$\frac{\partial}{\partial C} \left[\int \Phi(\mathbf{x}) d\mathcal{A} \right] = \frac{\partial}{\partial C} \left[\int_{C_0}^C dC' \oint \frac{\Phi dl}{|\nabla C'|} \right] \quad (48.204a)$$

$$= \oint \frac{\Phi dl}{|\nabla C|}, \quad (48.204b)$$

with Φ an arbitrary function. This is a remarkable identity with many useful applications such as those in [Marshall et al. \(2006\)](#).

EXERCISE 48.8: EVOLUTION OF TRACER CENTER OF MASS IN A STATIC DOMAIN

The exercise introduces us to how the tracer center of mass evolves within a Boussinesq fluid. We define the tracer center of mass as

$$\langle \mathbf{x} \rangle^C = \frac{\int \mathbf{x} C dV}{\int C dV}, \quad (48.205)$$

with C the tracer concentration, \mathbf{x} the coordinate of a point in the fluid, and integration is over the full fluid domain. For example, with a spherically symmetric tracer cloud, the center of mass position is at the sphere’s center. The center of mass position is not necessarily where the largest tracer concentration sits, in the same way that the center of mass of a massive object is not necessarily where the object is most dense. For example, a hollow spherical shell has its center of mass at the center of the sphere, even though there is no mass there.

For this exercise, assume the fluid is within a domain whose static boundaries are either material (no normal component to the boundary flux) or periodic. Hence, the total fluid volume and total tracer content remain constant

$$V = \int dV \quad \text{and} \quad C = \int C dV. \quad (48.206)$$

Furthermore, when computing the time derivative acting on the integral, make use of the kinematic results from Section 16.3.3, in which for any integrand φ

$$\frac{d}{dt} \int \varphi dV = \int \frac{\partial \varphi}{\partial t} dV, \quad (48.207)$$

which follows since the region boundaries are assumed to be static. Equivalently, since the region under consideration is material (no matter crosses the boundaries), we can make use of Reynold’s transport theorem from Section 16.3.4 to write

$$\frac{d}{dt} \int \varphi dV = \int \frac{D\varphi}{Dt} dV. \quad (48.208)$$

- (a) Consider a tracer concentration whose tendency at a point in space is affected only by advection

$$\frac{DC}{Dt} = 0 \implies \frac{\partial C}{\partial t} + \nabla \cdot (\mathbf{v} C) = 0, \quad (48.209)$$

with \mathbf{v} a non-divergent velocity, $\nabla \cdot \mathbf{v} = 0$. Show that the tracer center of mass position evolves according to the tracer center of mass velocity

$$\frac{d\langle \mathbf{x} \rangle^C}{dt} = \langle \mathbf{v} \rangle^C, \quad (48.210)$$

where the tracer center of mass velocity is given by

$$\langle \mathbf{v} \rangle^C = \frac{\int \mathbf{v} C dV}{\int C dV} = \frac{1}{C} \int \mathbf{v} C dV. \quad (48.211)$$

- (b) Consider a tracer concentration whose tendency at a point in space affected only by diffusion

$$\frac{\partial C}{\partial t} = \nabla \cdot (K \cdot \nabla C), \quad (48.212)$$

where $K = K(\mathbf{x}, t) > 0$ is a kinematic diffusivity (physical dimensions of squared length per time), and which is assumed to vanish at the domain boundaries. Show that the tracer center of mass drifts up the diffusivity gradient

$$\frac{d\langle \mathbf{x} \rangle^C}{dt} = \langle \nabla K \rangle^C. \quad (48.213)$$

Hint: use the product rule and drop boundary terms.

- (c) Consider an initial tracer concentration that is a function only of latitude,

$$C(x, y, z, t = 0) = C_0(y), \quad (48.214)$$

and assume a smooth spherical domain. Assume the diffusivity, K , is a turbulent diffusivity proportional to the eddy kinetic energy of the flow, so that large diffusivity occurs in regions with large eddy activity; i.e., there is a lot of turbulent mixing where turbulence is active. Introduce an stirring from the eddies that breaks the zonal symmetry. Qualitatively discuss the process whereby this turbulent diffusive mixing causes the tracer center of mass to drift towards the turbulent region.

- (d) Generalize the result from part (b) to the case of the diffusion equation

$$\frac{\partial C}{\partial t} = \nabla \cdot (\mathbb{K} \cdot \nabla C) = \partial_p (\mathbb{K}^{pq} \partial_q C), \quad (48.215)$$

where \mathbb{K} is a second order symmetric diffusion tensor.

EXERCISE 48.9: EVOLUTION OF TRACER CENTER OF MASS IN MOVING REGION

Consider a finite region of fluid with fixed mass that is moving with the fluid velocity field, $\mathcal{R}(\mathbf{v})$. The fluid is assumed to have a tracer whose concentration is affected by an irreversible process so that

$$\frac{DC}{Dt} = \dot{C} \neq 0. \quad (48.216)$$

For example, \dot{C} may represent a diffusive process, in which case the tracer content within the region changes due to diffusion of tracer across the region boundary.

Determine the evolution equation for the tracer center of mass position

$$\langle \mathbf{x} \rangle^C = \frac{\int_{\mathcal{R}(\mathbf{v})} \mathbf{x} C \rho dV}{\int_{\mathcal{R}(\mathbf{v})} C \rho dV}. \quad (48.217)$$

Hint: the region under consideration is moving with the fluid and has constant mass. Although the region boundaries are not material, we can make use of Reynold's transport theorem from Section 16.3.5 since the region has a constant mass. Consequently, we can set

$$\frac{d}{dt} \int_{\mathcal{R}(\mathbf{v})} \psi \rho dV = \int_{\mathcal{R}(\mathbf{v})} \frac{D\psi}{Dt} \rho dV. \quad (48.218)$$

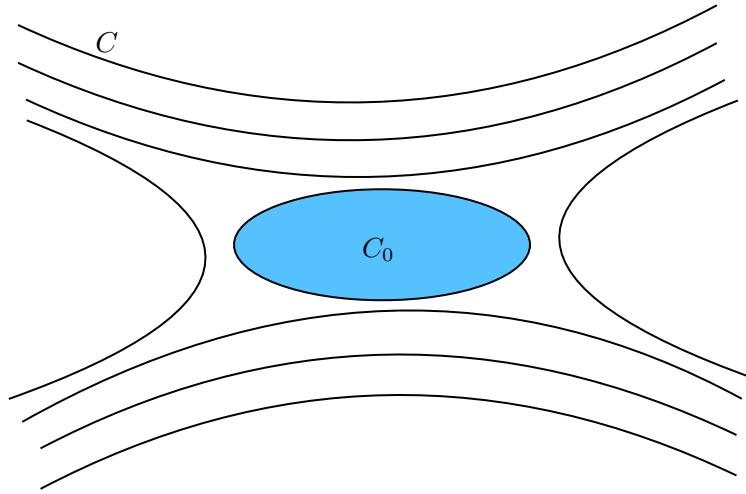


FIGURE 48.6: In a steady state flow, the tracer concentration within a constant C contour is constant. Diffusion has thus acted to remove all variations in tracer concentration within the region. In this figure, the concentration within the closed region has constant value $C = C_0$, whereas the region with open tracer contours remains non-homogeneous.

EXERCISE 48.10: DIFFUSIVE HOMOGENIZATION OF SCALARS INSIDE CLOSED CONTOURS

The advection-diffusion equation for a tracer concentration is given by

$$\frac{\partial(\rho C)}{\partial t} + \nabla \cdot (\rho \mathbf{v} C) = -\nabla \cdot \mathbf{J} \quad (48.219)$$

where

$$\mathbf{J} = -\rho \mathbb{K} \cdot \nabla C \quad (48.220)$$

is a downgradient diffusive flux with \mathbb{K} a symmetric positive-definite diffusion tensor. In the steady state, the divergence of the advective tracer flux balances the convergence of the diffusive flux

$$\nabla \cdot (\rho \mathbf{v} C) = -\nabla \cdot \mathbf{J}. \quad (48.221)$$

Consider a two-dimensional steady state flow and consider a region enclosed by a constant tracer contour. Prove that the tracer concentration is homogeneous (i.e., a spatially constant) within the contour of constant C , as shown in Figure 48.6. Hence, in the steady state, diffusion removes all tracer variations within closed tracer contours; i.e., there are no tracer extrema within a closed tracer contour.

Hint: make use of a *reductio ad absurdum* argument and study the material in Section 52.11.

EXERCISE 48.11: STEADY TWO DIMENSIONAL ADVECTION-DIFFUSION

Consider the steady state advection-diffusion equation for a scalar field, Q , in a two dimensional non-divergent flow

$$\nabla \cdot (\mathbf{u} Q) = \nabla \cdot (\mathbb{K} \cdot \nabla Q) \quad \text{with} \quad \mathbf{u} = \hat{\mathbf{z}} \wedge \nabla \psi, \quad (48.222)$$

and \mathbb{K} a diffusion tensor. Show that when evaluated along a contour of constant Q we can write

$$-(\hat{\mathbf{n}} \cdot \nabla Q)(\hat{\mathbf{t}} \cdot \nabla \psi) = \nabla \cdot (\mathbb{K} \cdot \nabla Q) \quad (48.223)$$

where $\hat{\mathbf{t}}$ is the unit tangent along the contour and $\hat{\mathbf{n}}$ is a unit vector pointing to the left of the tangent (e.g., see Figure 5.6). Assuming $\hat{\mathbf{n}} \cdot \nabla Q \neq 0$, this equation takes on the form

$$\hat{\mathbf{t}} \cdot \nabla \psi = -\frac{\nabla \cdot (\mathbb{K} \cdot \nabla Q)}{(\hat{\mathbf{n}} \cdot \nabla Q)}, \quad (48.224)$$

which provides a means to integrate the streamfunction, ψ , along contours of constant Q .

If Q is the quasi-geostrophic potential vorticity (Chapter 42), then contours of constant Q are known as *geostrophic contours*. Within this context, [Rhines and Holland \(1979\)](#) made use of the identity (48.224) in their study of ocean circulation in the presence of eddy diffusion of potential vorticity.

Hint: write the advection operator as a Jacobian and make use of Exercise 5.1.



Eddy and mean tracer kinematics

Geophysical fluid flows exhibit multiple scales in both space and time. In the analysis of these flows, it is useful to seek a description that decomposes fluid properties into a mean component and a fluctuation relative to the mean. We perform an eddy-mean decomposition when interest concerns the mean field and impacts on the mean by the fluctuating instantaneous flow, with such impacts often termed *rectified* effects. The mean field can be defined in many fashions with subjective choices based on particulars of the flow and the analysis goals. The definition for the mean in turn affects what we refer to as the fluctuation. Quite generally, fluctuations take the form of transient linear waves, nonlinear and/or breaking waves, coherent structures, and/or a chaotic/turbulent soup of eddying features. In this chapter we develop a kinematic framework originally motivated by the analysis of scalar transport induced by small amplitude wave-like eddying features, but is also of use for turbulent processes and their parameterizations (e.g., Chapter 50).

We consider two kinematic methods to decompose the flow into a mean and eddy. The first is the *generalized Lagrangian mean (GLM)*, which is a hybrid Eulerian/Lagrangian method that introduces an Eulerian disturbance field to measure the position of a fluid particle relative to its mean position ([Andrews and McIntyre, 1978a,b](#); [Bühler, 2014a](#)). For our purposes it is sufficient to use only a small piece of the GLM framework to help unpack the kinematics of eddy tracer fluxes. The second kinematic method makes use of isopycnal vertical coordinates. We connect an isopycnal description to the GLM by applying the GLM just in the vertical direction. In this sense the isopycnal approach is quasi-Lagrangian since it fixes the horizontal position (Eulerian) yet allows the vertical to follow an adiabatic fluid parcel (Lagrangian). The isopycnal approach is frequently used to help understand how ocean mesoscale eddies affect stratification and tracer transport in stably stratified flows. Our presentation follows the methods developed by [McDougall and McIntosh \(2001\)](#) and summarized in Chapter 9 of [Griffies \(2004\)](#).

READER'S GUIDE TO THIS CHAPTER

Material in this chapter relies on an understanding of the tracer equation as derived in Section 16.1 and the maths and physics of the advection-diffusion equation explored in Chapter 48. We focus most discussion on incompressible flows discussed in Chapter 17 and applicable to the Boussinesq fluid commonly assumed for the ocean (Chapter 24). Generalizations to compressible fluids are straightforward, with examples provided by [Griffies and Greatbatch \(2012\)](#). The kinematics of isopycnal fluid layers in a perfect fluid (Sections 49.4 and 49.5) are posed using the isopycnal vertical coordinates detailed in Chapter 43 and further pursued in Chapter 45.

49.2 Basic kinematics of the GLM	1117
49.2.1 Motivation	1117
49.2.2 Length scales and the small parameter	1119
49.2.3 Decomposing the particle trajectory	1119
49.2.4 GLM and the Stokes mean	1120
49.2.5 An example linear wave	1121
49.2.6 GLM with a materially constant scalar	1123
49.2.7 Further study	1124
49.3 Kinematics of eddy tracer fluxes	1124
49.3.1 Particle displacements and eddy tracer fluxes	1124
49.3.2 Symmetric and skew symmetric tracer fluxes	1124
49.3.3 Massaging the mean field tracer equation	1127
49.3.4 Connection to Stokes drift	1128
49.3.5 A linear rotating periodic wave example	1128
49.3.6 Further study	1130
49.4 Kinematics of volume transport in isopycnal layers	1130
49.4.1 Isopycnal mean	1131
49.4.2 Modified mean	1131
49.4.3 Transformed residual mean (TRM)	1132
49.4.4 Volume conservation and the thickness equation	1134
49.4.5 Ensemble mean kinematics in isopycnal coordinates	1135
49.4.6 Ensemble mean kinematics in geopotential coordinates	1136
49.4.7 Approximate ensemble mean kinematics in geopotential coordinates	1137
49.4.8 Further study	1138
49.5 Mean tracer equation	1139
49.5.1 Thickness weighted means	1139
49.5.2 Isopycnal mean thickness weighted tracer equation	1139
49.5.3 Subgrid scale tracer transport tensor	1140
49.5.4 Mean tracer transport beneath a density surface	1140
49.5.5 Summary of the tracer parameterization problem	1140
49.5.6 Comments	1141

49.1 Reynolds decomposition

At any point in space and time, we can decompose a field into a mean, $\bar{\Phi}(\mathbf{x}, t)$, and a departure from the mean, $\Phi'(\mathbf{x}, t)$

$$\Phi(\mathbf{x}, t) = \bar{\Phi}(\mathbf{x}, t) + \Phi'(\mathbf{x}, t). \quad (49.1)$$

The departure from the mean is generally termed the “eddy” or the “fluctuation”. The following offers a non-exhaustive list of mean operators.

- **TIME MEAN:** If the mean operator is based on a long time mean, then the mean fields are assumed to be time independent: $\bar{\Phi}(\mathbf{x}, t) = \bar{\Phi}(\mathbf{x})$. This is a common operator when interest is focused on the long term mean fluid properties.
- **PHASE AVERAGE:** Rather than a time mean, we may choose to average over the phase (or period) of a wave. This choice is particularly relevant when the fluctuating field involves quasi-linear waves.
- **ZONAL MEAN:** If the mean operator is based on an average along a particular coordinate direction (e.g., zonal average), then the mean tracer concentration is independent of the “averaged out” direction.

- COARSE GRAINING: If the mean operator is based on an average over a spatial and temporal region, such as the mesoscale, then such coarse-graining averages out smaller scales.
- ENSEMBLE MEAN: Rather than a space or time mean operation, we may consider the mean computed over an ensemble of many flow realizations. For many purposes this is the most analytically convenient operator.

If a mean operator satisfies the following properties then it is said to provide a “Reynolds decomposition”

$$\overline{\Phi'} = 0 \quad (49.2a)$$

$$\overline{\overline{\Phi}} = \overline{\Phi} \quad (49.2b)$$

$$\overline{\gamma \Phi} = \gamma \overline{\Phi} \quad \text{for } \gamma \text{ a constant.} \quad (49.2c)$$

Equation (49.2a) says that the mean of an eddy fluctuation vanishes. The equality (49.2b) says that the mean of a mean field returns the mean field. The final equality, (49.2c), says that a constant commutes with the mean operator. Notably, some or all of these properties are not satisfied by certain operators used for eddy-mean decompositions. However, in the following we assume they are satisfied.

49.2 Basic kinematics of the GLM

We here consider basic elements of generalized Lagrangian mean (GLM) theory. GLM is distinct from both the Eulerian mean and the Lagrangian mean. Rather, GLM is a hybrid between Lagrangian and Eulerian descriptions of fluid motions, so that it might be more appropriate to refer to it as the “hybrid Lagrangian-Eulerian mean theory”.

The GLM and the Eulerian mean for a fluid property are generally distinct, with their difference referred to as the *Stokes mean*

$$\text{Lagrangian mean} = \text{Eulerian mean} + \text{Stokes mean}. \quad (49.3)$$

This name is motivated from the *Stokes drift* introduced in Section 30.3, which we again encounter in Section 49.2.4. Note that the literature typically refers to the Stokes mean as the “Stokes correction”. We avoid that terminology in order to avoid the spurious notion that one type of mean operator is more correct than the other. We instead propose that a mean operator is subjectively chosen based on its suitability to a particular scientific question. Furthermore, no mean operator is suitable for all questions.

49.2.1 Motivation

Consider a materially constant scalar field

$$\frac{D\Phi}{Dt} = \frac{\partial \Phi}{\partial t} + \mathbf{v} \cdot \nabla \Phi = 0. \quad (49.4)$$

The scalar Φ is constant following fluid particles whose trajectories are integral curves of the fluid velocity \mathbf{v} . The question arises how to develop a mean operator that averages over fluctuations in the trajectories while preserving the material constancy nature of the instantaneous equation $D\Phi/Dt = 0$. This aspiration is not trivial.

Eulerian mean

An Eulerian mean operator considered in Section 49.1 leads to the mean field equation

$$\frac{\partial \bar{\Phi}}{\partial t} + \bar{\mathbf{v}} \cdot \nabla \bar{\Phi} = -\overline{\mathbf{v}' \cdot \nabla \Phi'}. \quad (49.5)$$

Whereas Φ is materially constant when following the instantaneous flow field \mathbf{v} , the Eulerian mean $\bar{\Phi}$ is not materially constant when following $\bar{\mathbf{v}}$ due to the source term $-\overline{\mathbf{v}' \cdot \nabla \Phi'}$ provided by the eddy correlation. Furthermore, when given information only about the mean fields, then we must develop a closure for the unresolved correlation. Such closures are the topic of extensive research typical of eddy-mean decompositions. Nonetheless, we ask whether there are methods that offer insights into mean field behaviour even without making a closure assumption. GLM is one such method.

Lagrangian mean

An alternative approach is to remain in the Lagrangian frame, where material constancy of Φ takes on the linear form

$$\frac{\partial \Phi(\mathbf{a}, t)}{\partial t} = 0. \quad (49.6)$$

Consider a mean operator computed as an average over a region of material space coordinate \mathbf{a} . For example, if \mathbf{a} is the initial fluid particle position, then an average coordinate, $\bar{\mathbf{a}}$, and corresponding averaged field, $\bar{\Phi}$, render a coarse-graining over the initial positions. Since each member of the Lagrangian average satisfies the linear equation (49.6), so too does the Lagrangian mean

$$\frac{\partial \bar{\Phi}(\bar{\mathbf{a}}, t)}{\partial t} = 0. \quad (49.7)$$

Although this equation retains the simplicity of the unaveraged version, it still requires information about trajectories. Trajectories are computed based on the flow map (i.e., the velocity field), with trajectories an impractical means for describing chaotic or turbulent fluids. GLM offers an alternative that aims to meld elements of the Eulerian (e.g., computability) to the Lagrangian (e.g., material constancy).

Generalized Lagrangian mean

The GLM approach produces a GLM field that remains constant following the GLM velocity

$$\frac{\partial \bar{\Phi}^{(L)}}{\partial t} + \bar{\mathbf{v}}^{(L)} \cdot \nabla \bar{\Phi}^{(L)} = 0. \quad (49.8)$$

Hence, GLM maintains the desirable properties of the Lagrangian mean. However, it does so using Eulerian methods which can prove to be more practical for many cases. Notably, even if the Eulerian velocity is non-divergent, as for a Boussinesq fluid, the GLM velocity is generally divergent. Although we will not prove the GLM result (49.8), we will motivate the GLM average from the analysis of small amplitude eddying motions.

49.2.2 Length scales and the small parameter

There are two length scales associated with an eddy or wave fluctuation. One characterizes the size of the eddy whose length scale we write as λ . If the eddy is a monochromatic wave, then λ is its wave length. The other length scale characterizes the size of particle displacements, $|\xi|$. In the following, we assume the particle displacements are small relative to λ

$$|\xi| \ll \lambda \quad \text{small amplitude waves.} \quad (49.9)$$

We thus introduce the small non-dimensional ratio of length scales for the following analysis

$$\alpha = \frac{|\xi|}{\lambda} \ll 1. \quad (49.10)$$

49.2.3 Decomposing the particle trajectory

Recall the discussion of fluid particle trajectories given in Chapter 13. In this description, the trajectory of a particle is determined by integrating the relation between the particle trajectory and the particle velocity

$$\left[\frac{\partial \mathbf{X}(\mathbf{a}, t)}{\partial t} \right]_{\mathbf{a}} = \mathbf{v}[\mathbf{X}(\mathbf{a}, t)] \implies \mathbf{X}(\mathbf{a}, t) = \mathbf{X}(\mathbf{a}, 0) + \int_0^t \mathbf{v}[\mathbf{X}(\mathbf{a}, t')] dt', \quad (49.11)$$

so that the trajectory measures the position of a particle relative to a chosen origin. The material coordinate, \mathbf{a} , distinguishes the continuum of fluid particles, thus making the trajectory a field in material space-time.

The GLM develops a hybrid Eulerian-Lagrangian method and it is motivated by linear or quasi-linear disturbances. Keeping this motivation in mind, we consider each point in space, \mathbf{x} , to be the mean position of a unique fluid particle. In turn, we introduce an Eulerian field, $\xi(\mathbf{x}, t)$, that measures the position of a fluid particle relative to its mean position. Correspondingly, the Eulerian mean of the disturbance field vanishes

$$\overline{\xi(\mathbf{x}, t)} = 0. \quad (49.12)$$

Note that the Eulerian mean operator can be any of the operators (or others) satisfying the Reynold's decomposition property discussed in Section 49.1

Specification of $\xi(\mathbf{x}, t)$ for large amplitude disturbances (i.e., nonlinear waves) requires the full machinery of GLM, which is beyond our scope. Instead, to expose the rudiments we assume linear waves such as shown in Figure 49.1, for which the particle displacement amplitude is much smaller than the wavelength of the disturbance

$$\alpha = \frac{|\xi|}{\lambda} \ll 1. \quad (49.13)$$

In this case the disturbance field is constructed by time integration of the eddy velocity field

$$\left[\frac{\partial \xi(\mathbf{x}, t)}{\partial t} \right]_{\mathbf{x}} = \mathbf{v}'(\mathbf{x}, t) \implies \xi(\mathbf{x}, t) = \int^t \mathbf{v}'(\mathbf{x}, t') dt'. \quad (49.14)$$

With this specification for the disturbance field, we see that if the eddy velocity is non-divergent then so is the disturbance particle position field

$$\nabla \cdot \mathbf{v}' = 0 \implies \nabla \cdot \xi = 0. \quad (49.15)$$

The definition (49.14) for the disturbance field, $\xi(\mathbf{x}, t)$, is directly analogous to the particle trajectory position, $\mathbf{X}(\mathbf{a}, t)$, given by equation (49.11). However, there are important distinctions. Namely, the disturbance, $\xi(\mathbf{x}, t)$, is an Eulerian field that measures the position of a fluid particle relative to its mean position, with each Eulerian position \mathbf{x} corresponding to the mean position for a distinct fluid particle. In contrast, the particle position, $\mathbf{X}(\mathbf{a}, t)$, is a Lagrangian field that is attached to each fluid particle and measures the position of that particle relative to a unique origin.

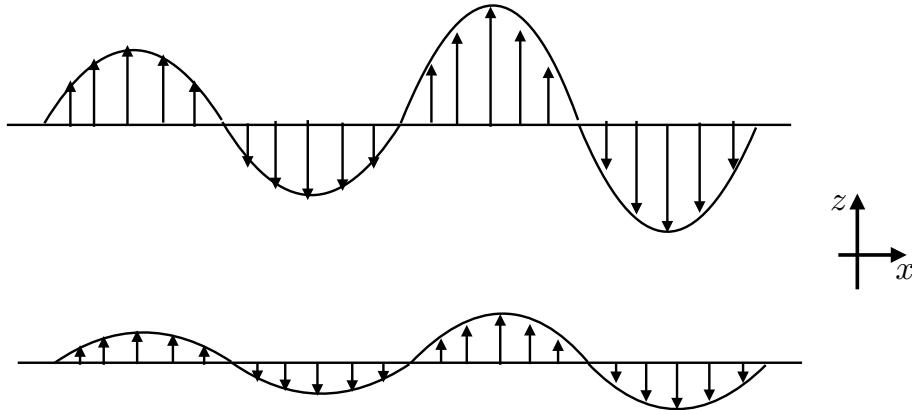


FIGURE 49.1: Illustrating the displacement of fluid particles at two selective vertical positions due to a linear transverse wave disturbance. The particle positions, $\mathbf{x}^{(\xi)} = \mathbf{x} + \xi(\mathbf{x}, t)$, have a disturbance field of the form $\xi(\mathbf{x}, t) = \hat{\mathbf{z}} \xi_0(x, z) \sin(kx - \omega t)$, with $\xi_0(x, z)$ a spatially dependent wave amplitude, $\lambda = 2\pi/k$ the wavelength, $\mathbf{k} = \hat{\mathbf{x}} k$ the wavevector, $\omega = ck$ the angular frequency, and c the wavespeed. Small amplitude waves satisfy $|\xi_0| \ll \lambda$. Note that this wave does not produce a Stokes drift since particle displacements are perpendicular to the wavevector: $\xi \cdot \mathbf{k} = 0$ (see Section 49.2.5), whereas Stokes drift requires particle motion to have a nonzero component in the wave direction (see Figures 30.4 and 30.5). Even so, it does generally produce a Stokes mean for an arbitrary field Φ (Section 49.2.4).

49.2.4 GLM and the Stokes mean

The mean of a fluid property, Φ , is generally a function of how the property is sampled to compute the mean. For example, the mean of Φ sampled on a fluctuating fluid particle differs from the mean sampled at the particle's mean position. Mathematically, this distinction means that

$$\underbrace{\overline{\Phi(\mathbf{x} + \xi(\mathbf{x}, t))}}_{\text{GLM}} \neq \underbrace{\overline{\Phi(\mathbf{x}, t)}}_{\text{Eulerian}}, \quad (49.16)$$

where it is common to make use of the shorthand¹

$$\mathbf{x}^{(\xi)}(\mathbf{x}, t) \equiv \mathbf{x} + \xi(\mathbf{x}, t) \quad (49.17)$$

for the instantaneous position of the fluid particle. The average

$$\overline{\Phi}^{(L)}(\mathbf{x}, t) \equiv \overline{\Phi(\mathbf{x} + \xi(\mathbf{x}, t), t)} = \overline{\Phi(\mathbf{x}^{(\xi)})} \quad (49.18)$$

defines the generalized Lagrangian mean. As defined, the GLM is computed by evaluating the property Φ at the position of a fluid particle, $\mathbf{x}^{(\xi)}(\mathbf{x}, t) = \mathbf{x} + \xi(\mathbf{x}, t)$, where \mathbf{x} is both an arbitrary

¹We place superscripts ξ, S, L, E inside parentheses to distinguish from tensor labels.

Eulerian field point and the mean position of a fluid particle, $\overline{\mathbf{x}^{(\xi)}} = \mathbf{x}$. In contrast, the Eulerian mean is determined by evaluating Φ at the fixed Eulerian point in space

$$\overline{\Phi}^{(E)}(\mathbf{x}, t) \equiv \overline{\Phi(\mathbf{x}, t)}. \quad (49.19)$$

Following our discussion at the start of Section 49.2, we define the difference between the GLM and Eulerian mean as the Stokes mean

$$\overline{\Phi}^{(S)}(\mathbf{x}, t) = \overline{\Phi}^{(L)}(\mathbf{x}, t) - \overline{\Phi}^{(E)}(\mathbf{x}, t). \quad (49.20)$$

The Stokes mean arises from inhomogeneities in Φ , which in turn lead to differences in its mean depending on how that field is sampled, whether sampled on a fluid particle, $\mathbf{x}^{(\xi)}(\mathbf{x}, t)$, or sampled at the mean position of the fluid particle, \mathbf{x} .

We mathematically expose the origin of the Stokes mean by performing a Taylor series expansion around the mean particle position

$$\Phi(\mathbf{x} + \boldsymbol{\xi}, t) = \Phi(\mathbf{x}, t) + \boldsymbol{\xi} \cdot \nabla \Phi(\mathbf{x}, t) + \frac{1}{2} \xi_m \xi_n \partial_m \partial_n \Phi(\mathbf{x}, t) + \mathcal{O}(\alpha^3). \quad (49.21)$$

The non-dimensional ratio $\alpha = |\boldsymbol{\xi}|/\lambda \ll 1$ was introduced in equation (49.13), which measures the ratio of the amplitude for particle displacements to the wavelength, λ , of fluctuations in the field Φ . Taking the mean of equation (49.21) then leads to an expression for the Stokes mean

$$\overline{\Phi}^{(S)}(\mathbf{x}, t) = \overline{\Phi}^{(L)}(\mathbf{x}, t) - \overline{\Phi}^{(E)}(\mathbf{x}, t) \quad (49.22a)$$

$$= \overline{\boldsymbol{\xi} \cdot \nabla \Phi} + \frac{1}{2} \overline{\xi_m \xi_n \partial_m \partial_n \Phi} + \mathcal{O}(\alpha^3). \quad (49.22b)$$

$$= \overline{\boldsymbol{\xi} \cdot \nabla \Phi'} + \frac{1}{2} \overline{\xi_m \xi_n} \partial_m \partial_n \overline{\Phi}^{(E)} + \mathcal{O}(\alpha^3), \quad (49.22c)$$

where we introduced the Eulerian fluctuation

$$\Phi'(\mathbf{x}, t) = \Phi(\mathbf{x}, t) - \overline{\Phi}^{(E)}(\mathbf{x}, t) \quad (49.23)$$

and all terms on the right hand side of equation (49.22c) are evaluated at (\mathbf{x}, t) . The Stokes drift (Section 30.3) associated with the GLM arises from setting Φ equal to one of the velocity components

$$\overline{v}_p^{(S)} = \overline{\boldsymbol{\xi} \cdot \nabla v'_p} + \frac{1}{2} \overline{\xi_m \xi_n} \partial_m \partial_n \overline{v}_p^{(E)} + \mathcal{O}(\alpha^3). \quad (49.24)$$

49.2.5 An example linear wave

We exemplify the previous discussion by considering the small amplitude linear wave

$$\boldsymbol{\xi} = -\frac{\mathbf{U}(\mathbf{x})}{\omega} \sin(\mathbf{k} \cdot \mathbf{x} - \omega t) \quad (49.25a)$$

$$\mathbf{v}' = \partial_t \boldsymbol{\xi} = \mathbf{U}(\mathbf{x}) \cos(\mathbf{k} \cdot \mathbf{x} - \omega t) \quad (49.25b)$$

$$\nabla v'_p = \nabla U_p \cos(\mathbf{k} \cdot \mathbf{x} - \omega t) - \mathbf{k} U_p \sin(\mathbf{k} \cdot \mathbf{x} - \omega t) \quad (49.25c)$$

$$\nabla \cdot \mathbf{v}' = (\nabla \cdot \mathbf{U}) \cos(\mathbf{k} \cdot \mathbf{x} - \omega t) - \mathbf{k} \cdot \mathbf{U} \sin(\mathbf{k} \cdot \mathbf{x} - \omega t) \quad (49.25d)$$

where \mathbf{U} is the velocity amplitude that is generally a function of space, \mathbf{k} is the wavevector, and

$$T = 2\pi/\omega \quad (49.26)$$

is the wave period. The wave renders an oscillatory motion to fluid particles, with the disturbance field specifying the instantaneous position of fluid particles whose mean position is \mathbf{x} . The disturbance field and velocity field both have a zero mean when time integrated over a wave period

$$\bar{\xi}(\mathbf{x}) = \frac{1}{T} \int_0^T \xi(\mathbf{x}, t') dt' = 0 \quad (49.27a)$$

$$\bar{\mathbf{v}}'(\mathbf{x}) = \frac{1}{T} \int_0^T \mathbf{v}'(\mathbf{x}, t') dt' = 0. \quad (49.27b)$$

To maintain a non-divergent eddy velocity at arbitrary times requires

$$\nabla \cdot \mathbf{v}' = 0 \implies \nabla \cdot \mathbf{U} = \mathbf{U} \cdot \mathbf{k} = 0. \quad (49.28)$$

The second condition means that the wave is transverse, so that particle displacements are orthogonal to the wavevector (e.g., Figure 49.1)

Stokes drift

Specializing to the velocity field (49.25b), substituting into the Stokes drift expression (49.24), and making use of an average over a wave period yields

$$\overline{\xi \cdot \nabla v'_p} = \frac{U_p \mathbf{U} \cdot \mathbf{k}}{2\omega} \quad (49.29a)$$

$$\bar{v}_p^{(E)} = 0. \quad (49.29b)$$

The second equality holds since the velocity at a point arises just from the wave field, which has a zero Eulerian mean. Hence, to $\mathcal{O}(\alpha^2)$, the Stokes drift velocity associated with the GLM is given by

$$\bar{\mathbf{v}}^{(S)} = \frac{\mathbf{U} (\mathbf{U} \cdot \mathbf{k})}{2\omega} + \mathcal{O}(\alpha^2). \quad (49.30)$$

The Stokes drift vanishes at this order of accuracy for transverse waves in which $\mathbf{U} \cdot \mathbf{k} = 0$.

As a check on our formalism we consider a one-dimensional longitudinal wave, in which the Stokes drift is given by

$$\bar{v}^{(S)} = \frac{U^2}{2c} + \mathcal{O}(\alpha^2). \quad (49.31)$$

This result agrees with that derived using Lagrangian trajectories in Section 30.3 (see Exercise 30.1). Use of the GLM displacement field offers a somewhat more streamlined method for computing Stokes drift.

Stokes mean for an arbitrary field

The Stokes mean for an arbitrary field is given by

$$\bar{\Phi}^{(S)}(\mathbf{x}, t) = -\omega^{-1} \mathbf{U} \cdot \overline{\nabla \Phi' \sin(\mathbf{k} \cdot \mathbf{x} - \omega t)} + \mathcal{O}(\alpha^2) \quad (49.32a)$$

$$= -\omega^{-1} \overline{\nabla \cdot (\mathbf{U} \Phi')} \sin(\mathbf{k} \cdot \mathbf{x} - \omega t) + \mathcal{O}(\alpha^2), \quad (49.32b)$$

where the second equality made use of the non-divergent nature of the wave field (49.28). To second order in wave amplitude, the Stokes mean is determined by the projection of the gradient of the Eulerian fluctuation, $\nabla \Phi'$, onto the wave amplitude, \mathbf{U} . For example, consider a transverse wave such as that shown in Figure 49.1. Even though the Stokes drift vanishes to order $\mathcal{O}(\alpha^2)$, the Stokes mean, $\bar{\Phi}^{(S)}(\mathbf{x}, t)$, can be nonzero so long as there is a nonzero vertical gradient in the Eulerian fluctuation.

49.2.6 GLM with a materially constant scalar

Consider a materially constant scalar field, such as a tracer concentration in the absence of mixing and sources

$$\frac{DC}{Dt} = 0. \quad (49.33)$$

How the GLM for C is related to the instantaneous C

The GLM for C equals to the value of C on a fluid particle

$$\bar{C}^{(L)}(\mathbf{x}, t) = C(\mathbf{x} + \boldsymbol{\xi}, t). \quad (49.34)$$

This is a very important identity that packs in a lot of information. In words, it says that when evaluated at the mean fluid particle position, \mathbf{x} , the GLM field $\bar{C}^{(L)}(\mathbf{x}, t)$ equals to the concentration, C , evaluated on a fluid particle at $\mathbf{x} + \boldsymbol{\xi}$. One means to understand this identity is to assume the GLM is an ensemble mean following fluid particles. Since C is constant on fluid particles, each ensemble member has the same value for C , in which case the GLM for C clearly equals the value of C for each ensemble member. We make particular use of the identity (49.34) when considering isopycnal kinematics in Sections 49.4 and 49.5.

Relating the particle disturbance field to Eulerian properties of C

There is a frequently used consequence of the identity (49.34) involving the disturbance field, the Eulerian fluctuation

$$C'(\mathbf{x}, t) = C(\mathbf{x}, t) - \bar{C}(\mathbf{x}, t) \quad (49.35)$$

and the Eulerian mean

$$C^{(E)}(\mathbf{x}, t) = \bar{C}(\mathbf{x}, t). \quad (49.36)$$

To derive it, recall the Taylor series expansion (49.21) truncated here to first order accuracy

$$C(\mathbf{x} + \boldsymbol{\xi}, t) = C(\mathbf{x}, t) + \boldsymbol{\xi} \cdot \nabla \bar{C}(\mathbf{x}, t) + \mathcal{O}(\alpha^2). \quad (49.37)$$

Taking the Eulerian mean of both sides renders

$$\overline{C(\mathbf{x} + \boldsymbol{\xi}, t)} = \bar{C}(\mathbf{x}, t) + \mathcal{O}(\alpha^2). \quad (49.38)$$

This identity means that the GLM equals to the Eulerian mean to order $\mathcal{O}(\alpha^2)$, which is a result consistent with the Stokes mean being an order $\mathcal{O}(\alpha^2)$ quantity as seen by equation (49.22c). From equation (49.34) we know that $C(\mathbf{x} + \boldsymbol{\xi}, t) = \bar{C}(\mathbf{x} + \boldsymbol{\xi}, t)$, so that we can subtract equations (49.37) and (49.38) to find

$$C'(\mathbf{x}, t) = -\boldsymbol{\xi} \cdot \nabla \bar{C}(\mathbf{x}, t) + \mathcal{O}(\alpha^2). \quad (49.39)$$

Hence, to first order accuracy, the Eulerian fluctuation equals to minus the disturbance field projected onto the gradient of the mean field; i.e., the Eulerian fluctuation in the tracer is first order in the disturbance. We make use of this result when discussing the kinematics of eddy tracer fluxes in Section 49.3. Furthermore, for the isopycnal kinematics in Sections 49.4 and 49.5, we focus on vertical particle displacements, $\boldsymbol{\xi} = \xi \hat{z}$, in which case the Eulerian fluctuation is given by

$$C'(z, t) = -\xi \partial_z \bar{C}(z, t) + \mathcal{O}(\alpha^2). \quad (49.40)$$

49.2.7 Further study

GLM was introduced in the seminal papers by [Andrews and McIntyre \(1978a,b\)](#). These papers offer a wealth of intellectual rewards after much study. GLM is also detailed in the monograph on waves and mean flows by [Bühler \(2014a\)](#).

49.3 Kinematics of eddy tracer fluxes

Consider the Eulerian eddy-mean decomposition for a materially constant tracer in an incompressible fluid. The advection equation for this tracer is given by

$$\frac{\partial C}{\partial t} + \nabla \cdot (\mathbf{v} C) = 0, \quad (49.41)$$

and its Eulerian mean is

$$\frac{\partial \bar{C}}{\partial t} + \nabla \cdot (\bar{\mathbf{v}} \bar{C}) = -\nabla \cdot \bar{\mathbf{v}}' C'. \quad (49.42)$$

The eddy advective flux, $\mathbf{v}' C'$, is the product of the eddy velocity and eddy tracer concentration. Its mean provides the correlation or mean eddy flux, $\bar{\mathbf{v}}' \bar{C}'$. The convergence of this mean eddy flux provides a source to the advection equation for the Eulerian mean tracer concentration.

In this section we make use of the particle disturbance field of Section 49.2 to unpack the kinematics of eddy tracer fluxes induced by small amplitude waves. Although not offering new dynamical information, the particle disturbance field is a very useful means to frame the kinematics of tracer eddy fluxes.

49.3.1 Particle displacements and eddy tracer fluxes

Following Section 49.2, we here introduce a particle disturbance vector corresponding to small amplitude eddy fluctuations

$$\partial_t \boldsymbol{\xi}(\mathbf{x}, t) = \mathbf{v}'(\mathbf{x}, t) + \mathcal{O}(\alpha^2) \quad (49.43a)$$

$$\bar{\boldsymbol{\xi}} = 0. \quad (49.43b)$$

Correspondingly, each spatial point, \mathbf{x} , is the mean position of a fluid particle whose instantaneous position is $\mathbf{x} + \boldsymbol{\xi}(\mathbf{x}, t)$. Following the results from Section 49.2.6, to leading order we can write the Eulerian fluctuation in terms of the particle displacement (equation (49.39))

$$C'(\mathbf{x}, t) = -\boldsymbol{\xi} \cdot \nabla \bar{C}(\mathbf{x}, t) + \mathcal{O}(\alpha^2) \quad (49.44)$$

Notice that if the particle displacement is oriented along a mean tracer iso-surface, then $\boldsymbol{\xi} \cdot \nabla \bar{C}(\mathbf{x}, t) = 0$ and there is no tracer fluctuation, $C' = 0$, to order $\mathcal{O}(\alpha^2)$. More general eddy motions lead to a nonzero tracer fluctuation with the eddy tracer flux taking on the form

$$\mathbf{v}' C' = -\partial_t \boldsymbol{\xi} \cdot (\boldsymbol{\xi} \cdot \nabla) \bar{C} + \mathcal{O}(\alpha^2). \quad (49.45)$$

We unpack this expression for the purpose of characterizing kinematic properties of the eddy tracer flux

49.3.2 Symmetric and skew symmetric tracer fluxes

From equation (49.45), the m 'th component of the eddy tracer flux is given by

$$v'_m C' = -[(\partial_t \xi_m) \xi_n] \partial_n \bar{C}. \quad (49.46)$$

We here decompose this flux in order to characterize its kinematic properties.

Decomposing the tracer flux

Let us decompose the second order tensor $(\partial_t \xi_m) \xi_n$ into its symmetric and anti-symmetric components²

$$2(\partial_t \xi_m) \xi_n = [(\partial_t \xi_m) \xi_n + (\partial_t \xi_n) \xi_m] + [(\partial_t \xi_m) \xi_n - (\partial_t \xi_n) \xi_m] \quad (49.47a)$$

$$= \partial_t(\xi_m \xi_n) + [(\partial_t \xi_m) \xi_n - (\partial_t \xi_n) \xi_m]. \quad (49.47b)$$

Introducing the symmetric and anti-symmetric correlation tensors

$$2\mathbb{K}_{mn} \equiv \overline{\partial_t(\xi_m \xi_n)} \quad (49.48a)$$

$$2\mathbb{A}_{mn} \equiv \overline{(\partial_t \xi_m) \xi_n} - \overline{(\partial_t \xi_n) \xi_m} \quad (49.48b)$$

allows us to write the mean eddy tracer flux

$$\overline{v'_m C'} = -(\mathbb{K}_{mn} + \mathbb{A}_{mn}) \partial_n \overline{C} \quad (49.49)$$

and the mean field tracer equation (49.42)

$$\frac{\partial \overline{C}}{\partial t} + \nabla \cdot (\bar{\mathbf{v}} \overline{C}) = \nabla \cdot [(\mathbb{K} + \mathbb{A}) \cdot \nabla \overline{C}]. \quad (49.50)$$

The right hand side of this equation equals to the convergence of the symmetric and skew-symmetric tracer fluxes³

$$\nabla \cdot [(\mathbb{K} + \mathbb{A}) \cdot \nabla \overline{C}] = -\nabla \cdot (\mathbf{F}^{(\text{sym})} + \mathbf{F}^{(\text{skew})}), \quad (49.51)$$

where

$$\mathbf{F}^{(\text{sym})} = -\mathbb{K} \cdot \nabla \overline{C} \quad (49.52a)$$

$$\mathbf{F}^{(\text{skew})} = -\mathbb{A} \cdot \nabla \overline{C} \quad (49.52b)$$

$$\overline{\mathbf{v}' C'} = \mathbf{F}^{(\text{sym})} + \mathbf{F}^{(\text{skew})}. \quad (49.52c)$$

The symmetric flux

In terms of particle displacements, the symmetric flux (49.52a) is given by

$$F_m^{(\text{sym})} = -\mathbb{K}_{mn} \partial_n \overline{C} = -\frac{1}{2} \overline{\partial_t(\xi_m \xi_n)} \partial_n \overline{C}. \quad (49.53)$$

The symmetric tensor \mathbb{K} vanishes when the average is over the period of a periodic wave, in which the particle displacements undergo reversible periodic excursions (see Section 49.3.5). For waves that decay in amplitude over the averaging period, particle displacements decrease in magnitude thus leading to an upgradient symmetric flux. In contrast, particle displacements increase in magnitude for waves that grow over the averaging period, in which case the flux is downgradient just as for diffusion. Furthermore, growing nonlinear waves generally break and then develop into turbulence, with turbulence leading to further particle separation and dispersive tracer mixing. Dispersive mixing is well parameterized by diffusion, and we have more to say about diffusive parameterizations of lateral dispersion in Section 50.3.

²See Section 14.2.5 for a similar decomposition of the velocity gradient tensor.

³We place parentheses around “skew” and “sym” to distinguish the name for these vectors from what may otherwise appear to be tensor labels.

The skew, advective, and rotational fluxes

Following our discussion in Section 48.7, we write the skew flux as

$$F_m^{(\text{skew})} = -\mathbb{A}_{mn} \partial_n \bar{C} = -\epsilon_{mnp} \Psi_p \partial_n \bar{C} = -(\nabla \bar{C} \wedge \Psi)_m, \quad (49.54)$$

where we introduced the vector streamfunction (dimensions squared length per time)⁴

$$\Psi = \frac{1}{2} \partial_t \xi \wedge \xi = \frac{1}{2} \bar{v}' \wedge \xi. \quad (49.55)$$

The vector streamfunction is half the angular momentum per mass of a fluid particle undergoing eddying motion, with the angular momentum computed relative to the mean particle position. The vector streamfunction is nonzero only if the eddy has a preferred sense of rotation, in which case the wave field is said to be *polarized*.

The skew flux can be written

$$\mathbf{F}^{(\text{skew})} = -\nabla \bar{C} \wedge \Psi \quad (49.56a)$$

$$= (\nabla \wedge \Psi) \bar{C} - \nabla \wedge (\bar{C} \Psi) \quad (49.56b)$$

$$= \mathcal{U}^{(\Psi)} \bar{C} - \nabla \wedge (\bar{C} \Psi) \quad (49.56c)$$

$$= \mathbf{F}^{(\text{adv})} - \mathbf{F}^{(\text{rot})}, \quad (49.56d)$$

so that the skew flux equals to an advective flux minus a rotational flux. We here introduced the non-divergent velocity

$$\mathcal{U}^{(\Psi)} = \nabla \wedge \Psi \quad (49.57)$$

and the non-divergent rotational flux

$$\mathbf{F}^{(\text{rot})} = \nabla \wedge (\bar{C} \Psi). \quad (49.58)$$

The divergence of the skew flux equals to the divergence of the advective flux

$$\nabla \cdot \mathbf{F}^{(\text{skew})} = \nabla \cdot (\mathbf{F}^{(\text{adv})} - \mathbf{F}^{(\text{rot})}) = \nabla \cdot \mathbf{F}^{(\text{adv})}, \quad (49.59)$$

so that the rotational flux has no impact on evolution of the mean tracer field.

What does a point measurement estimate?

From equation (49.52c), we see that a point measurement of the correlation $\bar{v}' \bar{C}'$ provides an estimate of the diffusive and skew diffusive tracer fluxes

$$\bar{v}' \bar{C}' = \mathbf{F}^{(\text{sym})} + \mathbf{F}^{(\text{skew})} = -(\mathbb{K} + \mathbb{A}) \cdot \nabla \bar{C}. \quad (49.60)$$

Furthermore, for a periodic wave field, where the symmetric tensor vanishes, the correlation, $\bar{v}' \bar{C}'$, provides a direct estimate of the skew flux, $-\nabla \bar{C} \wedge \Psi$. One might instead presume that the point measurement offers a direct estimate of the advective flux, $\bar{C} \mathcal{U}^{(\Psi)}$, rather than the skew flux. But that presumption is wrong. Instead, since the skew flux equals to a rotational flux plus the advective flux, we have

$$\bar{v}' \bar{C}' = -\mathbb{K} \cdot \nabla \bar{C} - \nabla \bar{C} \wedge \Psi \quad (49.61a)$$

$$= -\mathbb{K} \cdot \nabla \bar{C} - \nabla \wedge (\bar{C} \Psi) + \bar{C} \nabla \wedge \Psi \quad (49.61b)$$

$$= -\mathbb{K} \cdot \nabla \bar{C} - \nabla \wedge (\bar{C} \Psi) + \bar{C} \mathcal{U}^{(\Psi)}. \quad (49.61c)$$

⁴ Middleton and Loder (1989) and Garrett (2006) introduce a skew-diffusivity, \mathbf{D} , which is opposite in sign to the vector streamfunction: $\Psi = -\mathbf{D}$.

The rotational flux is generally nontrivial for polarized waves and so cannot be ignored. As detailed by *Fox-Kemper et al.* (2003), there is no general method for removing the rotational flux. We therefore find it more convenient to work directly with the skew flux than the advective flux.

Area integrated tracer flux

We now offer an interpretation for the rotational contribution by considering the mean of the tracer flux integrated over a static area \mathcal{S}

$$\mathcal{T} = \overline{\int_{\mathcal{S}} \mathbf{v} C \cdot \hat{\mathbf{n}} dS} = \int_{\mathcal{S}} \overline{\mathbf{v} C} \cdot \hat{\mathbf{n}} dS = \int_{\mathcal{S}} [\overline{\mathbf{v}} \overline{C} + \overline{\mathbf{v}' C'}] \cdot \hat{\mathbf{n}} dS. \quad (49.62)$$

In terms of particle displacements, the eddy correlation, $\overline{\mathbf{v}' C'}$, equals to the sum of the symmetric flux and the skew flux as in equation (49.60). Introducing the diffusive, advective, and rotational flux then renders

$$\mathcal{T} = \int_{\mathcal{S}} [\overline{\mathbf{v}} \overline{C} + \mathcal{U}^{(\Psi)} \overline{C} - \nabla \wedge (\overline{C} \Psi) - \mathbb{K} \cdot \nabla \overline{C}] \cdot \hat{\mathbf{n}} dS. \quad (49.63)$$

Use of Stokes' Theorem transforms the rotational term to a line integral around the boundary of the area

$$\mathcal{T} = \int_{\mathcal{S}} [\overline{\mathbf{v}} \overline{C} + \mathcal{U}^{(\Psi)} \overline{C} - \mathbb{K} \cdot \nabla \overline{C}] \cdot \hat{\mathbf{n}} dS - \oint_{\partial\mathcal{S}} \overline{C} \Psi \cdot d\mathbf{l}. \quad (49.64)$$

Following Section 2b of *Middleton and Loder* (1989), we interpret the boundary term as a Stokes contribution associated with the correlation of particle motion and perturbation velocity along the boundary

$$\oint_{\partial\mathcal{S}} \overline{C} \Psi \cdot d\mathbf{l} = (1/2) \oint_{\partial\mathcal{S}} \overline{C} (\overline{\mathbf{v}' \wedge \xi}) \cdot d\mathbf{l}. \quad (49.65)$$

We further this interpretation when considering the transport beneath a fluctuating isopycnal surface in Section 49.5.4.

49.3.3 Massaging the mean field tracer equation

We here write the mean tracer equation (49.50) in various forms that can be found throughout the literature. For this purpose, write the right hand side of equation (49.50) in the form

$$\nabla \cdot [(\mathbb{K} + \mathbb{A}) \cdot \nabla \overline{C}] = \partial_m [(\mathbb{K}_{mn} + \mathbb{A}_{mn}) \partial_n \overline{C}] \quad (49.66a)$$

$$= \partial_m (\mathbb{K}_{mn} + \mathbb{A}_{mn}) \partial_n \overline{C} + (\mathbb{K}_{mn} + \mathbb{A}_{mn}) \partial_m \partial_n \overline{C} \quad (49.66b)$$

$$= \partial_m (\mathbb{K}_{mn} + \mathbb{A}_{mn}) \partial_n \overline{C} + \mathbb{K}_{mn} \partial_m \partial_n \overline{C}. \quad (49.66c)$$

The final equality follows from the identity

$$\mathbb{A}_{mn} \partial_m \partial_n \overline{C} = 0, \quad (49.67)$$

which results from the contraction of the anti-symmetric, \mathbb{A}_{mn} , to the symmetric operator $\partial_m \partial_n$. The second term, $\mathbb{K}_{mn} \partial_m \partial_n \overline{C}$, is a diffusion operator if symmetric tensor \mathbb{K} is also positive-definite. The first term in equation (49.66c) can be interpreted as an advection operator through the action of a non-divergent plus a divergent advection velocity

$$\partial_m (\mathbb{K}_{mn} + \mathbb{A}_{mn}) \partial_n \overline{C} = [\mathcal{U}^{(K)} + \mathcal{U}^{(\Psi)}] \cdot \nabla \overline{C}, \quad (49.68)$$

where⁵

$$\mathcal{U}_n^{(K)} \equiv -\partial_m \mathbb{K}_{mn} \implies \nabla \cdot \mathcal{U}^{(K)} = -\partial_n \partial_m \mathbb{K}_{mn} \neq 0 \quad (49.69a)$$

$$\mathcal{U}_n^{(\Psi)} \equiv -\partial_m \mathbb{A}_{mn} \implies \nabla \cdot \mathcal{U}^{(\Psi)} = \partial_n \partial_m \mathbb{A}_{mn} = 0. \quad (49.69b)$$

Bringing the above results together allows us to write the mean field tracer equation (49.50) in the following equivalent forms

$$\frac{\partial \bar{C}}{\partial t} + [\bar{\mathbf{v}} + \mathcal{U}^{(\Psi)} + \mathcal{U}^{(K)}] \cdot \nabla \bar{C} = \mathbb{K}_{mn} \partial_m \partial_n \bar{C} \quad \text{advective form} \quad (49.70a)$$

$$\frac{\partial \bar{C}}{\partial t} + \nabla \cdot ([\bar{\mathbf{v}} + \mathcal{U}^{(\Psi)}] \bar{C}) = \nabla \cdot (\mathbb{K} \cdot \nabla \bar{C}) \quad \text{flux form}, \quad (49.70b)$$

where we made use of the identities

$$\nabla \cdot \bar{\mathbf{v}} = 0 \quad \nabla \cdot \mathcal{U}^{(\Psi)} = 0 \quad \nabla \cdot \mathcal{U}^{(K)} \neq 0. \quad (49.71)$$

49.3.4 Connection to Stokes drift

From equation (49.24) we have the leading order expression for the Stokes drift

$$\bar{v}_p^{(S)} = \overline{\xi_n \partial_n \partial_t \xi_p} + \mathcal{O}(\alpha^2). \quad (49.72)$$

As noted in equation (49.15), with $\partial_t \boldsymbol{\xi} = \mathbf{v}'$ and with $\nabla \cdot \mathbf{v}' = 0$, the corresponding particle displacements are non-divergent, $\nabla \cdot \mathbf{v}' = 0 \Rightarrow \nabla \cdot \boldsymbol{\xi} = 0$. Consequently, to second order accuracy, the Stokes drift velocity can be written

$$\bar{v}_p^{(S)} = \overline{\xi_n \partial_n \partial_t \xi_p} \quad (49.73a)$$

$$= \partial_n \overline{[(\partial_t \xi_p) \xi_n]} \quad (49.73b)$$

$$= \partial_n (\mathbb{K}_{pn} + \mathbb{A}_{pn}) \quad (49.73c)$$

$$= \partial_n (\mathbb{K}_{np} - \mathbb{A}_{np}) \quad (49.73d)$$

$$= -\mathcal{U}_p^{(K)} + \mathcal{U}_p^{(\Psi)}. \quad (49.73e)$$

For the case of periodic waves, the Stokes drift velocity equals to the non-divergent skew velocity

$$\bar{\mathbf{v}}^{(S)} = \mathcal{U}^{(\Psi)} \Rightarrow \nabla \cdot \bar{\mathbf{v}}^{(S)} = 0 \quad \text{periodic waves}. \quad (49.74)$$

More generally, for non-periodic waves the divergent velocity is non-zero so that the Stokes velocity is also divergent

$$\bar{\mathbf{v}}^{(S)} = \mathcal{U}^{(\Psi)} - \mathcal{U}^{(K)} \Rightarrow \nabla \cdot \bar{\mathbf{v}}^{(S)} = -\nabla \cdot \mathcal{U}^{(K)} \neq 0 \quad \text{non-periodic waves}. \quad (49.75)$$

49.3.5 A linear rotating periodic wave example

We illustrate some of the previous analysis by considering Consider a displacement vector comprised of a linear periodic wave in two-dimensions

$$\boldsymbol{\xi}(\mathbf{x}, t) = \Gamma [\hat{\mathbf{x}} \cos(kx - \omega t) + \hat{\mathbf{y}} \sin(kx - \omega t)] \quad (49.76a)$$

$$\partial_t \boldsymbol{\xi}(\mathbf{x}, t) = \omega \Gamma [\hat{\mathbf{x}} \sin(kx - \omega t) - \hat{\mathbf{y}} \cos(kx - \omega t)], \quad (49.76b)$$

⁵Note that [Middleton and Loder \(1989\)](#) define $\mathcal{U}_n^{(K)} \equiv +\partial_m \mathbb{K}_{mn}$, which is the opposite sign to that used here in equation (49.69a), whereas they define $\mathcal{U}_n^{(\Psi)} = -\partial_m \mathbb{A}_{mn}$ as in equation (49.69b).

where Γ the time-independent wave amplitude, $T = 2\pi/\omega$ is the wave period, and $\lambda = 2\pi/k$ is the wavelength. The fluid particles exhibit counter-clockwise circular motion in the horizontal plane with squared radius

$$\xi \cdot \xi = \Gamma^2. \quad (49.77)$$

We are motivated to let the mean operator be a phase average

$$\bar{\phi} = \frac{1}{T} \int_0^T \phi(t) dt, \quad (49.78)$$

which is the traditional operator used when examining the impacts of linear waves on mean fields. For spatially constant wave amplitude, we will see the the mean tracer concentration, \bar{C} , remains unchanged by these waves. The absence of a rectified change to \bar{C} reflects the linear periodic nature of the wave field.

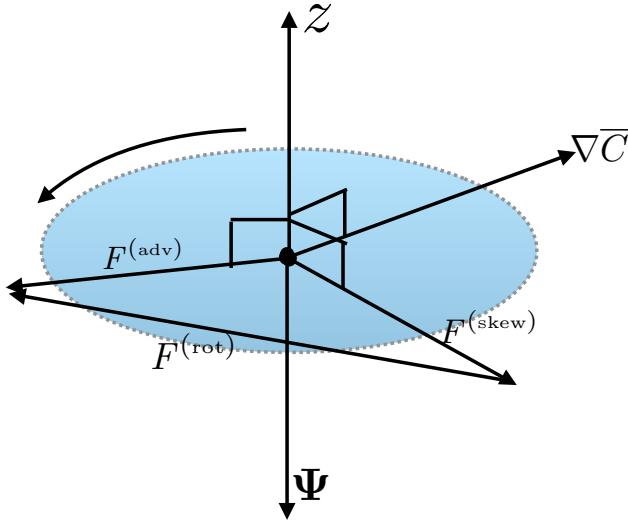


FIGURE 49.2: Sketch of the various tracer fluxes associated with the polarized displacement vector (49.76a). The particles are moving on the horizontal plane in a circle with time-independent radius Γ . The vector streamfunction (49.81) points in the negative \hat{z} direction. The mean concentration gradient, $\nabla \bar{C}$, generally points outside of the horizontal plane. However, it is only the horizontal components that contribute since the displacement vector is in the horizontal plane, thus resulting in horizontal skew, advective, and rotational fluxes.

Symmetric mixing tensor

The symmetric mixing tensor

$$\mathbb{K}_{mn} = \frac{\Gamma^2}{2T} \int_0^T dt \frac{\partial}{\partial t} \begin{bmatrix} \cos^2(kx - \omega t) & \cos(kx - \omega t) \sin(kx - \omega t) \\ \cos(kx - \omega t) \sin(kx - \omega t) & \sin^2(kx - \omega t) \end{bmatrix} \quad (49.79)$$

vanishes identically since the wave field is periodic so that the particle motion has an amplitude whose growing phase is exactly matched by its decaying phase.

Skew symmetric stirring tensor

In contrast, the skew-symmetric tensor has non-zero components

$$\mathbb{A}_{12} = -\mathbb{A}_{21} = \frac{\Gamma^2 \omega}{2} [\sin^2(kx - \omega t) + \cos^2(kx - \omega t)] = \frac{\Gamma^2 \omega}{2}, \quad (49.80)$$

which reflects the counter-clockwise polarization. The corresponding vector streamfunction is vertical

$$\Psi = \frac{\Gamma^2 \omega}{2} \hat{z}, \quad (49.81)$$

and the skew flux is horizontal

$$\mathbf{F}^{(\text{skew})} = \frac{\Gamma^2 \omega}{2} (\hat{z} \wedge \nabla \bar{C}). \quad (49.82)$$

Finally, the advective velocity is given by

$$\nabla \wedge \Psi = \omega \Gamma \nabla \Gamma \wedge \hat{z}, \quad (49.83)$$

with the advective velocity vanishing when the wave amplitude, Γ , is a constant. In this case, the advective tracer flux is zero, although the skew flux is non-zero. With a constant Γ , the skew flux has a zero divergence (it is a purely rotational flux when Γ is constant). Hence, for a constant wave amplitude, neither the skew flux nor the advective flux affect the evolution of \bar{C} . Figure 49.2 offers a schematic of the skew, advective, and rotational fluxes induced by a linear rotating particle wave in the horizontal plane.

49.3.6 Further study

Much of this section follows [Plumb \(1979\)](#), [Middleton and Loder \(1989\)](#), and [Garrett \(2006\)](#), each of whom considered elements of tracer dispersion by waves and nonlinear eddies. Additional treatments can be found in the review article of [Moffatt \(1983\)](#), who considers rotating fluids and magnetic fluids.

49.4 Kinematics of volume transport in isopycnal layers

In this section we consider the reversible stirring of fluid parcels by turbulent flow in a perfect (i.e., no mixing or sources) stratified Boussinesq fluid. As the fluid parcels are stirred, they preserve their volume while changing their shape and stretching into finer scale filaments. Stirring by ocean mesoscale/baroclinic eddies offers the canonical example of such stirring. Eventually, small-scale processes, such as those summarized in Section 51.1.4 mix properties irreversibly. We are here focused just on the stirring part of this scenario.

Over space and time scales larger than the mesoscale, the stirring by ocean mesoscale eddies can be considered chaotic, which in turn motivates a stochastic perspective in which an ensemble of eddies is considered. The goal is to describe the ensemble mean properties of the perfect fluid, with a focus in this section on the kinematics of parcel rearrangement. Hence, eddy correlations in the present section appear between the thickness of a fluid layer and the velocity. We introduce trace matter in Section 49.5, at which point we also consider eddy correlations between velocity and tracer as in Section 49.3.

The material in this section is rather detailed. However, its mastery comes readily by keeping in mind the more general (and somewhat simpler) presentation of GLM offered in Section 49.2. We are motivated to provide full details in this section since the kinematics of isopycnal ensembles appears throughout the study of wave-mean flow interactions in adiabatic geophysical fluid mechanics, such as in the study of ocean mesoscale eddies.

49.4.1 Isopycnal mean

Each fluid parcel in a stably stratified perfect Boussinesq fluid preserves its potential density. We are interested in following the vertical motion of potential density layer interfaces as waves and turbulent processes transport layer thickness from one region to another. In contrast, we are not concerned with following the lateral position of a fluid parcel within a layer. Here we introduce the isopycnal mean, which is based on describing ensembles of perfect fluid parcels using isopycnal coordinates. In Section 49.4.3, we relate this isopycnal approach to the vertical/isopycnal GLM.

Defining the isopycnal ensemble

An overbar with a potential density label, $\overline{(\cdot)}^{(\sigma)}$, denotes a mean over an ensemble of fluid parcels, each having the same potential density, σ , the same horizontal position, (x, y) , and the same time, t . Isopycnals undulate in space and time, which means that each ensemble member has a vertical position that is generally distinct from the ensemble mean depth, z . Furthermore, when the context is clear, it is useful to drop the dependence on (x, y, t) to highlight the dependence on potential density and/or the vertical position.

Isopycnal ensemble mean

The isopycnal ensemble mean makes use of potential density as a vertical coordinate (Chapters 43 and 45), with the mean field denoted by

$$\overline{\Phi}^{(\sigma)}(x, y, \sigma, t) \equiv \text{ensemble mean using isopycnal vertical coordinates.} \quad (49.84)$$

This average is straightforward to compute when using isopycnal coordinates, thus producing an isopycnal mean that is a function of the potential density, σ .

49.4.2 Modified mean

As a complement to the isopycnal approach in Section 49.4.1, we here introduce the vertical/isopycnal GLM, also known as the modified mean.

Vertical/isopycnal GLM

The discussion in Section 49.2 considered a three-vector particle displacement vector $\xi(\mathbf{x}, t)$. In contrast, we are here interested just in the vertical displacement of an isopycnal layer interface

$$\xi(x, y, \sigma, t) = \hat{z}\xi(x, y, \sigma, t). \quad (49.85)$$

The displacement field $\xi(x, y, \sigma, t)$ measures the vertical position of a potential density interface, σ , relative to its ensemble mean depth. For any particular ensemble member with potential density σ , we write its vertical position as (dropping x, y, t dependence for brevity)

$$z(\sigma) = \overline{z}^{(\sigma)} + \xi(\sigma), \quad (49.86)$$

where

$$\overline{z}^{(\sigma)} = \overline{z(\sigma)}^{(\sigma)} \quad (49.87)$$

is the isopycnal ensemble mean depth, and the displacement field has a zero ensemble mean

$$\overline{\xi(\sigma)}^{(\sigma)} = 0. \quad (49.88)$$

FIGURE 49.3: Schematic of the ensemble mean depth $\bar{z}^\rho(x, y, \rho, t)$ of a particular potential density surface ρ . In general, different members of an isentropic ensemble live at different depths. Therefore, when considering ensemble members with the same potential density, the ensemble mean depth is the average over the different members. For the case of a two-member ensemble, as shown here, $2\bar{z}^\rho(\rho) = z(1, \rho) + z(2, \rho)$, where the depth $z(1, \rho)$ is generally different from $z(2, \rho)$.

Given the above definitions for the vertical position, we define the *vertical/isopycnal GLM* for an arbitrary function

$$\tilde{\Phi}(x, y, \bar{z}^{(\sigma)}, t) \equiv \overline{\Phi(x, y, \bar{z}^{(\sigma)} + \xi(x, y, \sigma, t), t)}^{(\sigma)}. \quad (49.89)$$

As defined, the vertical/isopycnal GLM, $\tilde{\Phi}$, is a function of the ensemble mean vertical position, $\bar{z}^{(\sigma)}$ (left hand side), and is determined by an ensemble mean of Φ sampled at the depth of each ensemble member, $\bar{z}^{(\sigma)} + \xi(\sigma)$. [McDougall and McIntosh \(2001\)](#) refer to the vertical/isopycnal GLM (49.89) as the *modified mean*.

Relating the modified mean to the isopycnal mean

Following the general result (49.34), we know that the modified mean potential density, $\tilde{\sigma}(x, y, \bar{z}^{(\sigma)}, t)$, equals to the potential density of each ensemble member so that

$$\tilde{\sigma}(x, y, \bar{z}^{(\sigma)}, t) = \sigma(x, y, \bar{z}^{(\sigma)} + \xi(x, y, \sigma, t), t). \quad (49.90)$$

This relation means that the modified mean potential density is the functional inverse of the isopycnal ensemble mean vertical position. Consequently, the isopycnal ensemble mean of a function, $\overline{\Phi}^{(\sigma)}$ (equation (49.84)), when evaluated at the modified mean potential density, $\tilde{\sigma}$, equals to the modified mean $\tilde{\Phi}$ when evaluated at the vertical position of the mean density

$$\overline{\Phi}^{(\sigma)}(x, y, \tilde{\sigma}, t) = \tilde{\Phi}(x, y, \bar{z}^{(\sigma)}, t). \quad (49.91)$$

This is a very important identity that will be used in the following.

49.4.3 Transformed residual mean (TRM)

When working with isopycnal layers, it is very useful to use thickness weighting to account for the net amount of material within a layer, or to measure the net transport in the layer. We make use of thickness weighted fields, $h\Phi$, and the corresponding thickness weighted isopycnal ensemble mean

$$\widehat{\Phi}(\sigma) = \frac{\overline{h\Phi}^{(\sigma)}}{\overline{h}}. \quad (49.92)$$

The identity (49.90) then renders

$$\overline{\Phi}^{\#}(x, y, \bar{z}^{(\sigma)}, t) \equiv \widehat{\Phi}(x, y, \tilde{\sigma}, t), \quad (49.93)$$

where $\overline{\Phi}^{\#}$ is the *transformed residual mean* (TRM) evaluated at the isopycnal ensemble mean vertical position. This is yet another important identity that will be used in the following.

Depth integrated TRM transport

A particularly key TRM field is the TRM horizontal velocity

$$\hat{\mathbf{u}}(x, y, \tilde{\sigma}, t) = \bar{\mathbf{u}}^\#(x, y, \bar{z}^{(\sigma)}, t). \quad (49.94)$$

Following the discussion of the vertical gauge in Section 48.7.1 (see in particular equation (48.77)), we are led to define the depth integrated TRM transport

$$\bar{\mathbf{U}}^\#(\bar{z}^{(\sigma)}) = \int_{-H}^{\tilde{\sigma}(\bar{z}^{(\sigma)})} \bar{\mathbf{u}}^\#(z) dz = \int_{\sigma(-H)}^{\tilde{\sigma}(\bar{z}^{(\sigma)})} \hat{\mathbf{u}}(\gamma) \bar{h}^{(\gamma)} d\gamma, \quad (49.95)$$

with the second equality following from a change of coordinates from geopotential to isopycnal. We can go further with this expression by writing

$$\bar{\mathbf{U}}^\#(\bar{z}^{(\sigma)}) = \int_{\sigma(-H)}^{\tilde{\sigma}(\bar{z}^{(\sigma)})} \hat{\mathbf{u}}(\gamma) \bar{h}^{(\gamma)} d\gamma \quad \text{from equation (49.95)} \quad (49.96a)$$

$$= \int_{\sigma(-H)}^{\tilde{\sigma}(\bar{z}^{(\sigma)})} \bar{\mathbf{u}} h^{(\gamma)} d\gamma \quad \text{from equation (49.92)} \quad (49.96b)$$

$$= \int_{\sigma(-H)}^{\sigma(\bar{z}^{(\sigma)} + \xi)} \bar{\mathbf{u}} h^{(\gamma)} d\gamma \quad \text{from equation (49.90).} \quad (49.96c)$$

The final equality makes it clear that the TRM transport, $\bar{\mathbf{U}}^\#(\bar{z}^{(\sigma)})$, is the ensemble mean volume transport for fluid denser than $\sigma(\bar{z}^{(\sigma)} + \xi) = \tilde{\sigma}(\bar{z}^{(\sigma)})$. This transport can also be written using geopotential coordinates

$$\bar{\mathbf{U}}^\#(\bar{z}^{(\sigma)}) = \overline{\int_{-H}^{\bar{z}^{(\sigma)} + \xi} \mathbf{u} dz}. \quad (49.97)$$

The transport from each ensemble member is determined by integrating from the bottom to the depth, $\bar{z}^{(\sigma)} + \xi$, and then the TRM transport is determined by computing the ensemble mean for this transport.

Quasi-Stokes transport

The TRM transport (49.97) can be decomposed into an Eulerian mean plus the correlation of a fluctuation

$$\bar{\mathbf{U}}^\#(\bar{z}^{(\sigma)}) \equiv \bar{\mathbf{U}}(\bar{z}^{(\sigma)}) + \bar{\mathbf{U}}^{\text{qs}}(\bar{z}^{(\sigma)}). \quad (49.98)$$

The first term,

$$\bar{\mathbf{U}}(\bar{z}^{(\sigma)}) = \int_{-H}^{\bar{z}^{(\sigma)}} \mathbf{u} dz \quad (49.99)$$

is the ensemble mean transport between the bottom and the ensemble mean depth, $\bar{z}^{(\sigma)}$. We interpret this transport as an Eulerian mean since the depth ranges are fixed. In contrast, the *quasi-Stokes* transport

$$\bar{\mathbf{U}}^{\text{qs}}(\bar{z}^{(\sigma)}) \equiv \overline{\int_{\bar{z}^{(\sigma)}}^{\bar{z}^{(\sigma)} + \xi} \mathbf{u} dz} \quad (49.100)$$

measures the ensemble mean transport between the mean vertical position of an isopycnal, $\bar{z}^{(\sigma)}$, and that of each ensemble member, $\bar{z}^{(\sigma)} + \xi(\sigma)$. We refer to transport as “quasi-Stokes” given that

is the difference between an isopycnal (i.e., quasi-Lagrangian) mean and an Eulerian mean (see Section 49.2)

$$\bar{U}^{\text{qs}} = \bar{U}^\# - \bar{U}. \quad (49.101)$$

As for the traditional Stokes drift discussed in Sections 30.3, and 49.2.5, which arises from a correlation between larger velocity when a wave crest is present, so too does the quasi-Stokes transport arise from a correlation between a larger velocity and a larger undulation in isopycnal thickness.

Three-component TRM velocity

Following from the vertical gauge expression (48.76), we introduce the TRM vector streamfunction

$$\bar{\Psi}^\# = \bar{U}^\# \wedge \hat{z}, \quad (49.102)$$

and the corresponding three-dimensional non-divergent TRM velocity

$$\bar{v}^\# = \nabla \wedge \bar{\Psi}^\#. \quad (49.103)$$

The vertical component,

$$\bar{w}^\# = \hat{z} \cdot (\nabla \wedge \bar{\Psi}^\#), \quad (49.104)$$

has no corresponding component in an isopycnal description, which only requires the horizontal thickness weighted transport, \hat{u} . However, the TRM vector streamfunction only requires the horizontal TRM transport, $\bar{U}^\#$, so the two descriptions in effect make use of the same number of degrees of freedom.

49.4.4 Volume conservation and the thickness equation

Consider two perspectives on volume conservation: one based on isopycnal coordinates and the other based on geopotential coordinates.

Isopycnal coordinates

In isopycnal vertical coordinates, the volume of a fluid parcel is written

$$\delta V = \delta x \delta y \delta z = \delta x \delta y \delta \sigma h, \quad (49.105)$$

where we introduced the specific thickness

$$h = \frac{\partial z}{\partial \sigma}. \quad (49.106)$$

As discussed in Section 43.8.1, specific thickness is the Jacobian of transformation between geopotential coordinates, (x, y, z, t) , and isopycnal coordinates, (x, y, σ, t) . For stably stratified ideal fluids, h is one-signed, hence making the coordinate transformation well defined. It is also related to the buoyancy frequency through (Section 25.5.5)

$$N^2 = -\frac{g}{\rho_0} \frac{\partial \sigma}{\partial z} = -\frac{g}{\rho_0 h} \quad (49.107)$$

Geometrically, the product $|h \delta \sigma|$ represents the vertical distance, or *thickness*, between the two infinitesimally close density classes σ and $\sigma + \delta \sigma$. Material conservation of both volume and potential

density implies conservation of the product of specific thickness and horizontal area $\delta x \delta y h$, which leads to the thickness equation (Section 46.1.3)

$$\frac{\partial h}{\partial t} + \nabla_\sigma \cdot (h \mathbf{u}) = 0, \quad (49.108)$$

with \mathbf{u} the horizontal velocity field, the time derivative is computed with σ held fixed, and

$$\nabla_\sigma = \nabla_z + \mathbf{S} \frac{\partial}{\partial z} \quad (49.109)$$

is the horizontal derivative operator with σ held fixed and

$$\mathbf{S} = \nabla_\sigma z \quad (49.110)$$

is the horizontal slope of the potential density surface.

Geopotential coordinates

An Eulerian z -coordinate description of volume stirring within isopycnal layers is rendered via a combination of volume conservation, $\nabla \cdot \mathbf{v} = 0$, and material conservation of potential density, $D\sigma/Dt = 0$. When written as skewson rather than advection, the natural gauge is the vertical gauge introduced in Section 48.7.1 (equation (48.74)), since this gauge only requires the same horizontal velocity field \mathbf{u} used with the isopycnal coordinate description. This gauge has an associated skew flux of potential density $\mathbf{F}^{(\text{skew})} = -\nabla\sigma \wedge \Psi$, which leads to the evolution

$$\frac{\partial \sigma}{\partial t} = \nabla \cdot (\nabla\sigma \wedge \Psi), \quad (49.111)$$

where all derivatives are here taken with fixed Eulerian (geopotential) coordinates, (x, y, z) , and the divergence operator is three-dimensional.

49.4.5 Ensemble mean kinematics in isopycnal coordinates

Consider an ensemble of stably stratified (so that the layer specific thickness h is single-signed and nonvanishing) perfect Boussinesq fluid parcels with the same infinitesimal volume, $\delta V = \delta x \delta y \delta z = \delta x \delta y h \delta\sigma$, and same potential density, σ . Lacking any other marker, such as a tracer concentration, the ensemble members are distinguished from one another by values of their horizontal area, $\delta A = \delta x \delta y$, and their specific thickness, h , that is, their geometric attributes. The ensemble members are assumed to be stirred by different stochastic realizations of the fluid flow. Since each flow realization alters the geometric properties of the parcels, a mean field description focuses on the mean of these geometric properties.

In isopycnal coordinates, (x, y, σ, t) , the thickness equation (49.108) is satisfied by each ensemble member

$$\frac{\partial h}{\partial t} + \nabla_\sigma \cdot (h \mathbf{u}) = 0. \quad (49.112)$$

The ensemble mean computed over these fluid parcels with potential density σ satisfies

$$\partial_t \bar{h}^{(\sigma)} + \nabla_\sigma \cdot \left(\bar{h}^{(\sigma)} \bar{\mathbf{u}}^{(\sigma)} + \bar{h}' \mathbf{u}'^{(\sigma)} \right) = 0, \quad (49.113)$$

where primed variables represent deviations from the isopycnal mean. The mean specific thickness $\bar{h}^{(\sigma)}$ of parcels with potential density σ therefore satisfies the conservation equation

$$\partial_t \bar{h}^{(\sigma)} + \nabla_\sigma \cdot (\bar{h}^{(\sigma)} \hat{\mathbf{u}}) = 0. \quad (49.114)$$

FIGURE 49.4: Schematic of the ensemble averaged potential density as measured by an observer at a fixed point (x, y, z, t) in space-time. In general, different members of the ensemble have potential density surfaces that live at different depths. That is, a fixed Eulerian space-time observer measures an ensemble mean potential density as the average over different potential density surfaces. For the case of a two-member ensemble as shown here, $2\bar{\rho}^z(z) = \rho(1, z) + \rho(2, z)$, where $\rho(1, z)$ is generally different from $\rho(2, z)$.

In this equation we introduced the thickness weighted isopycnal ensemble mean horizontal velocity

$$\hat{\mathbf{u}} = \frac{\bar{h}\mathbf{u}^{(\sigma)}}{\bar{h}^{(\sigma)}} = \bar{\mathbf{u}}^{(\sigma)} + \frac{\bar{h}'\mathbf{u}'^{(\sigma)}}{\bar{h}^{(\sigma)}} \equiv \bar{\mathbf{u}}^{(\sigma)} + \mathbf{u}^{\text{bolus}}, \quad (49.115)$$

along with the isopycnal ensemble mean horizontal velocity, $\bar{\mathbf{u}}^{(\sigma)}$, and the horizontal *bolus velocity*, $\mathbf{u}^{\text{bolus}}$ originally introduced by [Rhines \(1982\)](#). The bolus velocity for an isopycnal layer corresponds to the transport

$$\bar{h}^{(\sigma)} \mathbf{u}^{\text{bolus}} = \bar{h}^{(\sigma)} (\hat{\mathbf{u}} - \bar{\mathbf{u}}^{(\sigma)}) = \bar{h}'\mathbf{u}'^{(\sigma)} \quad (49.116)$$

arises from the along-isopycnal correlations between specific thickness and horizontal velocity.

Quite conveniently, the mean conservation equation (49.114) takes the *same* mathematical form as the conservation equation (49.112) satisfied by each ensemble member. The key difference is that the isopycnal ensemble mean thickness $\bar{h}^{(\sigma)}$ is stirred by the thickness weighted isopycnal ensemble mean horizontal velocity $\hat{\mathbf{u}}$, whereas the thickness of each ensemble member is stirred by a randomly different realization of the horizontal velocity \mathbf{u} . The simplicity of the mean field description (49.114) is afforded by use of the Lagrangian vertical coordinate σ .

49.4.6 Ensemble mean kinematics in geopotential coordinates

We now consider a geopotential coordinate description of the isopycnal ensemble. For this purpose, we interpret a vertical position, z , as the ensemble mean vertical position, $\bar{z}^{(\sigma)}$. Consequently, mean fields defined at the fixed vertical position correspond to either modified mean fields when not thickness weighted (equation (49.89)), or TRM fields when thickness weighted (equation (49.93)).

Evolution of modified mean density

Following the skewson formulation from Section 48.7, at the ensemble mean depth $z = \bar{z}^\rho$, the streamfunction $\bar{\Psi}^\#$ defines an effective skew flux of the modified mean potential density given by

$$\bar{\mathbf{F}}^\# = -\nabla \tilde{\sigma} \wedge \bar{\Psi}^\#. \quad (49.117)$$

Using the identity $\bar{\Psi}^\# = \bar{\mathbf{U}}^\# \wedge \hat{\mathbf{z}}$, we can write this expression in one of the forms

$$\bar{\mathbf{F}}^\# = -\bar{\mathbf{U}}^\# \partial_z \tilde{\sigma} + \hat{\mathbf{z}} \bar{\mathbf{U}}^\# \cdot \nabla_z \tilde{\sigma} \quad (49.118a)$$

$$= -(\bar{\mathbf{U}}^\# + \hat{\mathbf{z}} \mathbf{S} \cdot \bar{\mathbf{U}}^\#) \partial_z \tilde{\sigma}, \quad (49.118b)$$

where

$$\mathbf{S} = -\frac{\nabla_z \tilde{\sigma}}{\partial_z \tilde{\sigma}} \quad (49.119)$$

is the slope of the modified mean density field and $\nabla_z = (\partial_x, \partial_y, 0)$ is the horizontal gradient operator taken with constant depth $z = \bar{z}^{(\sigma)}$. The convergence of the effective skew flux leads to a stirring of the modified mean density $\tilde{\sigma}$ at the mean depth $z = \bar{z}^{(\sigma)}$,

$$\frac{\partial \tilde{\sigma}}{\partial t} = \nabla \cdot (\nabla \tilde{\sigma} \wedge \overline{\Psi}^{\#}). \quad (49.120)$$

This equation represents an geopotential coordinate specification of the evolution of the modified mean density due to stirring by the mean eddies. It corresponds directly to the evolution equation (49.111) satisfied at depth z by a single member of the ensemble.

49.4.7 Approximate ensemble mean kinematics in geopotential coordinates

Equation (49.120) represents an exact z -coordinate description of the stirring of modified mean potential density. However, when working in geopotential coordinates, all that is available is Eulerian information. Hence, the Lagrangian information used to realize this exact description must be approximated.

Estimating the quasi-Stokes transport

The approximation requires us to estimate the quasi-Stokes transport \overline{U}^{qs} defined by equation (49.100). We addressed a similar estimation in Section 49.2.4 when discussing the Stokes mean. Here, we expand the TRM transport in a Taylor series about the geopotential $z = \bar{z}^{(\sigma)}$

$$\overline{U}^{\#}(z) = \overline{\int_{-H}^{z+\xi} \mathbf{u}(s) ds} \quad (49.121a)$$

$$= \overline{U}(z) + \overline{\mathbf{u} \xi}^{(z)} + \frac{1}{2} \overline{\partial_z \mathbf{u} \xi \xi}^{(z)} + \mathcal{O}(\alpha^3), \quad (49.121b)$$

where neglected terms are third order in deviation quantities. Note that all ensemble means are taken at fixed vertical position, which accords with taking a Taylor series about the ensemble mean depth $z = \bar{z}^{(\sigma)}$.

The ensemble means in equation (49.121b) are interpreted as follows. The first term is the Eulerian mean horizontal transport passing beneath the ensemble mean depth, $z = \bar{z}^{(\sigma)}$. The second term, $\overline{\mathbf{u} \xi}$ is the horizontal velocity evaluated at the ensemble mean depth and multiplied by the deviation, ξ , of the potential density surface from its mean depth, all averaged at fixed depth. An Eulerian split of the horizontal velocity \mathbf{u} into its Eulerian mean $\overline{\mathbf{u}}^{(z)}$ and deviation \mathbf{u}' leads to the correlation

$$\overline{\mathbf{u} \xi}^{(z)} = \overline{\mathbf{u}' \xi}^{(z)}. \quad (49.122)$$

For the second order term, similar considerations lead to

$$\overline{\partial_z \mathbf{u} \xi \xi}^{(z)} \approx \partial_z \overline{\mathbf{u}^z} \overline{\xi \xi}^{(z)}, \quad (49.123)$$

where neglected terms are third order and higher. Combining these relations leads to the second order accurate expression

$$\overline{U}^{\#} \approx \overline{U} + \overline{\mathbf{u}' \xi}^{(z)} + \frac{1}{2} \overline{\xi \xi}^z \partial_z \overline{\mathbf{u}^{(z)}}. \quad (49.124)$$

The disturbance field

Following the discussion in Section 49.2.6, we here determine the disturbance field, ξ , in terms of fields at constant depth. For this purpose, use the identity (49.90) to give

$$\tilde{\sigma}(z) = \sigma(z + \xi) \quad (49.125a)$$

$$= \sigma(z) + \partial_z \sigma(z) \xi + \frac{1}{2} \partial_{zz} \sigma(z) \xi^2 + \mathcal{O}(\alpha^3). \quad (49.125b)$$

Subtracting the Eulerian mean of equation (49.125b) from the unaveraged equation (49.125b), and noting that $\tilde{\sigma}$ is already a mean field, leads to the second order accurate expression for the deviation

$$\xi = -\sigma'(z)/\partial_z \bar{\sigma}^{(z)} + \mathcal{O}(\alpha^2), \quad (49.126)$$

where

$$\sigma(z) = \bar{\sigma}^{(z)} + \sigma'(z). \quad (49.127)$$

To within the same order, the deviation can be written

$$\xi = -\sigma'(z)/\partial_z \tilde{\sigma}(z) + \mathcal{O}(\alpha^2). \quad (49.128)$$

Approximate quasi-Stokes transport

Substituting the deviation (49.128) into the approximate expression (49.121b) for the TRM transport yields an approximate expression for the Stokes transport

$$\mathbf{U}^{\text{qs}} = -\frac{\overline{\mathbf{u}' \sigma'}^{(z)}}{\partial_z \tilde{\sigma}} + \frac{\overline{\phi}^{(z)} \partial_z \bar{\mathbf{u}}^{(z)}}{(\partial_z \tilde{\sigma})^2} + \mathcal{O}(\alpha^3), \quad (49.129)$$

where

$$\overline{\phi}^{(z)} = \frac{1}{2} \overline{\sigma' \sigma'}^{(z)} \quad (49.130)$$

is the mean potential density variance. [McDougall and McIntosh \(2001\)](#) noted that the [Gent et al. \(1995\)](#) scheme offers a parameterization of the two correlations on the right hand side of equation (49.129). We have more to say regarding this parameterization in Section 50.1.

Substituting the deviation (49.128) into the approximate expression (49.125b) yields, to within terms of third order, the relation

$$\tilde{\sigma} = \bar{\sigma}^{(z)} - \partial_z \left[\frac{\overline{\phi}^{(z)}}{\partial_z \bar{\sigma}^{(z)}} \right] + \mathcal{O}(\alpha^3). \quad (49.131)$$

As for the Stokes transport, the modified mean density and Eulerian mean density, when evaluated at the same depth, differ by terms that are second order in eddy amplitude.

49.4.8 Further study

This section is largely based on approaches used by [DeSzoeke and Bennett \(1993\)](#), [McIntosh and McDougall \(1996\)](#), [Kushner and Held \(1999\)](#), and [McDougall and McIntosh \(2001\)](#) as summarized in Section 9.3 of [Griffies \(2004\)](#). Many other papers have applied this formalism to a variety of analyses, with examples including [Nurser and Lee \(2004a\)](#), [Nurser and Lee \(2004b\)](#), [Wolfe \(2014\)](#).

49.5 Mean tracer equation

We now include a tracer field to the ideal Boussinesq parcel and determine a mean field description for the tracer. The transport of tracer by eddies has both a reversible stirring component and an irreversible mixing component. The stirring arises from both the thickness correlation to velocity as well as the velocity correlated with tracer.

49.5.1 Thickness weighted means

Equation (49.115) introduced a specific thickness weighted mean operator, which will prove to be quite useful when considering the mean tracer equation. In general, for any field Φ associated with a potential density layer σ , we define the decomposition into thickness weighted mean and deviation

$$\Phi(\sigma) = \widehat{\Phi}(\sigma) + \Phi''(\sigma) \quad (49.132a)$$

$$= \frac{\overline{h \Phi}^{(\sigma)}}{\overline{h}^{(\sigma)}} + \Phi''. \quad (49.132b)$$

It follows by definition that

$$\overline{h \Phi''}^{(\sigma)} = 0. \quad (49.133)$$

49.5.2 Isopycnal mean thickness weighted tracer equation

When attaching a tracer to fluid parcels, each member of the ensemble satisfies the isopycnal tracer equation

$$\frac{\partial C}{\partial t} + \mathbf{u} \cdot \nabla_\sigma C = 0. \quad (49.134)$$

Combining the tracer and thickness equations leads to the thickness weighted tracer equation

$$\frac{\partial(hC)}{\partial t} + \nabla_\sigma \cdot (h \mathbf{u} C) = 0. \quad (49.135)$$

Hence, in isopycnal coordinates and in the absence of irreversible processes, the evolution of thickness weighted tracer occurs via the isopycnally oriented convergence of the two-dimensional thickness weighted horizontal advective flux, $h \mathbf{u} C$.

To address the problem of describing the ensemble mean tracer equation in isopycnal coordinates, decompose the tracer and velocity field into their thickness weighted average and deviation to give

$$\partial_t[h(\widehat{C} + C'')] + \nabla_\sigma \cdot [h(\widehat{\mathbf{u}} + \mathbf{u}'')](\widehat{C} + C'') = 0. \quad (49.136)$$

Taking an ensemble average over fluid elements with the same potential density, and using equation (49.133), yield the mean thickness weighted tracer equation

$$\partial_t(\overline{h}^{(\sigma)} \widehat{C}) + \nabla_\sigma \cdot (\overline{h}^{(\sigma)} \widehat{C} \widehat{\mathbf{u}}) = -\nabla_\sigma \cdot (\overline{h C'' u''}^{(\sigma)}). \quad (49.137)$$

Now introduce the correlation

$$\overline{h C'' u''}^{(\sigma)} = \overline{h}^{(\sigma)} \widehat{C'' u''} \quad (49.138)$$

(see equation (49.132b)), and recall that the mean thickness $\overline{h}^{(\sigma)}$ satisfies the mean thickness equation (49.114). These two points lead to the evolution equation for the mean thickness weighted tracer concentration

$$(\partial_t + \widehat{\mathbf{u}} \cdot \nabla_\sigma) \widehat{C} = -\frac{1}{\overline{h}^{(\sigma)}} \nabla_\sigma \cdot (\overline{h}^{(\sigma)} \widehat{C'' u''}). \quad (49.139)$$

49.5.3 Subgrid scale tracer transport tensor

The correlation between tracer and velocity found on the right-hand side of the mean thickness weighted tracer equation (49.139) is typically written in terms of a subgrid scale tracer transport tensor

$$\widehat{C'' \mathbf{u}''} = -\mathbb{J} \cdot \nabla_{\sigma} \widehat{C}. \quad (49.140)$$

This definition leads to the evolution equation

$$(\partial_t + \widehat{\mathbf{u}} \cdot \nabla_{\sigma}) \widehat{C} = \frac{1}{\bar{h}^{(\sigma)}} \nabla_{\sigma} \cdot (\bar{h}^{(\sigma)} \mathbb{J} \cdot \nabla_{\sigma} \widehat{C}). \quad (49.141)$$

The subgrid scale operator on the right hand side has the same general form as the diffusion operator written in isopycnal coordinates as derived in Section 43.14. However, in addition to symmetric diffusion processes, this operator includes skewed fluxes that lead to skew diffusion as discussed in Section 49.3.2. Whereas the diffusive aspect is commonly parameterized as dianeutral diffusion and neutral diffusion (Section 50.1), there is no parameterization for the skewed correlations for use in ocean models. We comment further on this situation in Section 50.3.8.

49.5.4 Mean tracer transport beneath a density surface

It is useful to further elucidate the relevance of mean thickness weighted fields. For this purpose, consider the mean horizontal tracer transport occurring beneath a particular potential density surface $\sigma = \tilde{\sigma}$,

$$\overline{\mathbf{C}}^{\#}(\bar{z}^{(\sigma)}) = \overline{\int_{-H}^{\bar{z}^{(\sigma)} + \xi} C \mathbf{u} \, dz}. \quad (49.142)$$

Setting tracer concentration to unity recovers the expression (49.97) for the TRM transport. Changing coordinates and making use of the tracer correlation tensor renders

$$\overline{\mathbf{C}}^{\#}(\bar{z}^{(\sigma)}) = \int_{\sigma(-H)}^{\tilde{\sigma}(\bar{z}^{(\sigma)})} \overline{C \mathbf{u} h^{(\sigma)}} \, d\sigma \quad (49.143a)$$

$$= \int_{\sigma(-H)}^{\tilde{\sigma}(\bar{z}^{(\sigma)})} \bar{h}^{(\sigma)} \, d\sigma (\widehat{C} \widehat{\mathbf{u}} + \widehat{C'' \mathbf{u}''}) \quad (49.143b)$$

$$= \int_{\sigma(-H)}^{\tilde{\sigma}(\bar{z}^{(\sigma)})} \bar{h}^{(\sigma)} \, d\sigma (\widehat{C} \widehat{\mathbf{u}} - \mathbb{J} \cdot \nabla_{\sigma} \widehat{C}) \quad (49.143c)$$

$$= \int_{-H}^{\bar{z}^{(\sigma)}} dz (\widehat{C} \widehat{\mathbf{u}} - \mathbb{J} \cdot \nabla_{\sigma} \widehat{C}). \quad (49.143d)$$

Hence, the mean thickness weighted fields naturally appear when considering such physically interesting quantities as the mean horizontal transport of a tracer beneath the modified mean potential density surface.

49.5.5 Summary of the tracer parameterization problem

Traditionally, the isopycnal parameterization problem for the evolution of the mean thickness weighted tracer requires a parameterization of the bolus velocity $\mathbf{u}^{\text{bolus}}$, which again is related to the thickness weighted horizontal velocity via

$$\widehat{\mathbf{u}}((\sigma)) = \frac{\overline{h' \mathbf{u}'^{(\sigma)}}}{\bar{h}^{(\sigma)}} = \overline{\mathbf{u}}^{(\sigma)} + \frac{\overline{h' \mathbf{u}'^{(\sigma)}}}{\bar{h}^{(\sigma)}} = \overline{\mathbf{u}}^{(\sigma)} + \mathbf{u}^{\text{bolus}}. \quad (49.144)$$

In addition to the bolus velocity, it is necessary to parameterize the subgrid scale tracer transport tensor

$$\widehat{C'' \mathbf{u}''} = -\mathbb{J} \cdot \nabla_\sigma \widehat{C}, \quad (49.145)$$

which generally has symmetric (diffusive) and antisymmetric (stirring) components (Section 49.3).

For a geopotential coordinate description, equation (49.93) is used to relate thickness weighted mean fields, defined as a function of σ , and TRM fields, defined as a function of the mean vertical position of σ , to write for the tracer field

$$\widehat{C}(x, y, \tilde{\sigma}, t) = \overline{C}^\#(x, y, \bar{z}^{(\sigma)}, t). \quad (49.146)$$

Equation (49.146), and the developed formalism, leads to the mean field tracer equation in geopotential coordinates

$$\partial_t \overline{C}^\# = \nabla \cdot (\nabla \overline{C}^\# \wedge \overline{\Psi}^\#) + R(\overline{C}^\#), \quad (49.147)$$

where $R(\overline{C}^\#)$ is the geopotential coordinate form of the mixing/stirring operator on the right-hand side of equation (49.141). Details for the transformation of the mixing/stirring opertor from isopycnal to geopotential coordinates are provided in Section 43.14.

49.5.6 Comments

Much in this section follows from [Smith \(1999\)](#), [McDougall and McIntosh \(2001\)](#), and [Young \(2012\)](#), each of which focused on the hydrostatic primitive equations assuming a vertically stable buoyancy stratification. The paper by [Young \(2012\)](#) is a milestone in the literature as he succeeded in formulating the ensemble mean primitive equations in a form where only the thickness weighted (residual mean) velocity appears. Prior attempts failed due to their insufficient mathematical framework. Hence, the formulation of [Young \(2012\)](#) eliminates the need to parameterize the bolus velocity or the quasi-Stokes transport since neither appear as separately identified terms.

Even so, realistic ocean general circulation models are not formulated as “residual mean” models. The key reason is that outside of the stably stratified interior, as in boundary layers, thickness weighted averaging is inappropriate. Instead, we need Eulerian averaged fields when formulating boundary layer closures (e.g., [Large et al., 1994](#)). [Young \(2012\)](#) thus provides a compelling method to decompose the flow into eddies and mean within the stably stratified interior. However, it is not sufficient to capture the full suite of flow regimes represented or parameterized by realistic ocean circulation models.



Subgrid scale tracer transport

We are here concerned with the general properties of parameterization of processes affecting tracer distributions. These parameterizations aim to summarize physical processes too small to observe and/or simulate and how they impact on the larger scales. Such parameterizations in the tracer equation generally take the form of subgrid scale advection and diffusion. This *subgrid scale parameterization problem* is broader and deeper than available from a single chapter. Instead, we here aim to synthesize a range of physical and mathematical topics associated with ocean tracers, offering a platform for further study of a vast and growing literature.

READER'S GUIDE FOR THIS CHAPTER

Tracers evolve according to the advection-diffusion equation discussed in Chapter 48 and further unpacked in the tracer kinematics Chapter 49.

- add section on anisotropic neutral diffusion
- add section on anisotropic GM
- Mention the *Veronis effect* as per [Veronis \(1975\)](#), [Böning et al. \(1995\)](#), and [Roberts and Marshall \(1998\)](#).

50.1	Parameterizing eddy induced tracer transport	1144
50.1.1	Framework based on tracer variance cascade	1144
50.1.2	Density changes from molecular diffusion	1144
50.1.3	Mixing from small (or fine) scale processes	1145
50.1.4	Mesoscale eddy-induced stirring and mixing	1145
50.1.5	Synthesis	1148
50.1.6	Lateral versus diapycnal diffusion	1148
50.2	Quasi-Stokes induced tracer stirring	1149
50.2.1	Gent and McWilliams skewson	1150
50.2.2	Local adiabatic dissipation of potential energy	1150
50.2.3	Meridional overturning streamfunction	1152
50.2.4	Connection to form stress	1152
50.2.5	Isopycnal thickness diffusion and GM	1154
50.3	Neutral diffusion	1155
50.3.1	Redi neutral diffusion	1155
50.3.2	Small slope neutral diffusion	1156
50.3.3	Neutral tangent plane neutral diffusion	1157

50.3.4	Neutrality condition	1158
50.3.5	Symmetry condition	1158
50.3.6	GM skewson plus small slope neutral diffusion	1158
50.3.7	Small slope neutral diffusion in generalized vertical coordinates	1159
50.3.8	Comments	1160
50.4	Cabbeling and thermobaricity	1160
50.4.1	Basic manipulations	1161
50.4.2	A tidy form	1161
50.4.3	Cabbeling	1162
50.4.4	Thermobaricity	1162
50.4.5	Comments	1163

50.1 Parameterizing eddy induced tracer transport

In this section, we present a theoretical framework commonly used for parameterizing tracer transport. In turn, we build on the discussion from Section 51.1 to further detail how *in situ* density evolves in the presence of eddy parameterizations.

50.1.1 Framework based on tracer variance cascade

In the presence of turbulent processes, tracer variance directly cascades to the small scales. This cascade is facilitated by reversible stirring from balanced and unbalanced fluctuations (e.g., mesoscale eddies, submesoscale eddies, breaking gravity and lee waves, turbulent boundary layer processes). The cascade to progressively smaller scales eventually reaches the Batchelor scale (order millimetres; e.g., Section 11.5.1 of [Vallis \(2017\)](#)). At this scale, tracer gradients are sufficiently large in magnitude that molecular diffusion can readily act to dissipate tracer variance through irreversible diffusive mixing. Hence, tracer transport at scales larger than the Batchelor scale is dominated by reversible stirring, whereas transport at and below the Batchelor scale is dominated by irreversible mixing from molecular diffusion. This phenomenology provides a constraint on the form of the tracer equation to be used for coarse grained numerical models, where the model grid scale, Δ , is generally much larger than the Batchelor scale.

50.1.2 Density changes from molecular diffusion

Ignoring cross-diffusion processes (see [IOC et al. \(2010\)](#), Section 2.5 of [Olbers et al. \(2012\)](#), and [Graham and McDougall \(2013\)](#) for discussion), the molecular diffusion of Θ and S lead to the material evolution equations

$$\rho \frac{D\Theta}{Dt} = \nabla \cdot [\rho \kappa_\Theta \nabla \Theta] \quad (50.1a)$$

$$\rho \frac{DS}{Dt} = \nabla \cdot [\rho \kappa_S \nabla S], \quad (50.1b)$$

where $\kappa_\Theta > 0$ and $\kappa_S > 0$ are the molecular kinematic diffusivities for Θ and S , respectively. Following equation (51.10), we see that these molecular tracer fluxes lead to the material evolution of *in situ* density

$$\frac{D\rho}{Dt} - \frac{\omega}{c_s^2} = -\alpha \nabla \cdot (\rho \kappa_\Theta \nabla \Theta) + \beta \nabla \cdot (\rho \kappa_S \nabla S) \quad (50.2a)$$

$$= -\nabla \cdot [\rho (\kappa_\Theta \alpha \nabla \Theta - \kappa_S \beta \nabla S)] + \rho (\kappa_\Theta \nabla \Theta \cdot \nabla \alpha - \kappa_S \nabla S \cdot \nabla \beta). \quad (50.2b)$$

Density evolves from molecular tracer diffusion through the convergence of a buoyancy flux as well as through processes associated with the nonlinear equation of state that give rise to spatial dependence for α and β .

50.1.3 Mixing from small (or fine) scale processes

For a model grid scale, Δ , larger than the scale where gravity waves break and dissipate kinetic energy (i.e., tens to hundreds of metres), diffusion is commonly used to parameterize the associated irreversible tracer mixing (e.g., [MacKinnon et al., 2013](#)). Diffusion is also used to parameterize mixing from other small scale processes, such as turbulent boundary layer processes, double-diffusion, breaking leewaves, etc.

Small scale mixing generally takes place in an isotropic manner. Its parameterization thus appears just as for isotropic molecular diffusion given by equation (50.1b), yet with a far larger eddy diffusivity $\kappa \gg \kappa_\Theta, \kappa_s$ that is a function of the flow. This eddy tracer diffusion dissipates tracer variance at the grid scale, and in turn it mixes *in situ* density according to

$$\frac{D\rho}{Dt} - \frac{\omega}{c_s^2} = -\alpha \nabla \cdot (\rho \kappa \nabla \Theta) + \beta \nabla \cdot (\rho \kappa \nabla S) \quad (50.3a)$$

$$= -\nabla \cdot [\rho \kappa (\alpha \nabla \Theta - \beta \nabla S)] + \rho \kappa (\nabla \Theta \cdot \nabla \alpha - \nabla S \cdot \nabla \beta). \quad (50.3b)$$

Note that since vertical stratification is generally much larger than horizontal stratification, the isotropic diffusion operator is commonly approximated by a vertical or diapycnal diffusion operator (see [McDougall et al. \(2014\)](#) for further discussion).

50.1.4 Mesoscale eddy-induced stirring and mixing

Stirring from turbulent scales smaller than the grid scale is commonly parameterized by an eddy-induced stirring velocity, v^* . For mesoscale eddies, such parameterized stirring generally follows a variant of [Gent et al. \(1995\)](#). In addition, mixing is promoted by the direct cascade from stirring. This mixing is parameterized by a diffusion operator distinct from that used for the small scale mixing discussed in Section 50.1.3. The general form of the diffusion operator is inferred in this section.

We mathematically frame our discussion by introducing a second order subgrid scale transport tensor, \mathbb{M} , meant to parameterize both subgrid scale eddy stirring and eddy mixing. With this tensor, the evolution of salinity and Conservative Temperature takes the form

$$\rho \frac{DS}{Dt} = \nabla \cdot (\rho \mathbb{M} \cdot \nabla S) \quad (50.4a)$$

$$\rho \frac{D\Theta}{Dt} = \nabla \cdot (\rho \mathbb{M} \cdot \nabla \Theta). \quad (50.4b)$$

Note that we use the same transport tensor for both S and Θ . This assumption follows the general approach for turbulent transport parameterizations (e.g, [Vallis, 2017](#)), whereby eddies are assumed to act in the same manner on any conserved scalar tracer.

As discussed in Chapter 48, it is useful to decompose the second order transport tensor into the sum of its symmetric and anti-symmetric components

$$\mathbb{M} = \mathbb{K} + \mathbb{A}. \quad (50.5)$$

The symmetric tensor, \mathbb{K} , gives rise to downgradient diffusion whereas the anti-symmetric tensor, \mathbb{A} , gives rise to skew-diffusion or eddy-induced advection.

Mesoscale eddy-induced stirring

The anti-symmetric tensor, \mathbb{A} , contributes to the parameterized transport according to

$$\nabla \cdot (\rho \mathbb{A} \cdot \nabla S) = \partial_m (\rho A^{mn} \partial_n S) \quad (50.6a)$$

$$= \partial_m (\rho A^{mn}) \partial_n S + \rho A^{mn} \partial_m \partial_n S \quad (50.6b)$$

$$= -\rho v^{*n} \partial_n S, \quad (50.6c)$$

where we made use of the Einstein index notation with repeated indices summed over their range $m, n = 1, 2, 3$, and where A^{mn} are the components to the anti-symmetric transport tensor \mathbb{A} . Additionally, we noted that

$$\rho A^{mn} \partial_m \partial_n S = 0 \quad (50.7)$$

since A^{mn} is anti-symmetric whereas $\partial_m \partial_n S$ is symmetric. Finally, we introduced a density-weighted eddy-induced velocity

$$\rho v^{*n} = -\partial_m (\rho A^{mn}). \quad (50.8)$$

Importantly, $\rho \mathbf{v}^*$ has a zero divergence, again due to anti-symmetry of A^{mn}

$$\nabla \cdot (\rho \mathbf{v}^*) = \partial_n (\rho v^{*n}) = -\partial_n \partial_m (\rho A^{mn}) = 0. \quad (50.9)$$

A zero-divergence for $\rho \mathbf{v}^*$ means that it contributes no mass sources or sinks to the fluid.¹

Transport from the anti-symmetric tensor thus adds a means to stir tracers due to unresolved eddy processes. The mathematical form of the stirring can be either through skew-diffusion or through advection (see Section 48.6). Choosing to make use of the advection form allows us to combine the contribution from the anti-symmetric transport tensor with the resolved advection operator, thus resulting in a residual mean material transport equation

$$\rho \frac{D^\dagger S}{Dt} = \nabla \cdot (\rho \mathbb{K} \cdot \nabla S) \quad (50.10a)$$

$$\rho \frac{D^\dagger \Theta}{Dt} = \nabla \cdot (\rho \mathbb{K} \cdot \nabla \Theta), \quad (50.10b)$$

where the residual mean material time derivative is given by

$$\frac{D^\dagger}{Dt} = \frac{\partial}{\partial t} + \mathbf{v}^\dagger \cdot \nabla \quad (50.11)$$

and the residual mean velocity is

$$\mathbf{v}^\dagger = \mathbf{v} + \mathbf{v}^*. \quad (50.12)$$

Making use of the residual mean velocity then leads to the material evolution of the *in situ* density

$$\frac{D\rho}{Dt} - \frac{1}{c_s^2} \frac{Dp}{Dt} = -\rho \mathbf{v}^* \cdot (-\alpha \nabla \Theta + \beta \nabla S) - \alpha \nabla \cdot (\rho \mathbb{K} \cdot \nabla \Theta) + \beta \nabla \cdot (\rho \mathbb{K} \cdot \nabla S), \quad (50.13)$$

which can be written in terms of the residual mean material time derivative

$$\frac{D^\dagger \rho}{Dt} - \frac{1}{c_s^2} \frac{D^\dagger p}{Dt} = -\alpha \nabla \cdot (\rho \mathbb{K} \cdot \nabla \Theta) + \beta \nabla \cdot (\rho \mathbb{K} \cdot \nabla S). \quad (50.14)$$

¹For a Boussinesq fluid, the density factor is replaced by the constant reference density, ρ_0 , so that $\nabla \cdot \mathbf{v}^* = 0$ in the Boussinesq fluid. See section 7 of [Griffies and Greatbatch \(2012\)](#) for more details of the Boussinesq and non-Boussinesq forms for the parameterized eddy-induced transport.

Mesoscale eddy-induced diffusion

Transport from the symmetric tensor, \mathbb{K} , corresponds to diffusion so long as the tensor is positive semi-definite. The diffusion operator in the residual mean evolution equation (50.14) can be written

$$-\alpha \nabla \cdot (\rho \mathbb{K} \cdot \nabla \Theta) + \beta \nabla \cdot (\rho \mathbb{K} \cdot \nabla S) = \nabla \cdot [\rho \mathbb{K} \cdot (-\alpha \nabla \Theta + \beta \nabla S)] + \rho \nabla \alpha \cdot \mathbb{K} \cdot \nabla \Theta - \rho \nabla \beta \cdot \mathbb{K} \cdot \nabla S, \quad (50.15)$$

so that the *in situ* density evolves according to

$$\frac{D^\dagger \rho}{Dt} - \frac{1}{c_s^2} \frac{D^\dagger p}{Dt} = -\underbrace{\nabla \cdot [\rho \mathbb{K} \cdot (\alpha \nabla \Theta - \beta \nabla S)]}_{\text{conservative processes}} + \underbrace{\rho \nabla \alpha \cdot \mathbb{K} \cdot \nabla \Theta - \rho \nabla \beta \cdot \mathbb{K} \cdot \nabla S}_{\text{sources from nonlinear EOS processes}}. \quad (50.16)$$

We now discuss the physical processes associated with the right hand side terms.

- NONLINEAR EQUATION OF STATE: A nonlinear equation of state is characterized by spatially dependent thermal expansion and haline contraction coefficients. Mixing of Θ and S in the presence of a nonlinear equation of state generally gives rise to material evolution of *in situ* density through cabbeling and thermobaricity ([McDougall, 1987b](#)). We offer a summary of these processes in Section 50.4.
- CONSERVATIVE PROCESSES AND NEUTRAL DIFFUSION: A linear equation of state is independent of pressure, so that the evolution equation (50.16) takes the form

$$\frac{D^\dagger \rho}{Dt} = -\nabla \cdot [\rho \mathbb{K} \cdot (\alpha \nabla \Theta - \beta \nabla S)]. \quad (50.17)$$

Under the residual mean transport, density remains materially constant in the absence of any diffusion. Additionally, it remains constant if the diffusive fluxes of Θ and S are density-compensated so that

$$\mathbb{K} \cdot (\alpha \nabla \Theta - \beta \nabla S) = 0. \quad (50.18)$$

As detailed in Section 50.3, various forms of this *neutral diffusion* satisfy this *neutrality condition* even for general equations of state that have pressure dependence.

We thus propose that the mesoscale eddy stirring-induced mixing of Θ and S be parameterized as density-compensated diffusion, otherwise known as neutral diffusion. Neutral diffusion leaves *in situ* density unchanged in the absence of nonlinear equation of state processes so that

$$\frac{D^\dagger \rho}{Dt} - \frac{1}{c_s^2} \frac{D^\dagger p}{Dt} = \underbrace{\rho \nabla \alpha \cdot \mathbb{K} \cdot \nabla \Theta - \rho \nabla \beta \cdot \mathbb{K} \cdot \nabla S}_{\text{sources from nonlinear EOS processes}}. \quad (50.19)$$

What is the evidence for a diffusion operator oriented according to neutral directions? To answer this question, consider a diffusion tensor that does *not* maintain the constraint (50.18). In this case, additional diffusive mixing appears, adding to that already parameterized from small scale mixing processes such as breaking gravity waves. As discussed in Section 14.1.5 of [Griffies \(2004\)](#) as well as Section 1 of [McDougall et al. \(2014\)](#), the extra mixing induced by this non-neutral orientation of the diffusive fluxes is proportional to the squared tangent of the angle between the proposed new direction and the neutral tangent plane. Estimates for interior ocean mixing constrain the magnitude of the tangent to be less than 10^{-4} . This number is very small, indeed it is zero within error bars of field measurements. Measurements thus support the use of a neutral diffusion operator oriented so to respect the constraint (50.18). In Section 50.3 we dive into the details of neutral diffusion.

50.1.5 Synthesis

In summary, the proposed evolution equation for *in situ* density in the presence of subgrid scale processes takes the form

$$\frac{D\rho}{Dt} - \underbrace{\frac{1}{c_s^2} \frac{Dp}{Dt}}_{\text{compressibility}} = -\underbrace{\mathbf{v}^* \cdot (-\alpha \nabla \Theta + \beta \nabla S)}_{\text{eddy-induced stirring}} - \underbrace{\alpha \nabla \cdot (\rho \kappa \nabla \Theta) + \beta \nabla \cdot (\rho \kappa \nabla S)}_{\text{small scale mixing and nonlinear EOS}} + \underbrace{\rho \nabla \alpha \cdot \mathbb{K} \cdot \nabla \Theta - \rho \nabla \beta \cdot \mathbb{K} \cdot \nabla S}_{\text{nonlinear EOS processes from eddy mixing}}. \quad (50.20)$$

We thus have the following physical processes contributing to the evolution of *in situ* density.

- **SMALL SCALE MIXING:** To parameterize mixing induced by the suite of subgrid small scale processes (e.g., breaking gravity waves, lee waves, turbulent boundary layer processes, double diffusion) we introduce an isotropic diffusion operator (50.3b) with an eddy-diffusivity, κ , that is a function of the flow. This diffusivity is the same for all tracers, with the exception of double-diffusive processes whereby material tracers (e.g., salinity, nutrients) have a diffusivity distinct from temperature ([Schmitt, 1994](#)). Given the dominance of vertical stratification over horizontal, it is common to approximate the isotropic diffusion operator with a vertical diffusion operator.
- **EDDY-INDUCED STIRRING:** For subgrid scale stirring, such as from mesoscale (and submesoscale) eddies, we introduce a parameterized eddy-induced advection operator. When combined with the resolved advection, we are led to a residual mean material time derivative, D^\dagger/Dt .
- **EDDY-INDUCED MIXING:** Subgrid scale eddy-induced stirring leads to a direct cascade of Θ and S variance to the small scales. Mixing arising from this cascade is parameterized by neutral diffusion, whereby the diffusive fluxes of Θ and S are density compensated according to the constraint (50.18).
- **NONLINEAR EOS PROCESSES:** Mixing of Θ and S in the presence of a nonlinear equation of state means that *in situ* density evolves due to cabbeling and thermobaricity (Section 50.4). The dominant contributions to these processes arise from eddy-stirring induced mixing (i.e., neutral diffusion) ([McDougall, 1987b](#)), though small scale mixing also has a contribution as seen by writing

$$-\alpha \nabla \cdot (\rho \kappa \nabla \Theta) + \beta \nabla \cdot (\rho \kappa \nabla S) = -\nabla \cdot [\rho \kappa (\alpha \nabla \Theta - \beta \nabla S)] + \rho \kappa (\nabla \alpha \cdot \nabla \Theta - \nabla \beta \cdot \nabla S). \quad (50.21)$$

50.1.6 Lateral versus diapycnal diffusion

What is more important for setting tracer distributions: lateral or diapycnal diffusion? Although the lateral eddy diffusivity is many orders larger than the eddy diapycnal diffusivity, the gradients on which they act are very different. So to help answer the question, consider a scaling in which we consider a constant coefficient lateral diffusivity and a constant coefficient isotropic diffusivity. Furthermore, to simplify the analysis assume Cartesian orientation of the diffusion operators and assume the isotropic diffusion is dominated by vertical diffusion

$$\text{horizontal diffusion} = \kappa_h \nabla_z^2 C \quad \text{vertical diffusion} = \kappa_v \partial_{zz} C. \quad (50.22)$$

Now introduce a vertical scale H and horizontal scale L over which the tracer concentration changes by δC . Doing so leads to the scaled diffusion operators

$$\text{horizontal diffusion} \sim (\kappa_h/L^2) \delta C \quad \text{vertical diffusion} \sim (\kappa_v/H^2) \delta C. \quad (50.23)$$

These operators have the same scale when

$$\kappa_v = (H/L)^2 \kappa_h. \quad (50.24)$$

Choosing $L = 10^5$ m and $H = 10^1$ m leads to

$$\kappa_v = 10^{-8} \kappa_h. \quad (50.25)$$

Furthermore, if $\kappa_h = 10^3$ m² s⁻¹, then the two operators provide a similar contribution to tracer evolution if $\kappa_v = 10^{-5}$ m² s⁻¹. This is a rather small diffusivity that is generally thought to be on the order of that afforded by the background of breaking gravity waves in the ocean interior ([MacKinnon et al., 2013](#)).² This scaling is crude since the length scales are dependent on details of the flow regime as are the eddy diffusivities. Even so, the scaling indicates that even a relatively small turbulent diapycnal diffusivity arising from the background gravity wave spectrum can contribute to tracer distributions a similar amount as from lateral diffusion.

50.2 Quasi-Stokes induced tracer stirring

As mentioned in Section 49.5.3, there are two processes that contribute to eddy-induced stirring. One involves the correlations between eddy fluctuations in the velocity and tracer fields. In Section 49.3, we considered the kinematics of correlations induced by small amplitude eddying motions, where we found that the eddy-induced motion of fluid particles leads to both a symmetric (mixing) and anti-symmetric (stirring) dispersion of tracer concentrations. There is currently no method available for parameterizing this form of eddy-induced stirring when it arises from subgrid scale processes, thus leaving unanswered its importance to large-scale tracer distributions.

The second process leading to eddy-induced stirring arises from correlations between fluctuations in isopycnal layer thickness and horizontal velocity. As detailed in Section 49.4, this second effect leads to a movement of volume between isopycnal layers, or equivalently we can conceive of it as the quasi-Stokes transport of volume arising from transient eddy motion. This eddy-induced volume stirring in turn affects an eddy-induced tracer stirring within isopycnal layers. Transient mesoscale eddies are the canonical dynamical process leading to this form of stirring. For simulations that do not resolve transient mesoscale eddies, we commonly parameterize the subgrid scale stirring through variants of the [Gent et al. \(1995\)](#) scheme. Elements of this scheme are detailed in this section.

In this section we refer to ρ as the potential density, with the assumption of a linear equation of state so that buoyancy is linearly proportional to potential density (Section 25.3.5). This assumption is for convenience only, and can be readily generalized to an arbitrary equation of state. Furthermore, for convenience we assume a Boussinesq fluid, with extensions to non-Boussinesq fluids discussed in section 7 of [Griffies and Greatbatch \(2012\)](#).

²Molecular diffusivities are roughly $\approx 10^{-6}$ m² s⁻¹ for temperature and $\approx 10^{-7}$ m² s⁻¹ for salinity and other material tracers. (SMG: check these values).

50.2.1 Gent and McWilliams skewson

Gent et al. (1995) parameterize the divergent part of the quasi-Stokes transport by setting

$$\mathbf{U}^{\text{qs}} = -\kappa \mathbf{S}. \quad (50.26)$$

In this expression, \mathbf{S} is the slope of the potential density surfaces (equation (49.119)), and $\kappa > 0$ is a kinematic diffusivity (dimensions of velocity times a length). The corresponding three-dimensional non-divergent eddy-induced velocity is given by

$$\mathbf{v}^* = -\partial_z (\kappa \mathbf{S}) + \hat{\mathbf{z}} \nabla_z \cdot (\kappa \mathbf{S}), \quad (50.27)$$

and the antisymmetric stirring tensor is

$$\mathbb{A}_{mn} = \begin{bmatrix} 0 & 0 & -\kappa S_x \\ 0 & 0 & -\kappa S_y \\ \kappa S_x & \kappa S_y & 0 \end{bmatrix}. \quad (50.28)$$

The parameterized skew flux of potential density, ρ , due to the quasi-Stokes transport is given by

$$\mathbf{F}^{\text{skew}} = -\mathbf{U}^{\text{qs}} \partial_z \rho + \hat{\mathbf{z}} (\mathbf{U}^{\text{qs}} \cdot \nabla_z) \rho \quad (50.29a)$$

$$= -\kappa \nabla_z \rho + \hat{\mathbf{z}} S^2 \kappa \partial_z \rho \quad (50.29b)$$

$$= -\kappa \nabla_z \rho - \hat{\mathbf{z}} (\kappa \rho_0 / g) (S N)^2. \quad (50.29c)$$

This parameterization yields horizontal downgradient diffusion of potential density, combined with a vertical upgradient diffusion. So long as the stratification is stable ($N^2 > 0$), the vertical component to the skew flux is vertically downward. Additionally, *Gent et al.* (1995) prescribe a diffusivity that vanishes on all boundaries, including the ocean surface. *McIntosh and McDougall* (1996) and *McDougall and McIntosh* (2001) present more discussion of vertical boundary conditions, which can be understood by considering the exact form of the quasi-Stokes transport defined by equation (49.100).

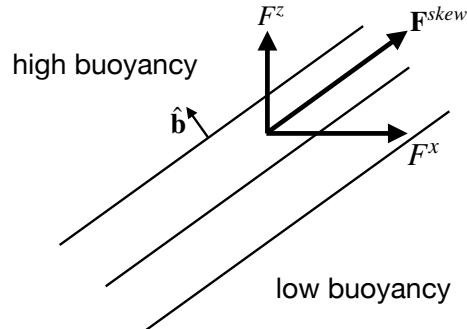


FIGURE 50.1: Orientation of the skew flux for buoyancy as proposed by *Gent et al.* (1995) and described by *Griffies* (1998), where buoyancy is $b = -(g/\rho_0)(\rho - \rho_0)$ so that $\nabla b = -(g/\rho_0)\nabla\rho$. The horizontal flux is downgradient (moving high buoyancy water to low buoyancy) whereas the vertical flux component is upgradient. The net effect is a flux that is oriented parallel to constant buoyancy lines (i.e., skewed relative to the buoyancy gradient).

50.2.2 Local adiabatic dissipation of potential energy

We here consider the effects from the *Gent et al.* (1995) scheme on the potential energy and available potential energy (APE). We express the behavior using both skew fluxes and advective fluxes. Note that since we are assuming the parameterization is adiabatic, the change in potential energy is identical to the change in available potential energy (APE).

Skew flux approach

Let us approach the parameterization problem from the perspective of satisfying two general properties: (I) the subgrid scale operator adiabatically stirs while maintaining the same amount of fluid within isopycnal layers, (II) the subgrid operator locally dissipates potential energy through an adiabatic rearrangement of the density surfaces, with the dissipation vanishing when there is zero baroclinicity. That is, the scheme dissipates available potential energy. What is the form of the stirring operator implied by these two assumptions?

Adiabatic stirring of potential density can be realized via the convergence of a skew flux oriented parallel to potential density surfaces

$$\mathbf{F} = -\nabla\rho \wedge \Psi, \quad (50.30)$$

where Ψ remains to be found. To see what the local dissipation of APE imposes, consider the gravitational potential energy of the adiabatic Boussinesq system

$$P = g \int \rho z \, dV, \quad (50.31)$$

where we assume the *in situ* density equals to the potential density as per a linear equation of state (Section 25.3.5). Assuming all boundaries are material and static allows us to focus on the time tendency of potential energy associated with the unknown flux

$$\frac{dP}{dt} = g \int z \frac{\partial \rho}{\partial t} \, dV \quad (50.32a)$$

$$= -g \int (z \nabla \cdot \mathbf{F}) \, dV \quad (50.32b)$$

$$= -g \int (z \partial_z F^{(z)}) \, dV \quad (50.32c)$$

$$= g \int F^{(z)} \, dV, \quad (50.32d)$$

where $F^{(z)}$ is the vertical flux component. We drop boundary effects by assuming the subgrid scale flux vanishes on all boundaries. To provide a *local* APE sink requires

$$F^{(z)} \leq 0, \quad (50.33)$$

where zero occurs when the isopycnals are flat. It is sufficient to construct the vertical flux component using only the potential density field itself. For a stably stratified fluid in which $\partial_z \rho < 0$, the following form provides a local APE sink

$$F^{(z)} = \kappa S^2 \frac{\partial \rho}{\partial z} = -(\kappa \rho_0/g) (S N)^2 \leq 0, \quad (50.34)$$

where $\kappa > 0$ is a kinematic diffusivity setting the strength of the flux and S^2 is the squared isopycnal slope. The corresponding horizontal flux is given by a downgradient diffusive flux

$$\mathbf{F}^{(h)} = -\kappa \nabla_z \rho. \quad (50.35)$$

We have thus recovered the skew flux (50.29c) as proposed by Gent et al. (1995). Note that Aiki et al. (2004) proceed in a similar manner yet do not assume locality of the APE sink, thus deriving a more general subgrid scale operator.

Advection flux approach

The impacts on potential energy should be the same when representing the parameterization as an advective flux. To verify this result, return to equation (50.32d) and make use of the vertical component of the advective flux rather than the skew flux

$$\frac{dP}{dt} = g \int F^{(z)} dV \quad (50.36a)$$

$$= g \int \rho w^* dV \quad (50.36b)$$

$$= g \int \rho \nabla_z \cdot (\kappa \mathbf{S}) dV \quad (50.36c)$$

$$= g \int \nabla_z \cdot (\rho \kappa \mathbf{S}) dV - g \int \nabla_z \rho \cdot \kappa \mathbf{S} dV \quad (50.36d)$$

$$= -\rho_0 \int \kappa (S N)^2 dV, \quad (50.36e)$$

which is the same result as for the skew flux.

50.2.3 Meridional overturning streamfunction

It is often of interest to compute the net transport of volume across a portion of the ocean. In particular, meridional-depth or meridional-density streamfunctions allow one to visualize and quantify the zonally integrated transport occurring in a closed basin or over the full globe. The quasi-Stokes transport provides a transport in addition to that from the resolved scale Eulerian mean transport, and the parameterization of [Gent et al. \(1995\)](#) leads to a straightforward computation of the quasi-Stokes contribution. For this purpose, write the net meridional transport of fluid across a basin and passing beneath a particular depth in the form (the minus sign is conventional)

$$\mathcal{T}(y, z, t) = - \int dx \int_{-H}^z (v + v^*) dz \quad (50.37a)$$

$$= - \int dx \int_{-H}^z v dz + \int dx \int_{-H}^z \partial_z(\kappa S_y) dz \quad (50.37b)$$

$$= - \int dx \int_{-H}^z v dz + \int \kappa S_y dx \quad (50.37c)$$

$$\equiv \mathcal{T}^{\text{eulerian}}(y, z, t) + \mathcal{T}^{\text{qs}}(y, z, t). \quad (50.37d)$$

For the penultimate step we set the parameterized quasi-Stokes transport to zero at the ocean bottom. We thus see that the parameterized quasi-Stokes transport adds a contribution that scales linearly with basin size, isopyncal slope, and diffusivity,

$$\mathcal{T}^{\text{qs}} \sim L S \kappa. \quad (50.38)$$

As an example, let $\kappa = 10^3 \text{ m}^2 \text{s}^{-1}$, $S = 10^{-3}$, and $L = 10^7 \text{ m}$, which yields $\mathcal{T} \approx 10 \times 10^6 \text{ m}^3 \text{s}^{-1} = 10 \text{ Sv}$. Such transport can represent a nontrivial addition to that from the resolved scale velocity field.

50.2.4 Connection to form stress

We now connect the [Gent et al. \(1995\)](#) closure, normally implemented in the tracer equation, to vertical transfer of momentum through form stress. For this purpose we anticipate our general

discussion of form stress in Section 20.1 and more detailed discussions in Sections 26.7 and 32.4. In those discussions, we identify form stress as the horizontal pressure force acting on a sloped surface, with our present concern with surfaces of constant isopycnals as discussed in Sections 26.7 and 32.4.

[Young \(2012\)](#) provides a general means for making the connection between [Gent et al. \(1995\)](#) and form stress for a continuously stratified fluid. For our more schematic purposes, we follow the treatment in [Greatbatch and Lamb \(1990\)](#), [Gent et al. \(1995\)](#) (their Section 4), [Ferreira and Marshall \(2006\)](#) (their Section 2), and [Zhao and Vallis \(2008\)](#) (their Section 2.2). For this purpose, assume the fluid is in Boussinesq planetary geostrophic balance (detailed in Section 26.5) whereby the horizontal momentum satisfies

$$\rho_o f (\hat{\mathbf{z}} \wedge \mathbf{u}) = -\nabla_z p + \partial_z \boldsymbol{\tau}. \quad (50.39)$$

The Coriolis acceleration balances the acceleration from horizontal pressure gradients plus a vertical transfer of horizontal stress. The horizontal stress term is generally quite small in the ocean interior, where the flow is in geostrophic balance, whereas it is large at the ocean surface where it arises from turbulent air-sea interactions; i.e., wind stress. Furthermore, it can be large at the bottom through turbulent bottom stresses.

To make the connection between [Gent et al. \(1995\)](#) and the vertical transfer of horizontal form stress, add $\rho_o f (\hat{\mathbf{z}} \wedge \mathbf{u}^*)$ to both sides of equation (50.39) to obtain

$$\rho_o f (\hat{\mathbf{z}} \wedge \mathbf{u}^\dagger) = -\nabla_z p + \partial_z \boldsymbol{\tau} + \rho_o f (\hat{\mathbf{z}} \wedge \mathbf{u}^*), \quad (50.40)$$

where $\mathbf{u}^\dagger = \mathbf{u} + \mathbf{u}^*$ is the horizontal residual mean velocity. This equation says that the Coriolis acceleration from the horizontal residual mean velocity balances pressure gradients, vertical frictional stresses, plus the Coriolis acceleration from the eddy-induced velocity. We further unpack the eddy Coriolis acceleration by noting that the planetary geostrophic velocity satisfies the thermal wind relation in the ocean interior (Section 26.5.5), whereby

$$f \frac{\partial \mathbf{u}}{\partial z} = -(g/\rho_0) \hat{\mathbf{z}} \wedge \nabla \rho = -\hat{\mathbf{z}} \wedge N^2 \mathbf{S}, \quad (50.41)$$

with

$$\mathbf{S} = -\frac{\nabla_z \rho}{\partial_z \rho} = \frac{g}{\rho_0} \frac{\nabla_z \rho}{N^2} \quad (50.42)$$

the isopycnal slope. We can thus write the Coriolis acceleration from the eddy-induced velocity in the form

$$f (\hat{\mathbf{z}} \wedge \mathbf{u}^*) = -f [\hat{\mathbf{z}} \wedge \partial_z (\kappa \mathbf{S})] \quad (50.43a)$$

$$= -\partial_z [\hat{\mathbf{z}} \wedge (f \kappa \mathbf{S})] \quad (50.43b)$$

$$= \frac{\partial}{\partial z} \left[\frac{\kappa f^2}{N^2} \frac{\partial \mathbf{u}}{\partial z} \right] \quad (50.43c)$$

$$= \frac{\partial}{\partial z} \left[\nu_e \frac{\partial \mathbf{u}}{\partial z} \right], \quad (50.43d)$$

where the final equality introduced an eddy-induced vertical viscosity

$$\nu_e \equiv \kappa (f^2/N^2). \quad (50.44)$$

Making use of this result in the planetary geostrophic equation (50.40) thus leads to

$$\rho_o f (\hat{\mathbf{z}} \wedge \mathbf{u}^\dagger) = -\nabla_z p + \partial_z (\boldsymbol{\tau} + \boldsymbol{\tau}_e), \quad (50.45)$$

where

$$\tau_e = \rho_0 \nu_e \frac{\partial \mathbf{u}}{\partial z} \quad (50.46)$$

defines a horizontal mesoscale eddy stress arising from the thermal wind shears. Equation (50.45) says that the Coriolis acceleration from the horizontal residual mean velocity is in balance with the horizontal pressure gradient plus the vertical transfer of horizontal shears arising from both friction/wind/bottom drag *plus* a contribution from parameterized mesoscale eddies.

We conclude that the [Gent et al. \(1995\)](#) parameterization appears in the planetary geostrophic residual mean momentum equation as a vertical transport of horizontal stress determined by a viscosity $\nu_e = \kappa(f/N)^2$. Notably, this vertical eddy transfer occurs in the absence of irreversible mixing. We thus interpret it as a parameterization of the vertical transfer of pressure form stress via mesoscale eddies that act between isopycnal layers. That is, the [Gent et al. \(1995\)](#) scheme offers a means to parameterize vertical transfer of horizontal form stress arising from undulating mesoscale eddies in the ocean interior. This interpretation is more thoroughly discussed in Section 26.7.

50.2.5 Isopycnal thickness diffusion and GM

Recall the ensemble mean thickness equation (49.114) derived in Section 49.4.5

$$\partial_t h + \nabla_\rho \cdot (h \hat{\mathbf{u}}) = 0, \quad (50.47)$$

where

$$\hat{\mathbf{u}} = \mathbf{u} + \mathbf{u}^{\text{bolus}} \quad (50.48)$$

is the thickness weighted transport velocity affecting evolution of the ensemble mean thickness h . Note that for brevity we here drop the nomenclature $(\)^{(\rho)}$ used in Section 49.4.5.

Isopycnal correlations of horizontal velocity and layer thickness define the bolus velocity via

$$h \mathbf{u}^{\text{bolus}} = \overline{h' \mathbf{u}'} \quad (50.49)$$

Now consider a downgradient diffusive closure for this correlation

$$h \mathbf{u}^{\text{bolus}} = \overline{h' \mathbf{u}'}^{(\rho)} = -\mathbb{K}_{\text{thick}} \cdot \nabla_\rho h, \quad (50.50)$$

with $\mathbb{K}_{\text{thick}}$ a symmetric and positive-definite 2×2 diffusion tensor. The mean thickness equation thus takes the form of an advection-diffusion equation in isopycnal coordinates

$$\partial_t h + \nabla_\rho \cdot (h \mathbf{u}) = \nabla_\rho \cdot (\mathbb{K}_{\text{thick}} \cdot \nabla_\rho h). \quad (50.51)$$

To make a connection between the thickness diffusion closure (50.50) and the [Gent et al. \(1995\)](#) closure discussed in Section 50.2.1, note that the specific thickness is the inverse of the vertical derivative of the potential density

$$h = (\partial_z \rho)^{-1}. \quad (50.52)$$

Correspondingly, using the relation between derivative operators, $\nabla_\rho = \nabla_z + \mathbf{S} \partial_z$, gives

$$h^{-1} \nabla_\rho h = -h \nabla_\rho (1/h) \quad (50.53a)$$

$$= -(\partial_z \rho)^{-1} (\nabla_z + \mathbf{S} \partial_z) \partial_z \rho \quad (50.53b)$$

$$= -\frac{\partial_z (\nabla_z \rho)}{\partial_z \rho} + \frac{\partial_{zz} \rho \nabla_z \rho}{(\partial_z \rho)^2} \quad (50.53c)$$

$$= \partial_z \mathbf{S}. \quad (50.53d)$$

Consequently, the bolus velocity takes the form

$$\mathbf{u}^{\text{bolus}} = -h^{-1} \mathbb{K}_{\text{thick}} \cdot \nabla_\rho h = -\mathbb{K}_{\text{thick}} \cdot \partial_z \mathbf{S}. \quad (50.54)$$

The special case of depth independent diffusivity

For the special case where $\mathbb{K}_{\text{thick}}$ is independent of depth and proportional to the 2×2 identity matrix, then

$$\mathbf{u}^{\text{bulus}} = -\partial_z (\kappa \mathbf{S}) = \mathbf{u}^*, \quad (50.55)$$

where the horizontal component of the [Gent et al. \(1995\)](#) velocity \mathbf{u}^* was identified from equation [\(50.27\)](#). Again, this identity holds only for the special case of a vertically independent diffusivity tensor proportional to the identity.

Further caveats

The relevance of a depth-independent diffusivity has been questioned by many authors, such as [Killworth \(1997\)](#), [Treguier et al. \(1997\)](#), [Smith and Vallis \(2002\)](#), [Smith and Marshall \(2009\)](#), and [Abernathay et al. \(2013\)](#). We conclude from these studies that a depth independent diffusivity is not the best choice for the [Gent et al. \(1995\)](#) parameterization, in which case where one places the vertical derivative is crucial.

The relation between thickness diffusion with the [Gent et al. \(1995\)](#) parameterization further breaks down near boundaries. The reason is that the eddy diffusivity vanishes next to boundaries and thus has a depth-dependence. Additionally, as noted by [Holloway \(1997\)](#) and [Griffies et al. \(2000\)](#), thickness diffusion next to solid earth boundaries leads to an increase in potential energy, with isopycnals creeping up the topographic slope. Such unphysical behavior motivates isopycnal modelers instead to use *interfacial height* diffusion to dissipate noise in the thickness field.

50.3 Neutral diffusion

Neutral diffusion parameterizes the mixing induced by mesoscale eddy stirring acting preferentially along neutral directions. By construction, the neutral diffusive flux of a tracer is oriented along a neutral direction or a *neutral tangent plane*. As detailed in Section [25.4](#), neutral directions are directions in a stratified fluid that allow for mixing of Θ and S without modifying the locally defined buoyancy. The neutral diffusive tracer flux for an arbitrary tracer, C , is perpendicular to the dianeutral unit vector

$$\mathbf{J} \cdot \hat{\gamma} = 0 \quad (50.56)$$

where (equation [\(25.29\)](#))

$$\hat{\gamma} = \frac{-\alpha \nabla \Theta + \beta \nabla S}{|-\alpha \nabla \Theta + \beta \nabla S|}. \quad (50.57)$$

50.3.1 Redi neutral diffusion

One diffusive flux satisfying the property [\(50.56\)](#) is given by

$$\mathbf{J}_{\text{redi}} = -\rho A [\nabla C - \hat{\gamma} (\hat{\gamma} \cdot \nabla C)], \quad (50.58)$$

where $A > 0$ is the neutral diffusivity (dimensions of squared length per time). We confirm that \mathbf{J}_{redi} is oriented down the tracer gradient by noting that

$$\mathbf{J}_{\text{redi}} \cdot \nabla C = -\rho A [|\nabla C|^2 - (\hat{\gamma} \cdot \nabla C)^2] \leq 0. \quad (50.59)$$

The flux \mathbf{J}_{redi} can be written in a downgradient flux-gradient relation

$$J^m = -\rho \mathbb{K}_{\text{redi}}^{mn} \partial_n C, \quad (50.60)$$

where the Redi diffusion tensor is given by

$$\mathbb{K}_{\text{redi}}^{mn} = \frac{A}{1 + S_x^2 + S_y^2} \begin{bmatrix} 1 + S_y^2 & -S_x S_y & S_x \\ -S_x S_y & 1 + S_x^2 & S_y \\ S_x & S_y & S_x^2 + S_y^2 \end{bmatrix} \quad (50.61)$$

with the corresponding neutral diffusion operator given by the three-dimensional flux convergence

$$\mathcal{R}_{\text{redi}} = -\nabla \cdot \mathbf{J}_{\text{redi}} = \partial_m (\rho \mathbb{K}_{\text{redi}}^{mn} \partial_n C). \quad (50.62)$$

In the Redi tensor (50.61) we introduced the components of the horizontal vector, $\mathbf{S} = (S_x, S_y, 0)$, with

$$\mathbf{S} = - \begin{bmatrix} -\alpha \nabla_z \Theta + \beta \nabla_z S \\ -\alpha \partial_z \Theta + \beta \partial_z S \end{bmatrix} = \frac{g (-\alpha \nabla_z \Theta + \beta \nabla_z S)}{N^2} \quad (50.63)$$

the slope of the neutral tangent plane relative to the (x, y) horizontal plane, with

$$N^2 = -g [-\alpha \partial_z \Theta + \beta \partial_z S] \quad (50.64)$$

the squared buoyancy frequency (Section 25.4). Notably, it is useful to introduce the slope vector only when the fluid is stratified in the vertical so that the slope magnitude, $|\mathbf{S}|$, is finite.

50.3.2 Small slope neutral diffusion

Another form of the neutral diffusion flux is based on assuming a small magnitude for the slope of the neutral tangent plane relative to the horizontal, which is the case for most of the ocean interior even in frontal regions. With this approximation, the small slope neutral diffusion tensor takes the form

$$\mathbb{K}_{\text{small}}^{mn} = A \begin{bmatrix} 1 & 0 & S_x \\ 0 & 1 & S_y \\ S_x & S_y & S_x^2 + S_y^2 \end{bmatrix}. \quad (50.65)$$

The corresponding small slope neutral diffusive flux is

$$\mathbf{J}_{\text{small}} = -\rho A [\nabla_{\gamma} + \hat{\mathbf{z}} (\mathbf{S} \cdot \nabla_{\gamma})] C \quad (50.66)$$

where

$$\nabla_{\gamma} = \nabla_z + \mathbf{S} \partial_z \quad (50.67)$$

is the horizontal derivative operator computed on the neutral tangent plane (see equation (43.64)). To show that $\mathbf{J}_{\text{small}} \cdot \hat{\gamma} = 0$, we make use of the identity

$$\hat{\gamma} = \frac{\mathbf{S} - \hat{\mathbf{z}}}{(1 + \mathbf{S} \cdot \mathbf{S})^{1/2}}, \quad (50.68)$$

so that

$$\mathbf{J}_{\text{small}} \cdot \hat{\gamma} = \frac{\mathbf{J}_{\text{small}} \cdot \mathbf{S} - \mathbf{J}_{\text{small}} \cdot \hat{\mathbf{z}}}{(1 + \mathbf{S} \cdot \mathbf{S})^{1/2}} = 0. \quad (50.69)$$

Furthermore, we confirm that $\mathbf{J}_{\text{small}}$ is oriented down the tracer gradient by noting that

$$\mathbf{J}_{\text{small}} \cdot \nabla C = -\rho A [\nabla_{\gamma} C \cdot \nabla_z C + (\mathbf{S} \cdot \nabla_{\gamma} C) \partial_z C] \quad (50.70a)$$

$$= -\rho A [|\nabla_z C|^2 + 2(\mathbf{S} \cdot \nabla_z C) \partial_z C + |\mathbf{S} \partial_z C|^2] \quad (50.70b)$$

$$= -\rho A |\nabla_z C + \mathbf{S} \partial_z C|^2 \quad (50.70c)$$

$$= -\rho A |\nabla_{\gamma} C|^2 \quad (50.70d)$$

$$\leq 0. \quad (50.70e)$$

The small slope approximation was proposed by [Cox \(1987\)](#). However, his form for the small slope neutral diffusion flux was incorrect as it did not satisfy $\mathbf{J}_{\text{small}} \cdot \hat{\gamma} = 0$. The corrected form given by equation (50.66) was first written by [Gent and McWilliams \(1990\)](#). The resulting small slope neutral diffusion operator is commonly used in ocean climate models ([Griffies et al., 1998](#); [Lemarié et al., 2012](#)), which results from computing the three-dimensional convergence

$$\mathcal{R}_{\text{small}} = -\nabla \cdot \mathbf{J}_{\text{small}} = \nabla_z \cdot (\rho A \nabla_\gamma C) + \partial_z (\rho A \mathbf{S} \cdot \nabla_\gamma C). \quad (50.71)$$

50.3.3 Neutral tangent plane neutral diffusion

A third method to compute neutral diffusion is motivated by the form of isopycnal diffusion in isopycnal layered models. Rather than isopycnal layers, we work with layers determined locally by neutral tangent planes. The neutral tangent frame makes use of projected non-orthogonal generalized vertical coordinates detailed in Chapter 43.

Following the derivations given in Section 43.14, the neutral diffusive flux in the neutral tangent frame is given by the horizontal flux

$$\mathbf{J}_{\text{ntp}} = -\rho A \nabla_\gamma C. \quad (50.72)$$

This flux is oriented down the tracer gradient as oriented along neutral directions

$$\mathbf{J}_{\text{ntp}} \cdot \nabla_\gamma C = -\rho A |\nabla_\gamma C|^2, \quad (50.73)$$

which is the same as equation (50.70d) for the small slope fluxes. However, as a purely horizontal flux, \mathbf{J}_{ntp} is not oriented along neutral directions

$$\mathbf{J}_{\text{ntp}} \cdot \hat{\gamma} \neq 0. \quad (50.74)$$

Nevertheless, rather than computing the neutral diffusion operator as a horizontal convergence of this flux, the neutral tangent plane diffusion operator is computed by taking the convergence of \mathbf{J}_{ntp} along the neutral tangent plane as per equation (43.83)

$$\mathcal{R}_{\text{ntp}} = -\frac{1}{h^\gamma} [\nabla_\gamma \cdot (h^\gamma \mathbf{J}_{\text{ntp}})] = \frac{1}{h^\gamma} [\nabla_\gamma \cdot (h^\gamma \rho A \nabla_\gamma C)], \quad (50.75)$$

where

$$h^\gamma = \frac{\partial z}{\partial \gamma} d\gamma = - \left[\frac{g}{\rho_0 N^2} \right] d\gamma \quad (50.76)$$

measures the thickness of a layer defined by two neutral tangent planes (see equation (43.80)).

As detailed in Section 43.14, \mathcal{R}_{ntp} is identical to the small slope neutral diffusion operator (50.71)

$$\mathcal{R}_{\text{ntp}} = \mathcal{R}_{\text{small}}. \quad (50.77)$$

In principle, it is a matter of convenience which form of the operator one uses. However, there are certain issues to consider when implementing these operators in a numerical model. Notably, a discrete realization of \mathcal{R}_{ntp} allows for a diagonal downgradient implementation of neutral diffusion, just as isopycnal diffusion in an isopycnal ocean model. In contrast, a discrete realization of either $\mathcal{R}_{\text{redi}}$ or $\mathcal{R}_{\text{small}}$ cannot guarantee downgradient fluxes due to the off-diagonal nature of its neutral diffusive flux components ([Griffies et al. \(1998\)](#), [Beckers et al. \(1998\)](#), [Gnanadesikan \(1999\)](#), [Beckers et al. \(2000\)](#) [Lemarié et al. \(2012\)](#)). As a result, discrete realizations of $\mathcal{R}_{\text{redi}}$ or $\mathcal{R}_{\text{small}}$ can produce extrema, which are distinctly not properties of diffusion in the continuum (see Exercise 48.3). Hence, even though the continuum identity holds $\mathcal{R}_{\text{ntp}} = \mathcal{R}_{\text{small}}$, there are important differences that arise upon realizing these operators on a discrete lattice.

50.3.4 Neutrality condition

Given the expression (50.57) for the dianeutral unit vector, $\hat{\gamma}$, it is straightforward to show that the neutral diffusive flux for Conservative Temperature balances that for salinity

$$\alpha \mathbf{J}(\Theta) = \beta \mathbf{J}(S). \quad (50.78)$$

We refer to this balance as the *neutrality condition*. It reflects the vanishing of the neutral diffusive flux when acting on locally referenced potential density. It is maintained by the diffusive flux (50.58) of Redi (1982), the small slope flux (50.66) of Gent and McWilliams (1990), and the neutral tangent frame neutral diffusive flux (50.72). However, it is not maintained by the small slope fluxes from Cox (1987). Indeed, Griffies et al. (1998) argued for the importance of maintaining this balance to avoid a nonlinear instability plaguing certain numerical realizations of neutral diffusion such as that from Cox (1987).

50.3.5 Symmetry condition

Since the neutral diffusion tensor is symmetric (as are all diffusion tensors; see Section 48.4), we have

$$\mathbf{J}(\Theta) \cdot \nabla S = -A \rho \mathbb{K}^{mn} \partial_n \Theta \partial_m S \quad (50.79a)$$

$$= -A \rho \mathbb{K}^{nm} \partial_n S \partial_m \Theta \quad (50.79b)$$

$$= -A \rho \mathbb{K}^{nm} \partial_n S \partial_m \Theta \quad (50.79c)$$

$$= \mathbf{J}(S) \cdot \nabla \Theta. \quad (50.79d)$$

This symmetry condition holds for both the Redi neutral diffusion tensor and its small slope limit. It will be useful in our discussion of cabbeling and thermobaricity in Section 50.4.

50.3.6 GM skewson plus small slope neutral diffusion

A parameterization of mesoscale eddy stirring and mixing often appears in geopotential coordinate ocean models in the form of GM skewson (Section 50.2.1) and small slope neutral diffusion (Section 50.3.2). The combined tracer flux takes the form

$$\mathbf{F} = -A \nabla_z C + (\kappa - A) \mathbf{S} \partial_z C - \hat{\mathbf{z}} [(A + \kappa) \mathbf{S} \cdot \nabla_z C + A S^2 \partial_z C], \quad (50.80)$$

which can be written in terms of a subgrid scale transport tensor (Griffies, 1998)

$$\begin{bmatrix} F^{(x)} \\ F^{(y)} \\ F^{(z)} \end{bmatrix} = \begin{bmatrix} A & 0 & (A - \kappa) S_x \\ 0 & A & (A - \kappa) S_y \\ (A + \kappa) S_x & (A + \kappa) S_y & A S^2 \end{bmatrix} \begin{bmatrix} \partial_x C \\ \partial_y C \\ \partial_z C \end{bmatrix}. \quad (50.81)$$

In the 1990s and throughout much of the 2000s, it was common to assume that $A = \kappa$, in which case the combined subgrid scale flux simplifies to

$$\mathbf{F} = -\kappa \nabla_z C - \hat{\mathbf{z}} \kappa (2 \mathbf{S} \cdot \nabla_z C + S^2 C_{,z}). \quad (50.82)$$

Notably, the 2×2 horizontal mixing tensor is diagonal. Hence, the horizontal tracer flux is the same as that which arises from downgradient horizontal tracer diffusion. The simplicity of the horizontal flux component was compelling and alluring to modelers. It was furthermore argued by Dukowicz and Smith (1997) to be a fundamental property of mesoscale turbulence. However, as

emphasized through the works of [Treguier et al. \(1997\)](#), [Ferrari et al. \(2008\)](#), and [Ferrari et al. \(2010\)](#), the boundary conditions for neutral diffusion and GM skewness are distinct, thus breaking their symmetry. Furthermore, studies such as [Smith and Marshall \(2009\)](#) and [Abernathy et al. \(2013\)](#) clearly point to the distinct vertical structure for the two diffusivities. Such distinctions are expected since the skew diffusivity and neutral diffusivity parameterize physically distinct processes: one parameterizes the quasi-Stokes transport and the other parameterizes downgradient diffusion along neutral directions.

50.3.7 Small slope neutral diffusion in generalized vertical coordinates

Thus far we have considered neutral diffusion as realized in geopotential coordinates or using neutral tangent plane coordinates. Here, we detail the steps needed to realize neutral diffusion using the generalized vertical coordinates (GVCs) detailed in Chapters 43 and 44. This formulation is relevant for the now common use of generalized vertical coordinates for ocean modeling as reviewed by [Griffies et al. \(2020\)](#).

We start by recalling the expression (43.82) for a general diffusion operator written in terms of the generalized vertical coordinate, $\sigma = \sigma(x, y, z, t)$

$$\mathcal{R} = -\frac{1}{h^\sigma} \left[\nabla_\sigma \cdot (h^\sigma \mathbf{J}^h) + \delta_\sigma(z_\sigma \nabla_\sigma \cdot \mathbf{J}) \right], \quad (50.83)$$

where $\delta_\sigma \equiv d\sigma / \partial_\sigma$ is the dimensionless derivative operator, and the thickness of a σ -layer is

$$h^\sigma = dz = z_\sigma d\sigma = \frac{\partial z}{\partial \sigma} d\sigma. \quad (50.84)$$

Now assume the flux, \mathbf{J} , is given by equation (50.66) for small slope neutral diffusion. Transforming to GVCs leads to the horizontal flux component

$$\mathbf{J}_{\text{small}}^h = -\rho A \nabla_\gamma C \quad (50.85a)$$

$$= -\rho A [\nabla_z + (\nabla_\gamma z) \partial_z] C \quad (50.85b)$$

$$= -\rho A [\nabla_\sigma + (-\nabla_\sigma z + \nabla_\gamma z) \partial_z] C \quad (50.85c)$$

$$= -\rho A [\nabla_\sigma + (-\mathbf{S}^{(\sigma/z)} + \mathbf{S}^{(\gamma/z)}) \partial_z] C \quad (50.85d)$$

$$= -\rho A (\nabla_\sigma + \mathbf{S}^{(\gamma/\sigma)} \partial_z) C, \quad (50.85e)$$

where the neutral slopes as shown in Figure 50.2 satisfy the identity

$$\mathbf{S}^{(\sigma/z)} = \mathbf{S}^{(\gamma/z)} - \mathbf{S}^{(\gamma/\sigma)}. \quad (50.86)$$

Furthermore, we made use of the identity (43.64) relating the partial derivative operators

$$\nabla_\gamma = \nabla_z + (\nabla_\gamma z) \partial_z \quad \nabla_z = \nabla_\sigma - (\nabla_\sigma z) \partial_z. \quad (50.87)$$

The horizontal flux (50.85e) has the same form as when written using geopotential coordinates, only now with the derivative operator ∇_σ and the slope $\mathbf{S}^{(\gamma/\sigma)}$. Correspondingly, the vertical flux component

$$J_{\text{small}}^z = \mathbf{J}_{\text{small}}^h \cdot \mathbf{S}^{(\gamma/z)} \quad (50.88)$$

takes the form

$$z_\sigma \nabla_\sigma \cdot \mathbf{J}_{\text{small}} = -\mathbf{S}^{(\sigma/z)} \cdot \mathbf{J}_{\text{small}}^h + J_{\text{small}}^z = \mathbf{J}_{\text{small}}^h \cdot \mathbf{S}^{(\gamma/\sigma)}, \quad (50.89)$$

which in turn yields the diffusion operator (50.83)

$$\mathcal{R} = -\frac{1}{h^\sigma} \left[\nabla_\sigma \cdot (h^\sigma \mathbf{J}_{\text{small}}^h) + \delta_\sigma (\mathbf{J}_{\text{small}}^h \cdot \mathbf{S}^{(\gamma/\sigma)}) \right]. \quad (50.90)$$

In the special case when σ is parallel to the neutral direction so that $\mathbf{S}^{(\gamma/\sigma)} = 0$, the diffusion operator (50.90) reduces to the neutral tangent plane version given by equation (50.75).

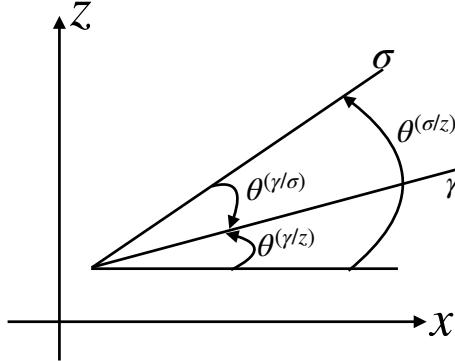


FIGURE 50.2: Slopes of neutral directions (denoted by γ isolines) relative to both the horizontal plane, $\tan \theta^{(\gamma/z)}$, and relative to a σ -isoline, $\tan \theta^{(\gamma/\sigma)}$. We assume positive angles as measure counter-clockwise relative to the horizontal and relative to the σ -isoline, respectively. Hence, for this example, $\theta^{(\gamma/z)} > 0$ yet $\theta^{(\gamma/\sigma)} < 0$. When extending to the two horizontal directions, the slopes generally satisfy $\mathbf{S}^{(\sigma/z)} = \mathbf{S}^{(\gamma/z)} - \mathbf{S}^{(\gamma/\sigma)}$, where $|\mathbf{S}^{(\gamma/z)}| = |\tan \theta^{(\gamma/z)}|$ and $|\mathbf{S}^{(\gamma/\sigma)}| = |\tan \theta^{(\gamma/\sigma)}|$. Note that this relation between slope vectors also holds for arbitrary orientations of the σ and γ isolines.

50.3.8 Comments

As noted in Section 49.5.3, there is presently no parameterization of subgrid scale stirring along neutral directions arising from the correlations between tracer and velocity fluctuations. Rather, the only parameterized subgrid scale stirring is associated with quasi-Stokes transport, with [Gent et al. \(1995\)](#) providing the canonical approach. To parameterize the skew fluxes arising from tracer-velocity correlations requires one to study the polarization of the eddies giving rise to the skew flux, as per the discussion in Section 49.3.2 and [Middleton and Loder \(1989\)](#).

50.4 Cabbeling and thermobaricity

We now return to the density equation (51.10)

$$\frac{D \ln \rho}{Dt} = \nabla \cdot [\nu_\Theta \mathbf{J}(\Theta) + \nu_S \mathbf{J}(S)] - (\mathbf{J}(\Theta) \cdot \nabla \nu_\Theta + \mathbf{J}(S) \cdot \nabla \nu_S) + \frac{\omega}{\rho c^2}. \quad (50.91)$$

We focus here on temperature and salinity fluxes just from neutral diffusion. The neutrality condition (50.78) takes the following form in terms of specific volume

$$\nu_\Theta \mathbf{J}(\Theta) + \nu_S \mathbf{J}(S) = 0. \quad (50.92)$$

Consequently, neutral diffusion affects density evolution only through the source term

$$\left[\frac{D \ln \rho}{Dt} \right]_{\text{ntral diff}} = -\mathbf{J}(\Theta) \cdot \nabla \nu_\Theta - \mathbf{J}(S) \cdot \nabla \nu_S. \quad (50.93)$$

50.4.1 Basic manipulations

As a first step, eliminate the salt flux by using the neutrality condition (50.92) so that

$$\mathbf{J}(\Theta) \cdot \nabla \nu_\Theta + \mathbf{J}(S) \cdot \nabla \nu_S = \mathbf{J}(\Theta) \cdot [\nu_S \nabla \nu_\Theta - \nu_\Theta \nabla \nu_S] / \nu_S. \quad (50.94)$$

Next, expand the gradients of the specific volume to write

$$\nabla \nu_\Theta = \nu_{\Theta\Theta} \nabla \Theta + \nu_{\Theta S} \nabla S + \nu_{\Theta p} \nabla p \quad (50.95a)$$

$$\nabla \nu_S = \nu_{SS} \nabla S + \nu_{\Theta S} \nabla \Theta + \nu_{Sp} \nabla p, \quad (50.95b)$$

so that

$$\begin{aligned} \nu_S \nabla \nu_\Theta - \nu_\Theta \nabla \nu_S &= \nabla \Theta (\nu_S \nu_{\Theta\Theta} - \nu_\Theta \nu_{\Theta S}) \\ &\quad + \nabla S (\nu_S \nu_{\Theta S} - \nu_\Theta \nu_{SS}) + \nabla p (\nu_S \nu_{\Theta p} - \nu_\Theta \nu_{Sp}). \end{aligned} \quad (50.96)$$

We again make use of the neutrality condition (50.92), as well as the symmetry condition (50.79d) to write

$$\mathbf{J}(\Theta) \cdot \nabla S (\nu_S \nu_{\Theta S} - \nu_\Theta \nu_{SS}) = -\mathbf{J}(\Theta) \cdot \nabla \Theta \left(\nu_\Theta \nu_{\Theta S} - \nu_{SS} \frac{(\nu_\Theta)^2}{\nu_S} \right). \quad (50.97)$$

Bringing these results together leads to

$$\begin{aligned} \mathbf{J}(\Theta) \cdot \nabla \nu_\Theta + \mathbf{J}(S) \cdot \nabla \nu_S &= \mathbf{J}(\Theta) \cdot \nabla p \left[\nu_{\Theta p} - \nu_{pS} \frac{\nu_\Theta}{\nu_S} \right] \\ &\quad + \mathbf{J}(\Theta) \cdot \nabla \Theta \left[\nu_{\Theta\Theta} - 2 \nu_{\Theta S} \frac{\nu_\Theta}{\nu_S} + \nu_{SS} \left(\frac{\nu_\Theta}{\nu_S} \right)^2 \right], \end{aligned} \quad (50.98)$$

which can be written in terms of density partial derivatives as

$$\begin{aligned} \mathbf{J}(\Theta) \cdot \nabla \nu_\Theta + \mathbf{J}(S) \cdot \nabla \nu_S &= -\rho^{-2} \mathbf{J}(\Theta) \cdot \nabla p \left[\rho_{\Theta p} - \rho_{pS} \frac{\rho_\Theta}{\rho_S} \right] \\ &\quad - \rho^{-2} \mathbf{J}(\Theta) \cdot \nabla \Theta \left[\rho_{\Theta\Theta} - 2 \rho_{\Theta S} \frac{\rho_\Theta}{\rho_S} + \rho_{SS} \left(\frac{\rho_\Theta}{\rho_S} \right)^2 \right]. \end{aligned} \quad (50.99)$$

50.4.2 A tidy form

We next write the bracket terms in forms consistent with those introduced by [McDougall \(1987b\)](#) in his classic paper discussing cabbeling and thermobaricity. For that purpose, introduce the *thermobaricity* parameter (dimensions of inverse temperature times inverse pressure) whose form is given by

$$\mathcal{T} = \beta \partial_p \left[\frac{\alpha}{\beta} \right] \quad (50.100a)$$

$$= \frac{\partial \alpha}{\partial p} - \frac{\alpha}{\beta} \frac{\partial \beta}{\partial p} \quad (50.100b)$$

$$= \rho \nu_S \partial_p \left[\frac{\nu_\Theta}{\nu_S} \right] \quad (50.100c)$$

$$= -\rho^{-1} \rho_S \partial_p \left[\frac{\rho_\Theta}{\rho_S} \right] \quad (50.100d)$$

$$= -\rho^{-1} \left[\rho_{\Theta p} - \rho_{pS} \left[\frac{\rho_\Theta}{\rho_S} \right] \right], \quad (50.100e)$$

and the *cabbeling* parameter (dimensions of squared inverse temperature)

$$\mathcal{C} = \frac{\partial\alpha}{\partial\Theta} + 2\frac{\alpha}{\beta}\frac{\partial\alpha}{\partial S} - \left(\frac{\alpha}{\beta}\right)^2\frac{\partial\beta}{\partial S} \quad (50.101a)$$

$$= -\rho^{-1} \left[\rho_{\Theta\Theta} - 2\rho_{\Theta S} \left[\frac{\rho_\Theta}{\rho_S} \right] + \rho_{SS} \left[\frac{\rho_\Theta}{\rho_S} \right]^2 \right] \quad (50.101b)$$

$$= \rho \left[\nu_{\Theta\Theta} - 2\nu_{\Theta S} \left[\frac{\nu_\Theta}{\nu_S} \right] + \nu_{SS} \left[\frac{\nu_\Theta}{\nu_S} \right]^2 \right] \quad (50.101c)$$

to render the very compact result

$$\mathbf{J}(\Theta) \cdot \nabla \nu_\Theta + \mathbf{J}(S) \cdot \nabla \nu_S = \rho^{-1} \mathbf{J}(\Theta) \cdot (\mathcal{T} \nabla p + \mathcal{C} \nabla \Theta) \quad (50.102)$$

which in turn yields the material evolution of *in situ* density due to neutral diffusion

$$\left[\frac{D\rho}{Dt} \right]_{\text{ntral diff}} = -\mathbf{J}(\Theta) \cdot (\mathcal{T} \nabla p + \mathcal{C} \nabla \Theta). \quad (50.103)$$

50.4.3 Cabbeling

Consider the mixing of two seawater elements. Let the fluid elements separately have distinct Conservative Temperature and/or salinity, but equal locally referenced potential density. For a linear equation of state, whereby density is a linear function of Θ and S , then the resulting mixed fluid element has the same density as the unmixed separate elements. However, for a nonlinear equation of state, the mixed element generally has a different density. Furthermore, a property of seawater is that the density of the mixed element is actually greater than the unmixed elements. This densification upon mixing is a physical process known as *cabbeling* ([McDougall, 1987b](#)).

The sign definite nature of cabbeling (i.e., cabbeling always results in denser fluid elements after mixing) is a direct result of the geometry of the locally referenced potential density surface when viewed in Conservative Temperature and salinity space. This property in turn manifests with the following inequality for the cabbeling parameter

$$\mathcal{C} = \frac{\partial\alpha}{\partial\Theta} + 2\frac{\alpha}{\beta}\frac{\partial\alpha}{\partial S} - \left[\frac{\alpha}{\beta}\right]^2\frac{\partial\beta}{\partial S} \geq 0. \quad (50.104)$$

Given the downgradient nature of the neutral diffusive fluxes, we have

$$\text{Cabbeling} \equiv -\mathcal{C} \mathbf{J}(\Theta) \cdot \nabla \Theta \geq 0, \quad (50.105)$$

thus providing a mathematical expression for the cabbeling source (with dimensions of density per time). That is, cabbeling results in a positive material evolution of density; i.e., density increases due to cabbeling. An increase in the density within a column of fluid results in the reduction of the sea level due to compression of the column.

50.4.4 Thermobaricity

The thermobaricity parameter

$$\mathcal{T} = \beta \frac{\partial}{\partial p} \left[\frac{\alpha}{\beta} \right] \quad (50.106)$$

is nonzero due to pressure dependence of the ratio of the thermal expansion coefficient to the haline contraction coefficient. As both thermal and haline effects are present, the parameter \mathcal{T} is more precisely split into two terms

$$\begin{aligned}\mathcal{T} &= \frac{\partial\alpha}{\partial p} - \frac{\alpha}{\beta} \frac{\partial\beta}{\partial p} \\ &= -\frac{\rho_{\Theta p}}{\rho} + \frac{\rho_\Theta}{\rho_S} \frac{\rho_{pS}}{\rho}\end{aligned}\tag{50.107}$$

Thermobaricity is the common name for the sum, since pressure variations in the thermal expansion coefficient dominate those of the haline contraction coefficient. The thermal expansion coefficient generally increases as pressure increases, thus making the thermobaric parameter positive.

Since neutral gradient of temperature need not be oriented in a special manner relative to the neutral gradient of pressure, there is no sign-definite nature to the thermobaricity source term (with units of density per time)

$$\text{Thermobaricity} \equiv -\mathcal{T} \mathbf{J}(\Theta) \cdot \nabla p\tag{50.108}$$

appearing in equation (50.102). Thus, thermobaricity can either increase or decrease density, depending on details of the density and fluxes. However, as noted by [McDougall and You \(1990\)](#), thermobaricity typically increases density in much of the World Ocean.

50.4.5 Comments

[Griffies and Greatbatch \(2012\)](#) discuss the impacts on global mean sea level from thermobaricity and cabbeling as diagnosed from an ocean model. Given that cabbeling always densifies and thermobaricity is also dominated by densification, these processes lead to a general reduction in global mean sea level. [Klocker and McDougall \(2010\)](#), [Groeskamp et al. \(2016\)](#), and [Groeskamp et al. \(2019\)](#) diagnose cabbeling and thermobaricity from observational based measurements, with [Groeskamp et al. \(2019\)](#) also offering a more robust numerical method for performing that diagnostic calculation.

Although cabbeling and thermobaricity lead to watermass transformation and associated transport of water across neutral directions, they are distinct from other mixing processes such as breaking gravity waves (Section 50.1). Namely, cabbeling and thermobaricity arise from the stirring by mesoscale eddies along neutral directions, which in turn leads to neutral diffusion acting on Conservative Temperature and salinity. Transient mesoscale eddies impart a downscale cascade of tracer variance that is ultimately halted by irreversible molecular mixing, or microscale processes active before reaching the molecular level. This mixing is the ultimate cause for cabbeling and thermobaricity, with the overall strength of the cabbeling and thermobaricity determined by the strength of the mesoscale stirring.



Ocean buoyancy and sea level

Conservative temperature, Θ , is the preferred means to measure the transport of heat in the ocean, and salinity, S , measures the concentration of dissolved salt matter. These two scalar fields are referred to as *active* tracers as they both impact buoyancy and in turn affect pressure and ocean currents. In this chapter we study how the evolution of Θ and S affects buoyancy. Θ and S are conservative tracers so that the net changes in potential enthalpy and salt over the global ocean domain arise from net imbalances in the enthalpy and salt boundary fluxes. Likewise, ocean mass is a conserved field, with global mass changes arising from imbalances in boundary mass fluxes such as those occurring from increases in land ice melt. However, ocean volume, and hence ocean density and buoyancy, are not conserved fields. These points have direct impact on how global mean sea level is affected by ocean processes such as mixing, with rudiments presented in this chapter.

READER'S GUIDE FOR THIS CHAPTER

Basic notions of thermodynamics, such as Section 22.8, motivate the use of Conservative temperature (or potential temperature) as a measure of ocean heat transfer, rather than *in situ* temperature. The material in this chapter considers how we measure air-sea buoyancy fluxes and analyze changes to global mean sea level. Both of these topics are increasingly important in our anthropogenically warming world.

51.1	Material evolution of <i>in situ</i> density	1166
51.1.1	Material changes to pressure	1166
51.1.2	Material changes to Θ and S	1167
51.1.3	Summary of density changes	1167
51.1.4	A synopsis of ocean mixing processes	1168
51.2	Salt and freshwater budgets	1168
51.2.1	Salt and freshwater	1168
51.2.2	Mass budgets	1169
51.3	Surface boundary conditions for S and Θ	1170
51.3.1	Salt and freshwater	1170
51.3.2	The non-advectional salt flux boundary condition	1171
51.3.3	Conservative Temperature boundary condition	1172
51.3.4	Comments and further reading	1173
51.4	Surface boundary fluxes of buoyancy	1173
51.4.1	Outlining the boundary fluxes of enthalpy and salt	1174
51.4.2	Evolution from surface boundary fluxes	1174
51.4.3	Buoyancy tendency from surface boundary fluxes	1175
51.4.4	Comments	1176

51.5 Global mean sea level	1176
51.5.1 Definitions and assumptions	1176
51.5.2 Budget for global mean sea level	1177
51.5.3 Changes due to mass input	1178
51.5.4 Steric changes due to changes in density	1178
51.5.5 Enthalpy flux imbalances giving rise to thermosteric sea level	1179
51.5.6 Global sea level in a Boussinesq fluid	1179
51.5.7 Global halosteric sea level changes are negligible	1180
51.5.8 Further study	1180

51.1 Material evolution of *in situ* density

Changes to the *in situ* density of seawater affects pressure forces in the ocean as well as the volume occupied by the ocean fluid (i.e., sea level). As discussed in Section 25.3.2, it is common to write the seawater equation of state for density as a function of potential temperature, salinity, and pressure. A somewhat more accurate approach makes use of the Conservative Temperature rather than the potential temperature, where the Conservative Temperature, Θ , is the potential enthalpy divided by a constant heat capacity (McDougall, 2003; IOC et al., 2010). We thus make use of the empirical relation for the seawater density in the functional form

$$\rho = \rho(\Theta, S, p), \quad (51.1)$$

where S is the salinity rather than the salt concentration ($S = 1000 \text{S}$).

We formulate the material evolution of density as weighted by the specific volume

$$\nu = \rho^{-1}, \quad (51.2)$$

so that we consider

$$\frac{D \ln \rho}{Dt} = \frac{\partial \ln \rho}{\partial \Theta} \frac{D\Theta}{Dt} + \frac{\partial \ln \rho}{\partial S} \frac{DS}{Dt} + \frac{\partial \ln \rho}{\partial p} \frac{Dp}{Dt} \quad (51.3a)$$

$$= -\alpha_\Theta \frac{D\Theta}{Dt} + \beta_s \frac{DS}{Dt} + \frac{\dot{p}}{\rho c_s^2}. \quad (51.3b)$$

In this equation we introduced the thermal expansion coefficient, the haline contraction coefficient, the squared speed of sound, and the vertical pseudo-velocity in pressure coordinates

$$\alpha_\Theta = - \left[\frac{\partial \ln \rho}{\partial \Theta} \right]_{p,S} \quad \beta_s = \left[\frac{\partial \ln \rho}{\partial S} \right]_{p,\Theta} \quad c_s^2 = \left[\frac{\partial p}{\partial \rho} \right]_{S,\Theta} \quad \dot{p} = \frac{Dp}{Dt}. \quad (51.4)$$

51.1.1 Material changes to pressure

To garner some exposure to the physics of \dot{p} as it appears in equation (51.3), we consider the special case of a hydrostatic fluid, where the volume per time per horizontal area of fluid crossing a surface of constant hydrostatic pressure is given by (see Section 44.4.6)

$$w^{(p)} = \frac{\partial z}{\partial p} \frac{Dp}{Dt} = -(\rho g)^{-1} \dot{p}. \quad (51.5)$$

The transport measured by $w^{(p)}$ is the pressure-coordinate analog of the vertical velocity component $w = Dz/Dt$ in a geopotential coordinate representation of the vertical. That is, fluid moving into

regions of increasing hydrostatic pressure ($\dot{p} > 0$) represents downward movement of fluid, with $w^{(p)} < 0$ in this case. Conversely, motion into decreasing hydrostatic pressure represents upward motion, with $w^{(p)} > 0$. This vertical movement generally occurs in the presence of waves, currents, and mixing; i.e., both reversible and irreversible processes give rise to vertical motion.

51.1.2 Material changes to Θ and S

We now focus on the salinity and temperature contributions to the evolution of *in situ* density. To do so, assume that the material evolution of Θ and S are given by the convergence of a subgrid scale flux

$$\rho \frac{D\Theta}{Dt} = -\nabla \cdot \mathbf{J}^{(\Theta)} \quad (51.6a)$$

$$\rho \frac{DS}{Dt} = -\nabla \cdot \mathbf{J}^{(S)}. \quad (51.6b)$$

This form for material changes in temperature and salinity then lead to

$$-\alpha_\Theta \frac{D\Theta}{Dt} + \beta_S \frac{DS}{Dt} = \nu_\Theta \nabla \cdot \mathbf{J}^{(\Theta)} + \nu_S \nabla \cdot \mathbf{J}^{(S)} \quad (51.7a)$$

$$= \nabla \cdot [\nu_\Theta \mathbf{J}^{(\Theta)} + \nu_S \mathbf{J}^{(S)}] - [\mathbf{J}^{(\Theta)} \cdot \nabla \nu_\Theta + \mathbf{J}^{(S)} \cdot \nabla \nu_S] \quad (51.7b)$$

where again $\nu = \rho^{-1}$ is the specific volume and its partial derivatives are

$$\nu_\Theta = \frac{\partial \nu}{\partial \Theta} = \frac{\alpha_\Theta}{\rho} \quad \text{and} \quad \nu_S = \frac{\partial \nu}{\partial S} = -\frac{\beta_S}{\rho}. \quad (51.8)$$

51.1.3 Summary of density changes

Bringing the above results together leads to the density equation

$$\frac{D \ln \rho}{Dt} - \frac{\dot{p}}{\rho c_s^2} = \nabla \cdot [\nu_\Theta \mathbf{J}^{(\Theta)} + \nu_S \mathbf{J}^{(S)}] - [\mathbf{J}^{(\Theta)} \cdot \nabla \nu_\Theta + \mathbf{J}^{(S)} \cdot \nabla \nu_S], \quad (51.9)$$

which has the equivalent form

$$\frac{D\rho}{Dt} - \frac{\dot{p}}{c_s^2} = \nabla \cdot [\alpha_\Theta \mathbf{J}^{(\Theta)} - \beta_S \mathbf{J}^{(S)}] - [\mathbf{J}^{(\Theta)} \cdot \nabla \alpha_\Theta - \mathbf{J}^{(S)} \cdot \nabla \beta_S]. \quad (51.10)$$

We brought the adiabatic source term from motion across pressure surfaces (Section 51.1.1) onto the left hand side, as this term appears in the absence of mixing whereas terms on the right hand side require mixing. The first term on the right hand side represents the divergence of a buoyancy flux due to subgrid scale fluxes of Conservative Temperature and salinity. In turn, density increases in regions where the buoyancy flux diverges (e.g., temperature Θ reducing and S increasing). These fluxes arise from a variety of mixing processes, some of which are surveyed in Section 51.1.4. The second term on the right hand side of equations (51.9) and (51.10) relates to properties of the locally referenced potential density surface. We study this source term arising from neutral diffusion in Section 50.4, where we encounter cabbeling and thermobaricity. Further effects arise from unresolved eddy-induced stirring, with that process contributing to the material time derivative operator to render a residual mean velocity (Section 50.1.4).

51.1.4 A synopsis of ocean mixing processes

Irreversible mixing in the ocean takes place at the millimeter scale through the process of molecular (Brownian) motion acting to dissipate property gradients. This mixing is generally represented by downgradient molecular diffusion ([Einstein, 1905](#)). The molecular diffusivity of matter (e.g., salt) in seawater is roughly $10^{-9} \text{ m}^2 \text{ s}^{-1}$, whereas the molecular thermal diffusivity is roughly 100 times larger (it is easier to diffuse heat than matter, [Gill, 1982](#)). Reversible stirring by turbulent eddies greatly increases the magnitude of property gradients upon which molecular diffusion acts ([Eckart, 1948](#); [Nakamura, 2001](#); [Müller and Garrett, 2002](#)), thereby increasing the total amount of irreversible mixing. Motivated by molecular diffusion, and following the pioneering work of [Taylor \(1921\)](#), it is common to parameterize mixing induced by eddy stirring as a diffusive closure with an eddy diffusivity that is far larger than molecular values. Furthermore, the eddy diffusivities are generally the same for all tracers since eddies generally act the same regardless the tracer. Double diffusive processes is the notable counter-example to this equivalence [Schmitt \(1994\)](#).

Mixing induced by eddies of length scale $\mathcal{O}(\text{centimeters-metres})$ is associated with, among other processes, gravitational instability, shear instability and breaking internal gravity waves ([MacKinnon et al., 2013](#)), as well as a suite of boundary layer processes ([Large et al., 1994](#)). This mixing is commonly parameterized by a flow dependent isotropic eddy diffusivity. The magnitude of the eddy diffusivity is typically $\mathcal{O}(10^{-3} - 10^{-2} \text{ m}^2 \text{ s}^{-1})$ in boundary layers, and $\mathcal{O}(10^{-5} \text{ m}^2 \text{ s}^{-1})$ in the quiescent ocean interior ([Polzin et al., 1997](#); [Whalen et al., 2012](#); [Waterhouse et al., 2014](#)).

Mesoscale eddies, with size $\mathcal{O}(10 - 100) \text{ km}$, preferentially stir tracers along neutral directions ([McDougall, 1987a,b](#); [McDougall et al., 2014](#)). The mesoscale eddy stirring in turn induces a mixing that is parametrized by downgradient diffusion along neutral directions (Section 50.3). When feeling the geometric constraints of the surface boundary, mesoscale stirring leads to horizontal oriented mixing across outcropped density surfaces ([Treguier et al., 1997](#); [Ferrari et al., 2008](#)). This mixing is parameterized by downgradient horizontal diffusion. The neutral and horizontal eddy diffusivities associated with mesoscale processes are typically $\mathcal{O}(10^2 - 10^3 \text{ m}^2 \text{ s}^{-1})$ in the ocean interior and can rise to $\mathcal{O}(10^4 \text{ m}^2 \text{ s}^{-1})$ in the ocean surface layer ([Abernathay et al., 2013](#); [Klocker and Abernathay, 2014](#); [Cole et al., 2015](#)).

Although the isotropic diffusivity is much smaller than the mesoscale diffusivity, the isotropic diffusivity multiplies the generally larger tracer gradients crossing neutral directions and thus supports a critical form of watermass transformation and an induced ocean circulation ([Munk, 1966](#); [Munk and Wunsch, 1998](#)).

51.2 Salt and freshwater budgets

We here specialize the kinematics of material tracers given in Section 16.5 to the case of seawater, which we treat as a two component fluid comprised of salt and freshwater concentrations. We extend this discussion in Section 51.4 by studying the role of surface boundary salt, heat, and water transports on changes in ocean buoyancy.¹

51.2.1 Salt and freshwater

Seawater is comprised of two material tracers: freshwater along with a suite of dissolved trace “salts”. The ratio of salts is roughly constant over the World Ocean. We are thus able to make use

¹We provide a theoretical discussion of Conservative Temperature in Section 22.8. For present purposes, we merely need to know it is the ocean tracer best suited to measuring heat transport.

of a single effective mass concentration known as the *salt concentration*²

$$S = \frac{\text{mass of salt}}{\text{mass of seawater}} = \frac{\text{mass of salt}}{\text{mass of freshwater} + \text{mass of salt}} \quad (51.11)$$

to specify the amount of salt within an element of seawater. In practice oceanographers choose to work with the *salinity*,³

$$S = 1000 S, \quad (51.12)$$

which converts from typical salt concentrations of $S = 0.035$ to a salinity of $S = 35$. The complement to salt concentration is the freshwater concentration or mass fraction for an element of seawater

$$F = \frac{\text{mass of freshwater}}{\text{mass of seawater}} = \frac{\text{mass of freshwater}}{\text{mass of freshwater} + \text{mass of salt}} = 1 - S. \quad (51.13)$$

Other trace matter occurs at very low concentrations so as to make seawater, in effect, a two-component fluid consisting of freshwater plus dissolved salt.⁴ We here derive the mass budget for salt and freshwater as well as the associated kinematic boundary conditions.

51.2.2 Mass budgets

Following our discussion of the tracer equation in Section 16.1, the mass budget equations for an element of seawater take the form

$$\frac{\partial \rho}{\partial t} + \nabla \cdot (\rho \mathbf{v}) = 0 \quad \text{seawater} \quad (51.14)$$

$$\frac{\partial(\rho S)}{\partial t} + \nabla \cdot (\rho \mathbf{v} S + \mathbf{J}^{(S)}) = 0 \quad \text{salt} \quad (51.15)$$

$$\frac{\partial(\rho F)}{\partial t} + \nabla \cdot (\rho \mathbf{v} F + \mathbf{J}^{(F)}) = 0 \quad \text{freshwater.} \quad (51.16)$$

Equation (51.14) is the mass budget for seawater and equation (51.15) is the mass budget for salt. The freshwater budget (51.16) is derived by subtracting the salt budget (51.15) from the seawater mass budget (51.14). Hence, only two of the three mass budget equations (51.14)-(51.16) are independent.

We make use of the barycentric velocity in the above conservation laws, where the barycentric velocity for the ocean is given by

$$\mathbf{v} = S \mathbf{v}^{(S)} + F \mathbf{v}^{(F)}. \quad (51.17)$$

The velocities $\mathbf{v}^{(S)}$ and $\mathbf{v}^{(F)}$ are, respectively, the molecular center of mass velocities for salt and freshwater within a fluid element, in which case

$$\frac{\partial S}{\partial t} + \mathbf{v}^{(S)} \cdot \nabla S = 0 \quad \text{and} \quad \frac{\partial F}{\partial t} + \mathbf{v}^{(F)} \cdot \nabla F = 0. \quad (51.18)$$

Furthermore, the fluxes $\mathbf{J}^{(S)}$ and $\mathbf{J}^{(F)}$ arise from the difference between the salt and freshwater velocities from the barycentric velocity

$$\mathbf{J}^{(S)} = \rho S (\mathbf{v}^{(S)} - \mathbf{v}) \quad \text{and} \quad \mathbf{J}^{(F)} = \rho F (\mathbf{v}^{(F)} - \mathbf{v}). \quad (51.19)$$

²We use the salt concentration, S , in this section to avoid 1/1000 factors needed if working with salinity, $S = 1000 S$.

³More precisely, the salinity, S , as defined by equation (51.12) is the *Absolute Salinity*. Absolute Salinity is distinct from the *practical salinity* determined by conductivity measurements. [IOC et al. \(2010\)](#) provides a full accounting of the theory and practice of ocean salinity.

⁴See [IOC et al. \(2010\)](#) for more discussion of the variations of salt concentration ratios over the ocean, as well as the impacts from biogeochemical tracers.

These fluxes are often parameterized by downgradient diffusive fluxes

$$\mathbf{J}^{(S)} = -\rho \mathbb{K} \cdot \nabla S \quad \text{and} \quad \mathbf{J}^{(F)} = -\rho \mathbb{K} \cdot \nabla F, \quad (51.20)$$

where \mathbb{K} is the kinematic diffusivity tensor for salt in seawater, which is a positive semi-definite symmetric tensor. We use the same diffusivity tensor for salt and freshwater since the diffusion of one is balanced by the other. When concerned with molecular processes, the diffusivity tensor is isotropic with diffusivities set by the molecular value of $10^{-9} \text{ m}^2 \text{ s}^{-1}$. However, as discussed in Section 50.1, the eddy diffusivity is far larger than the molecular diffusivity in the presence of turbulent eddy processes, which also introduces anisotropies to the diffusion tensor.

The advective flux of seawater is comprised of a salt flux plus a freshwater flux

$$\rho \mathbf{v} = \rho S \mathbf{v}^{(S)} + \rho F \mathbf{v}^{(F)}. \quad (51.21)$$

Conversely, the salt flux and freshwater flux can be represented as a *non-advective flux* plus an advective flux where advection is defined by the barycentric velocity

$$\rho S \mathbf{v}^{(S)} = \rho S (\mathbf{v}^{(S)} - \mathbf{v}) + \rho S \mathbf{v} = \mathbf{J}^{(S)} + \rho S \mathbf{v} \quad (51.22a)$$

$$\rho F \mathbf{v}^{(F)} = \rho F (\mathbf{v}^{(F)} - \mathbf{v}) + \rho F \mathbf{v} = \mathbf{J}^{(F)} + \rho F \mathbf{v}. \quad (51.22b)$$

The non-advective fluxes, $\mathbf{J}^{(S)}$ and $\mathbf{J}^{(F)}$, lead to an exchange of mass with zero net movement of mass. In contrast, the advective flux moves mass as determined by the barycentric velocity. Furthermore, note that the center of mass velocities, $\mathbf{v}^{(S)}$ and $\mathbf{v}^{(F)}$, offer a conceptual framework of use to formulate the kinematic boundary conditions. Even so, they offer no new information beyond that contained in the fluxes $\mathbf{J}^{(S)}$ and $\mathbf{J}^{(F)}$.

51.3 Surface boundary conditions for S and Θ

In this section we summarize the surface boundary conditions holding for the salinity and Conservative Temperature equations. This treatment complements that given in Section 16.5.3.

51.3.1 Salt and freshwater

In deriving the boundary condition (16.70) in Section 15.4.3, we made use of the barycentric velocity \mathbf{v} for an element of seawater. We can garner further kinematic insights into the two-component ocean system by decomposing the total mass flux into contributions from salt and freshwater

$$\mathcal{Q}_m = \mathcal{Q}_S + \mathcal{Q}_F, \quad (51.23)$$

and by introducing the center of mass velocities for salt and freshwater according to

$$-\mathcal{Q}_m = \rho (\mathbf{v} - \mathbf{v}^{(\eta)}) \cdot \hat{\mathbf{n}} \quad (51.24a)$$

$$= \rho [S \mathbf{v}^{(S)} + F \mathbf{v}^{(F)} - \mathbf{v}^{(\eta)}] \cdot \hat{\mathbf{n}} \quad (51.24b)$$

$$= \rho [S (\mathbf{v}^{(S)} - \mathbf{v}^{(\eta)}) + \mathbf{v}^{(\eta)}] + F \mathbf{v}^{(F)} - \mathbf{v}^{(\eta)}] \cdot \hat{\mathbf{n}} \quad (51.24c)$$

$$= S \rho (\mathbf{v}^{(S)} - \mathbf{v}^{(\eta)}) \cdot \hat{\mathbf{n}} + (1 - S) \rho (\mathbf{v}^{(F)} - \mathbf{v}^{(\eta)}) \cdot \hat{\mathbf{n}} \quad (51.24d)$$

$$= S \rho (\mathbf{v}^{(S)} - \mathbf{v}^{(\eta)}) \cdot \hat{\mathbf{n}} + F \rho (\mathbf{v}^{(F)} - \mathbf{v}^{(\eta)}) \cdot \hat{\mathbf{n}} \quad (51.24e)$$

$$\equiv -(\mathcal{Q}_S + \mathcal{Q}_F), \quad (51.24f)$$

where we wrote

$$S \rho (\mathbf{v}^{(S)} - \mathbf{v}^{(\eta)}) \cdot \hat{\mathbf{n}} = -\mathcal{Q}_S \quad (51.25a)$$

$$F \rho (\mathbf{v}^{(F)} - \mathbf{v}^{(\eta)}) \cdot \hat{\mathbf{n}} = -\mathcal{Q}_F. \quad (51.25b)$$

In these equations, we introduced the velocity, $\mathbf{v}^{(\eta)}$, of a point fixed to the free surface. We only need the projection of this velocity in the outward normal direction, which is written by equation (15.77)

$$\mathbf{v}^{(\eta)} \cdot \hat{\mathbf{n}} = \frac{\partial \eta / \partial t}{|\nabla(z - \eta)|} = \frac{\partial \eta / \partial t}{\sqrt{1 + |\nabla \eta|^2}} \implies \mathbf{v}^{(\eta)} \cdot \hat{\mathbf{n}} d\mathcal{S} = \partial_t \eta dA, \quad (51.26)$$

where $d\mathcal{S}$ is the area element on the free surface and dA is its horizontal projection. Note that in many regions, the ocean surface is impermeable to salt, in which case the ocean surface acts as a material surface in terms of the salt velocity

$$\rho (\mathbf{v}^{(S)} - \mathbf{v}^{(\eta)}) \cdot \hat{\mathbf{n}} = 0 \quad \text{zero surface salt flux.} \quad (51.27)$$

The key exception to this boundary condition concerns sea ice, whereby salt is exchanged between liquid seawater and sea ice upon the melting or freezing of ice.

For most applications, it is preferable to make use of equation (51.22a) to eliminate the salt velocity $\mathbf{v}^{(S)}$ in favor of the non-advection flux $\mathbf{J}^{(S)} = \rho S (\mathbf{v}^{(S)} - \mathbf{v})$, in which case the kinematic boundary condition (51.25a) takes the form

$$-\mathcal{Q}_S = S \rho (\mathbf{v}^{(S)} - \mathbf{v}^{(\eta)}) \cdot \hat{\mathbf{n}} = S \rho (\mathbf{v}^{(S)} - \mathbf{v} + \mathbf{v} - \mathbf{v}^{(\eta)}) \cdot \hat{\mathbf{n}} = \mathbf{J}^{(S)} \cdot \hat{\mathbf{n}} - S \mathcal{Q}_m. \quad (51.28)$$

Turning this equation around leads to the non-advection flux

$$\mathbf{J}^{(S)} \cdot \hat{\mathbf{n}} = S \mathcal{Q}_m - \mathcal{Q}_S = S \mathcal{Q}_F - F \mathcal{Q}_S, \quad (51.29)$$

which relates the mass transport crossing the ocean surface at $z = \eta$ (right hand side) to the non-advection salt transport on the ocean side of the surface boundary (left hand side). A form of this equation was also given by equation (16.72). To support intuition and to check signs, consider the case with $\mathcal{Q}_S = 0$ so that $\mathbf{J}^{(S)} \cdot \hat{\mathbf{n}} = S \mathcal{Q}_F$. This expression means there is an upward non-advection flux of salt ($\mathbf{J}^{(S)} \cdot \hat{\mathbf{n}} > 0$) on the ocean side of the $z = \eta$ boundary in the presence of an input of freshwater through the ocean surface ($S \mathcal{Q}_F > 0$). For the converse, let $\mathcal{Q}_F = 0$ so that $\mathbf{J}^{(S)} \cdot \hat{\mathbf{n}} = -F \mathcal{Q}_S$. Now, there is a downward non-advection flux of salt ($\mathbf{J}^{(S)} \cdot \hat{\mathbf{n}} < 0$) on the ocean side of the $z = \eta$ boundary in the presence of salt input through the ocean surface ($F \mathcal{Q}_S > 0$).

51.3.2 The non-advection salt flux boundary condition

The above properties of boundary mass transfer result from the kinematic property of a fluid element whose mass is constant, and so the transfer of freshwater across the boundary of a fluid element is compensated by an opposite transfer of salt. The ocean boundary interface acts as a boundary for the fluid elements adjacent to the surface. Hence, to move mass across the $z = \eta$ interface requires mass to be replenished to the surface fluid elements.

Diffusive closure for the non-advection flux

Consider an ocean without any mixing, such as for a perfect fluid. In this case, mass arriving to the ocean surface from $\mathcal{Q}_m > 0$ will not be incorporated into the ambient ocean fluid, but instead

will remain a separate unmixed region. Hence, when mass is exchanged across the ocean surface, mixing is required to incorporate the mass into the ambient ocean fluid. To determine the level of mixing, assume that $\mathbf{J}^{(S)}$ takes the form of a diffusive flux (51.20) so that the boundary condition (51.29) becomes

$$\mathbf{J}^{(S)} \cdot \hat{\mathbf{n}} = -\rho [\mathbb{K} \cdot \nabla S] \cdot \hat{\mathbf{n}} = S \mathcal{Q}_m - \mathcal{Q}_s = S \mathcal{Q}_f - F \mathcal{Q}_s. \quad (51.30)$$

This equation sets the level of diffusion on the ocean side of the surface boundary that is needed to generate the non-advection transport. The diffusive mixing of salt and freshwater mediate the transfer of mass across the ocean surface so to incorporate that mass into the ambient ocean fluid. For example, freshwater added to the ocean ($\mathcal{Q}_f > 0$) diffuses downward as salt diffuses upward toward the surface.

Salt dissolved within the mass transport

In the case when salt is transported across the ocean boundary, as occurs with sea ice melting and formation, it does so largely dissolved in the water that is transported. There can also be a non-advection transport, such as via parameterized turbulent fluxes, so that the net salt flux is given by

$$\mathcal{Q}_s = S_m \mathcal{Q}_m + \mathcal{Q}_s^{\text{non-adv}}. \quad (51.31)$$

If there are more sources of this transfer then a relation such as this holds for each process. We are thus led to the net salt flux

$$\mathcal{Q}_s = -[\rho S(\mathbf{v} - \mathbf{v}^{(\eta)}) + \mathbf{J}^{(S)}] \cdot \hat{\mathbf{n}} = S \mathcal{Q}_m - \mathbf{J}^{(S)} \cdot \hat{\mathbf{n}} = S_m \mathcal{Q}_m + \mathcal{Q}_s^{\text{non-adv}}. \quad (51.32)$$

which leads to the non-advection salt flux on the ocean side of the boundary

$$-\mathbf{J}^{(S)} \cdot \hat{\mathbf{n}} = \mathcal{Q}_s^{\text{non-adv}} + (S_m - S) \mathcal{Q}_m. \quad (51.33)$$

Figure 51.1 provides a schematic summary of the salt flux boundary condition. Furthermore, note that this boundary condition is consistent with that derived in Section 16.5.3 for a general tracer, in particular with equation (16.69).

Treatment in observational analyses and numerical models

In ocean climate modeling applications, the salt mass flux, \mathcal{Q}_s , typically does not affect the kinematic boundary conditions. This approximation is reasonable given that the dominant contributor to the mass flux, \mathcal{Q}_m , is the freshwater. Even so, there remains a net salt transported across the ocean surface in the presence of sea ice melt and formation. The above boundary conditions, in particular equations (51.32) and (51.33), remain unchanged. Furthermore, it is necessary to specify the boundary tracer concentration, $S(z = \eta)$. For salt, this value is typically set equal to that within the ocean model surface grid cell. This choice is also common for observation-based studies.

51.3.3 Conservative Temperature boundary condition

Conservative Temperature, potential temperature, potential vorticity, and passive tracers each satisfy the tracer equation (51.15), with distinct tracer flux vectors \mathbf{J} . However, they are not material tracers and so the kinematic constraints holding for salt do not hold for these other tracers. We describe the thermodynamic properties of Conservative Temperature in Section 22.8 and the processes affecting its boundary fluxes in Section 51.4. Here we begin our treatment of this tracer by outlining its surface boundary condition.

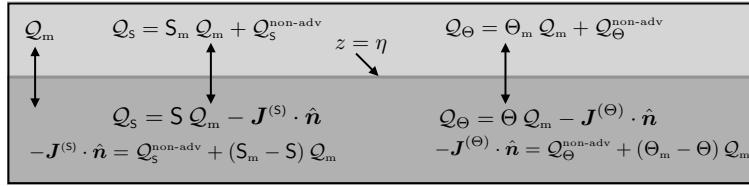


FIGURE 51.1: A schematic of an infinitesimal region of the ocean surface boundary at $z = \eta(x, y, t)$, with $z < \eta$ the ocean. $\mathcal{Q}_m d\mathcal{S}$ is the mass transport (mass per time) that crosses the interface and carries a tracer concentration. We depict the case for salt concentration, S , and Conservative Temperature, Θ , and the expressions for their net boundary fluxes.

As per the general discussion in Section 16.5.3, the net surface boundary flux of Conservative Temperature is written

$$\mathcal{Q}_\Theta = \Theta_m \mathcal{Q}_m + \mathcal{Q}_\Theta^{\text{non-adv}} = [-\rho \Theta (\mathbf{v} - \mathbf{v}^{(\eta)}) + \mathbf{J}^{(\Theta)}] \cdot \hat{\mathbf{n}} = \Theta \mathcal{Q}_m - \mathbf{J}^{(\Theta)} \cdot \hat{\mathbf{n}}. \quad (51.34)$$

In this equation, $\mathcal{Q}_\Theta^{\text{non-adv}}$ arises from the non-advective heat fluxes outside the ocean domain that impact on the upper ocean interface, such as from radiant and turbulent fluxes, whereas Θ_m is the Conservative Temperature of the boundary mass flux. Rearrangement leads to the net expression for the non-advective flux on the ocean side of the upper ocean boundary

$$-\mathbf{J}^{(\Theta)} \cdot \hat{\mathbf{n}} = \mathcal{Q}_\Theta^{\text{non-adv}} + (\Theta_m - \Theta) \mathcal{Q}_m, \quad (51.35)$$

where $\Theta = \Theta(z = \eta)$ is the Conservative Temperature at the surface interface. A common assumption made for models and observational studies is to set $\Theta_m - \Theta(z = \eta) = 0$, in which case

$$-\mathbf{J}^{(\Theta)} \cdot \hat{\mathbf{n}} = \mathcal{Q}_\Theta^{\text{non-adv}} \text{ if } \Theta_m - \Theta(z = \eta) = 0. \quad (51.36)$$

Figure 51.1 provides a schematic summary of the Θ flux boundary condition.

51.3.4 Comments and further reading

[Nurser and Griffies \(2019\)](#) offer further discussion of the kinematic boundary condition for salt and freshwater. We make use of many of the notions of this section when discussing surface ocean buoyancy fluxes in Section 51.4 and water mass transformation in Section 52.7.

51.4 Surface boundary fluxes of buoyancy

As introduced in Chapter 25, buoyancy measures the gravitational acceleration of a fluid element relative to that of the fluid environment surrounding the element. A reduction in density for the fluid element is associated with an increase in buoyancy; that is, the fluid element becomes more *buoyant*. Changes in buoyancy arise through changes in density associated with temperature and salinity changes, with buoyancy changes computed relative to a fixed pressure level. In this way, buoyancy changes are directly related to processes that impact locally referenced potential density through changes in the temperature and salinity of a fluid element.

In this section we derive the equation describing the changes in ocean buoyancy due to enthalpy (sometimes referred to as “heat”), salt, and water fluxes crossing the ocean boundaries. For this purpose, we expose certain of the issues associated with coupling numerical models of the ocean, atmosphere, and land. A detailed treatment of boundary layer physics is well outside of our scope. We thus take a phenomenological perspective, developing budget equations but not diving

into details of the turbulent exchange of matter and enthalpy across the ocean surface boundary. Furthermore, we are only concerned with the upper ocean boundary, so that we ignore geothermal fluxes crossing the ocean bottom boundary.

51.4.1 Outlining the boundary fluxes of enthalpy and salt

Broadly, the boundary fluxes are associated with the following physical processes.

- Turbulent processes transfer enthalpy through latent and sensible heating.
- Longwave radiation cools the upper ocean, with this radiation affected by the upper ocean skin temperature.
- Penetrative shortwave radiation is absorbed in seawater and so increases buoyancy in regions where the thermal expansion coefficient is positive.⁵
- All of the above processes are referred to as *non-advection* transports. They are not associated with a net mass transport across the ocean surface. In contrast, advective processes transfer enthalpy and salt across the ocean surface through the transfer of mass across the interface.
- Salt is transferred between the liquid ocean and sea ice when sea ice melts and forms. This transfer is proportional to the water mass flux and the difference in salinity between the liquid ocean and sea ice. There may be additional turbulent salt fluxes as well, but there is a negligible transfer of salt associated with precipitation, evaporation, or river runoff.

51.4.2 Evolution from surface boundary fluxes

We now develop finite volume budget equations for potential enthalpy (via Conservative Temperature, Θ), salt, and seawater mass for a grid cell region next to the ocean surface, with a focus on contributions due to surface boundary fluxes. For that purpose, introduce the following quantities for a grid cell,

$$M = \int_{\text{cell}} \rho dV = \langle \rho \rangle V \quad V = \int_{\text{cell}} dV = A \bar{h} \quad A = \int_{\text{cell}} dA \quad (51.37a)$$

$$\bar{h} A = \int_{\text{cell}} \left[\int_{\text{cell}} dz \right] dA \quad \langle C \rangle M = \int_{\text{cell}} C \rho dV, \quad (51.37b)$$

so that $\langle \rho \rangle$ is the cell averaged density, $\langle C \rangle$ is the cell mass weighted averaged tracer concentration, \bar{h} is the cell area averaged thickness, V is the cell volume, and A is the cell horizontal area. These definitions allow us to write

$$\frac{d}{dt} \left[\int_{\text{cell}} \rho C dV \right] = \frac{d}{dt} [\langle C \rangle M] = A \frac{d}{dt} [\bar{h} \langle C \rangle \langle \rho \rangle], \quad (51.38)$$

where the horizontal area of a cell is assumed to be constant in time. The surface boundary fluxes have similar grid cell area averages.

⁵The Baltic Sea is an outlier in the World Ocean, whose fresh and cold waters often realize a negative thermal expansion so that heating can increase rather than reduce density.

Focusing just on contributions from surface boundary transport leads to the budget equations

$$\frac{\partial(\rho h \Theta)}{\partial t} = Q_m \Theta_m + Q_\Theta^{\text{non-adv}} \quad (51.39\text{a})$$

$$\frac{\partial(\rho h S)}{\partial t} = Q_m S_m + Q_S^{\text{non-adv}} \quad (51.39\text{b})$$

$$\frac{\partial(\rho h)}{\partial t} = Q_m, \quad (51.39\text{c})$$

where we used a partial time derivative since we are holding the horizontal position fixed. Furthermore, we reduced notational clutter by dropping the angle brackets for volume average and the horizontal overline for area average. For a three-dimensional budget, the right hand side to these equations is combined with fluxes crossing interior cell boundaries. Finally, we wrote the fluxes as

$$Q_m A = \mathcal{Q}_m \mathcal{S} \quad \text{and} \quad Q_\Theta^{\text{non-adv}} A = \mathcal{Q}_\Theta^{\text{non-adv}} \mathcal{S} \quad \text{and} \quad Q_S^{\text{non-adv}} A = \mathcal{Q}_S^{\text{non-adv}} \mathcal{S}, \quad (51.40)$$

where \mathcal{S} is the area on the free surface and A is the corresponding horizontal area of the grid cell.

51.4.3 Buoyancy tendency from surface boundary fluxes

For many purposes, it is of interest to quantify the impacts on ocean buoyancy arising from surface boundary fluxes. For that purpose, we here develop the budget for buoyancy in a surface model grid cell region, focusing on surface flux contributions.

Buoyancy has a time tendency given by

$$-\frac{\rho_0}{g} \frac{\partial b}{\partial t} = \rho_\Theta \frac{\partial \Theta}{\partial t} + \rho_S \frac{\partial S}{\partial t}, \quad (51.41)$$

where we introduced the shorthand

$$\rho_\Theta = \left[\frac{\partial \rho}{\partial \Theta} \right]_{S,p} \quad \text{and} \quad \rho_S = \left[\frac{\partial \rho}{\partial S} \right]_{\Theta,p} \quad (51.42)$$

for the partial derivatives of density with respect to Conservative Temperature and salinity, respectively, each with pressure held constant. We wish to form an evolution equation for buoyancy at the ocean surface grid cell just due to the effects of surface forcing. For this purpose, multiply the temperature equation (51.39a) by ρ_Θ and add to the salinity equation (51.39b) multiplied by ρ_S

$$\rho_\Theta \left[\frac{\partial(\rho h \Theta)}{\partial t} \right] + \rho_S \left[\frac{\partial(\rho h S)}{\partial t} \right] = Q_m (\rho_\Theta \Theta_m + \rho_S S_m) + \rho_\Theta Q_\Theta^{\text{non-adv}} + \rho_S Q_S^{\text{non-adv}}. \quad (51.43)$$

Now use the mass budget (51.39c) and introduce the buoyancy tendency according to equation (51.41) to render an expression for the time tendency of the surface ocean buoyancy

$$-(\rho_0/g) \rho h \left[\frac{\partial b}{\partial t} \right]^{\text{surface}} = Q_m [\rho_\Theta (\Theta - \Theta_m) + \rho_S (S - S_m)] + \rho_\Theta Q_\Theta^{\text{non-adv}} + \rho_S Q_S^{\text{non-adv}}. \quad (51.44)$$

Introducing the thermal expansion and saline contraction coefficients

$$\alpha_\Theta = -\frac{1}{\rho} \left[\frac{\partial \rho}{\partial \Theta} \right]_{S,p} \quad \text{and} \quad \beta_S = \frac{1}{\rho} \left[\frac{\partial \rho}{\partial S} \right]_{\Theta,p} \quad (51.45)$$

yields

$$\left[\frac{\partial b}{\partial t} \right]_{\text{surface}} = \frac{g}{\rho_0 h} \left(Q_m [\alpha_\Theta (\Theta - \Theta_m) - \beta_S (S - S_m)] + \alpha_\Theta Q_\Theta^{\text{non-adv}} - \beta_S Q_S^{\text{non-adv}} \right) \quad (51.46a)$$

$$= \frac{g}{\rho_0 h} \left(\alpha_\Theta [Q_m (\Theta - \Theta_m) + Q_\Theta^{\text{non-adv}}] - \beta_S [Q_m (S - S_m) + Q_S^{\text{non-adv}}] \right). \quad (51.46b)$$

In regions where the thermal expansion coefficient is positive, we see that adding a boundary mass with $\Theta > \Theta_m$ (i.e., relatively warm water) increases surface ocean buoyancy. Likewise, where the haline contraction coefficient is positive, adding boundary mass with $S < S_m$ (i.e., relatively freshwater) increases surface ocean buoyancy. Finally, note that in some contexts it is useful to take the limit as the thickness, h , becomes vanishingly small and to introduce a Dirac delta distribution (see Section 4.2) and thus write⁶

$$\left[\frac{\partial b}{\partial t} \right]_{\text{surface}} = \frac{g \delta(z - \eta)}{\rho_0} \left(Q_m [\alpha_\Theta (\Theta - \Theta_m) - \beta_S (S - S_m)] + \alpha_\Theta Q_\Theta^{\text{non-adv}} - \beta_S Q_S^{\text{non-adv}} \right) \quad (51.47a)$$

$$= \frac{g \delta(z - \eta)}{\rho_0} \left(\alpha_\Theta [Q_m (\Theta - \Theta_m) + Q_\Theta^{\text{non-adv}}] - \beta_S [Q_m (S - S_m) + Q_S^{\text{non-adv}}] \right), \quad (51.47b)$$

with this form of use when organizing processes according to interior processes and surface boundary processes.

51.4.4 Comments

The buoyancy flux expression (51.46b) is of use for boundary layer parameterizations, such as the KPP scheme of [Large et al. \(1994\)](#) and [Van Roekel et al. \(2018\)](#). It is furthermore used when studying water mass transformations as reviewed by [Groeskamp et al. \(2019\)](#) and summarized in Chapter 52.

51.5 Global mean sea level

In this section we consider some basic features of global mean sea level by making use of the mass budget of liquid seawater. This analysis highlights the distinction between the mass budget and volume (sea level) budget.

51.5.1 Definitions and assumptions

Seawater mass is a conserved quantity so that the total liquid seawater mass, m , changes only via boundary mass fluxes

$$\frac{dm}{dt} = \mathcal{A} \overline{Q_m}, \quad (51.48)$$

where \mathcal{A} is the ocean surface area and $\overline{Q_m}$ is the area averaged surface mass flux. The global volume of liquid seawater

$$V = \frac{m}{\langle \rho \rangle} \quad (51.49)$$

changes due to mass changes *and* changes to the global mean density, $\langle \rho \rangle$. Throughout this section we assume the surface area is constant in time, thus neglecting the relatively small changes

⁶From Section 4.2, we know that the Dirac delta distribution, $\delta(z - \eta)$, has dimensions of inverse length.

associated with volume changes along sloping beaches. We also assume constant area averaged ocean bottom depth, \bar{H} . These two assumptions mean that changes in ocean volume arise just from changes in global mean sea level, $\bar{\eta}$. Since around the year 2000, measurements estimate that global area mean sea level has increased at a rate of

$$\left[\frac{d\bar{\eta}}{dt} \right]_{\text{observed}} \approx 3 \text{ mm yr}^{-1}, \quad (51.50)$$

and that this rate is increasing (positive sea level acceleration). As part of the analysis in this section we make use of the following phenomenological numbers and make a few assumptions to facilitate calculations.

- global seawater volume $\mathcal{V} \approx 1.3 \times 10^{18} \text{ m}^3$
- global ocean surface area $\mathcal{A} \approx 3.6 \times 10^{14} \text{ m}^2$
- global ocean mean density $\langle \rho \rangle \approx 1035 \text{ kg m}^{-3}$
- specific heat capacity for seawater $c_p \approx 3992 \text{ J kg}^{-1}\text{K}^{-1}$
- Ignore mass fluxes transported through the sea floor, which are small relative to surface mass fluxes.
- Ignore salinity and pressure effects on density, so that changes in global mean density arise just from changes in global mean Conservative Temperature.
- Assume a constant thermal expansion coefficient

$$\alpha_\Theta = -\frac{1}{\rho} \left[\frac{\partial \rho}{\partial \Theta} \right]_{S,p} \approx 2 \times 10^{-4} \text{ K}^{-1}. \quad (51.51)$$

This is not a great approximation, since the thermal expansion coefficient ranges over the ocean by a factor of 10. Nonetheless, for this section it is sufficient for deducing rough numbers that are consistent with errors in measurements for global boundary enthalpy and mass fluxes.

51.5.2 Budget for global mean sea level

Expression (51.49) for ocean volume leads to its time derivative

$$\frac{d\mathcal{V}}{dt} = \frac{1}{\langle \rho \rangle} \frac{dm}{dt} - \frac{m}{\langle \rho \rangle^2} \frac{d\langle \rho \rangle}{dt} \quad (51.52a)$$

$$= \frac{\mathcal{A} \bar{Q}_m}{\langle \rho \rangle} - \frac{\mathcal{V}}{\langle \rho \rangle} \frac{d\langle \rho \rangle}{dt}, \quad (51.52b)$$

where we used equation (51.48) to express mass changes in terms of the surface mass flux. Additionally, the ocean volume is given by

$$\mathcal{V} = \int dA \int_{-H}^{\bar{\eta}} dz = \mathcal{A} (\bar{H} + \bar{\eta}), \quad (51.53)$$

so that its time changes arise from changes in the global mean sea level

$$\frac{d\mathcal{V}}{dt} = \mathcal{A} \frac{d\bar{\eta}}{dt}. \quad (51.54)$$

Combining the two volume equations (51.52b) and (51.54) yields the budget equation for global mean sea level

$$\frac{d\bar{\eta}}{dt} = \frac{\overline{Q_m}}{\langle \rho \rangle} - \frac{\mathcal{V}}{\mathcal{A} \langle \rho \rangle} \frac{d\langle \rho \rangle}{dt}. \quad (51.55)$$

The first term arises from changes in ocean mass whereas the second term arises from changes in global mean seawater density.

51.5.3 Changes due to mass input

To ground these formula in phenomenology, assume that a surface mass flux gives one-half of the observed sea level rise

$$\frac{1}{2} \left[\frac{d\bar{\eta}}{dt} \right]_{\text{observed}} = \frac{\overline{Q_m}}{\langle \rho \rangle}. \quad (51.56)$$

This distribution of measured sea level is roughly correct. With $\langle \rho \rangle = 1035 \text{ kg m}^{-3}$ and $d\bar{\eta}/dt \approx 3 \text{ mm yr}^{-1}$, we need an area averaged mass flux across the ocean surface

$$\overline{Q_m} \approx 5 \times 10^{-8} \text{ kg m}^{-2} \text{ s}^{-1}. \quad (51.57)$$

Integrated over the global ocean area, this flux leads to a mass transport of

$$\mathcal{T} = \mathcal{A} \overline{Q_m} \approx 1.8 \times 10^7 \text{ kg s}^{-1} \approx 0.015 \times \mathcal{T}^{\text{river}}. \quad (51.58)$$

That is, global mean sea level rises at a rate of 1.5 mm yr^{-1} if there is a net additional mass added to the ocean equal to roughly 1.5% of the net river water entering the ocean, $\mathcal{T}^{\text{river}}$. This additional net mass is coming from the melting of land-ice.

51.5.4 Steric changes due to changes in density

Steric effects generally refer to properties of a substance associated with the space occupied by atoms. In the sea level context, steric effects refer to changes in sea level associated with density changes, with changes in density associated with changes in the volume occupied by seawater molecules. Changes in global mean sea level arising from changes in the global mean density are called global steric sea level changes. From the sea level budget equation (51.55) we know that steric changes are written mathematically as

$$\left[\frac{d\bar{\eta}}{dt} \right]_{\text{steric}} \equiv -\frac{\mathcal{V}}{\mathcal{A} \langle \rho \rangle} \frac{d\langle \rho \rangle}{dt}. \quad (51.59)$$

Global mean density will change primarily from changes in global mean Conservative Temperature. If we assume the ocean thermal expansion is constant, then

$$\frac{1}{\langle \rho \rangle} \frac{d\langle \rho \rangle}{dt} = -\alpha_\Theta \frac{d\langle \Theta \rangle}{dt}, \quad (51.60)$$

so that steric sea level changes are primarily driven by *thermosteric* effects

$$\left[\frac{d\bar{\eta}}{dt} \right]_{\text{thermosteric}} \equiv \frac{\alpha_\Theta \mathcal{V}}{\mathcal{A}} \frac{d\langle \Theta \rangle}{dt}. \quad (51.61)$$

With $\alpha_\Theta \approx 2 \times 10^{-4} \text{ K}^{-1}$ and $d\bar{\eta}/dt \approx 1.5 \text{ mm yr}^{-1}$, we have

$$\frac{d\langle \Theta \rangle}{dt} \approx 0.2 \text{ K century}^{-1}. \quad (51.62)$$

That is, a global thermosteric sea level rise of 1.5 mm yr^{-1} corresponds to a rate of increase in the global volume mean ocean temperature of roughly $0.2 \text{ K century}^{-1}$.

51.5.5 Enthalpy flux imbalances giving rise to thermosteric sea level

A global mean ocean temperature change can arise from an area averaged surface ocean enthalpy flux

$$\overline{Q_H} \approx \langle \rho \rangle c_p \overline{H} \frac{d\langle \Theta \rangle}{dt}. \quad (51.63)$$

Plugging in numbers leads to the enthalpy flux

$$\overline{Q_H} \approx 1 \text{ W m}^{-2}. \quad (51.64)$$

That is, a surface ocean enthalpy flux of roughly 1 W m^{-2} (ocean area normalized) gives rise to a global mean thermosteric sea level rise of roughly 1.5 mm yr^{-1} .

An enthalpy flux of 1 W m^{-2} is small compared to, say, that crossing the surface of a typical light bulb. However, 1 W m^{-2} is comparable to that accumulating within the earth system due to increases in greenhouse gases ([Otto et al., 2013](#)). That is, 1 W m^{-2} averaged over the global ocean, or 0.7 W m^{-2} averaged over the surface area of the planet,⁷ is roughly the net heating associated with anthropogenic climate change. Such seemingly small increases in surface heating represent a nontrivial increase in the earth's energy budget that are leading to the observed sizable climate changes and sea level rise.

A specific means to gauge the magnitude of 1 W m^{-2} distributed over the ocean surface area, \mathcal{A} , is to compare to the enthalpy flux due to blasting one atomic bomb per second ($\Delta t = 1 \text{ s}$) and uniformly distributing its released energy over the ocean surface area every second. $\mathcal{E}_{\text{bomb}} \approx 6.3 \times 10^{13} \text{ J}$ distributed over the ocean area each second corresponds to a surface ocean enthalpy flux of

$$Q_{\text{bomb}} = \frac{\mathcal{E}_{\text{bomb}}}{\mathcal{A} \Delta t} \approx 0.17 \text{ W m}^{-2}. \quad (51.65)$$

Hence, an enthalpy flux of 1 W m^{-2} due to anthropogenic climate warming corresponds to $1/0.17 \approx 6$ bombs per second. This way of framing the net heating of the ocean due to anthropogenic climate change dramatically illustrates the huge magnitude of the heating created by the burning of carbon based energy sources.

51.5.6 Global sea level in a Boussinesq fluid

The sea level for a Boussinesq ocean evolves according to the kinematic free surface equation for an incompressible fluid (equation (17.8))

$$\frac{\partial \eta^{\text{bouss}}}{\partial t} = -\nabla \cdot \mathbf{U} + Q_m/\rho_0, \quad (51.66)$$

where

$$\mathbf{U} = \int_{-H}^{\eta} \mathbf{u} dz \quad (51.67)$$

is the depth integrated horizontal velocity and ρ_0 is the Boussinesq reference density. This equation results from ignoring all changes to density, except for those related to the buoyancy force appearing in the momentum equations. Integrating the sea surface height equation (51.66) over the surface area of the global ocean reveals that the global mean Boussinesq sea level evolves according to

$$\frac{d\overline{\eta}^{\text{bouss}}}{dt} = \frac{\overline{Q}_m}{\rho_0}. \quad (51.68)$$

⁷The ocean covers about 70% of the earth surface.

Hence, $\bar{\eta}^{\text{bouss}}$ changes only so long as there are boundary mass fluxes. In contrast to the real ocean, the sea level computed from a Boussinesq ocean is unaffected by a surface enthalpy flux, $\bar{Q}_H \neq 0$. This result in turn means that we cannot use the prognostic sea surface height, η^{bouss} , level from a Boussinesq ocean model to compute changes the global mean sea level. Instead, corrections are required, as first identified by [Greatbatch \(1994\)](#) and further detailed in Appendix D of [Griffies and Greatbatch \(2012\)](#).

51.5.7 Global halosteric sea level changes are negligible

When freshwater enters the ocean, such as from melting continental ice sheets, it adds to the ocean mass and in turn increases global mean sea level. This increase is referred to as *barystatic sea level rise* according to the sea level terminology paper from [Gregory et al. \(2019\)](#). Although ocean salinity changes upon changing the freshwater content, the net effect on global mean sea level is almost entirely barystatic since the global halosteric effect is negligible. We can understand why the global halosteric effect is so tiny by recognizing that freshwater entering the ocean sees its salinity increase whilst the ambient seawater is itself freshened. These compensating salinity changes (which are often incorrectly ignored) have corresponding compensating sea level changes, thus bringing the global halosteric effect to near zero. Appendix B of [Gregory et al. \(2019\)](#) works through a two-bucket thought experiment where one bucket holds freshwater and the other holds seawater, with the Conservative Temperature and pressures identical for the two buckets. They compute how the total water volume changes upon homogenizing the water in the two buckets, while conserving the masses of freshwater and salt. Their result shows that the total volume of homogenized water equals to the sum of the volume initially in the two separate buckets to within 0.1%. Hence, the change in volume is almost entirely barystatic, with the global halosteric effect entirely negligible when considering global sea level changes.

51.5.8 Further study

The discussion of steric and thermosteric sea level changes are further explored in [Griffies and Greatbatch \(2012\)](#) and [Griffies et al. \(2014\)](#). The global halosteric discussion is based on Appendix B of the sea level terminology paper from [Gregory et al. \(2019\)](#), which presents a two-bucket thought experiment. The [Gregory et al. \(2019\)](#) paper is also notable for providing a conceptual rationalization of the often confusing terminology used in sea level studies.



Water mass analysis

In this chapter we develop the formalism of *water mass analysis* and its natural extension to *tracer mass analysis*. This formalism considers the budgets for seawater mass and tracer mass within layers or classes defined by properties such as buoyancy, Conservative Temperature, salinity, or biogeochemical tracers. Quite generally, a *water mass* refers to a region of seawater characterized by a suite of quasi-homogeneous properties used to distinguish this water from other water masses. Water masses are typically formed through surface boundary processes, such as the extremely large buoyancy fluxes at the high latitudes that form the Antarctic Bottom Water (AABW) and North Atlantic Deep Water (NADW). As these waters enter the ocean interior they are advected over basin scales while they are also eroded or *transformed* by irreversible mixing processes. Water masses offer a conceptual means to partition or bin the ocean into distinct classes whose origin, movement, and transformation can be measured, modeled, and studied. Furthermore, scalar properties generally used to classify water masses are simpler to measure than vector properties such as velocity. Hence, a water mass perspective offers the means to infer ocean circulation within water mass space without directly measuring the velocity.

READER'S GUIDE FOR THIS CHAPTER

The theoretical development in this chapter requires the vector calculus encountered in Chapter 2; elements of the generalized vertical coordinates from Chapters 43 and 44; kinematics from Chapter 15; tracer budgets from Chapter 16; advection and diffusion maths in Chapter 48, and features of parameterized tracer transport and mixing discussed in Chapter 50. For the discussion of generalized water mass analysis in Section 52.12, we make use of the tutorial on differential forms given in Chapter 9. Although some of this prior material is reviewed here to keep the discussion self-contained, the reader may find it necessary to revisit the earlier material to garner a full understanding of the concepts and manipulations. This is a rather long chapter that works through a number of related budget analyses with the aim to ground the theory with examples motivated from the growing literature.

When approaching this subject, particularly for the first time, many readers may appreciate the following paraphrase from A. Sommerfeld's quote in the thermodynamics Chapter 21.

Water mass analysis appears somewhat mysterious and puzzling on first encounter. On second encounter things start to fall into place, except perhaps for a few pesky math niceties. On third encounter, when deciding to do calculations, one returns to that unsettled feeling of the first encounter. However, by now, familiarity with the words and maths means that the mystery presents no practical bother. One simply turns the crank without thinking too much about the underlying foundations.

In hopes of partially dispelling the mystery, and maintaining an appreciation for the foundations, we here couple the many mathematical equations with schematics and conceptual descriptions. Even so, there should be no presumption that this subject is easy.

52.1	Loose ends	1183
52.2	Conceptual framework	1183
52.2.1	Water mass configuration space	1183
52.2.2	A novel kinematic lens	1184
52.2.3	Transformation and formation	1185
52.3	Buoyancy transformation and formation	1186
52.3.1	A three-layer thought experiment	1186
52.3.2	How processes lead to transformations	1187
52.4	Mathematical framework	1188
52.4.1	Seawater mass in an infinitesimal cylinder	1189
52.4.2	Seawater mass within a finite region	1189
52.4.3	Seawater mass distribution/density function	1190
52.4.4	Example regions	1190
52.4.5	Integrals of arbitrary functions	1193
52.4.6	Moments of λ	1193
52.4.7	Internal and external λ -moments	1194
52.4.8	Further study	1195
52.5	Water mass transformation across a λ -surface	1195
52.5.1	Dia-surface flux and interior transformation	1195
52.5.2	Transformation as the derivative of an integral	1196
52.5.3	Processes and kinematics	1197
52.5.4	Some details concerning interior transformation	1199
52.6	Budget for seawater mass in a $\Delta\lambda$ -layer	1200
52.6.1	Transport crossing interior open boundaries	1201
52.6.2	Mass transport crossing the ocean surface	1201
52.6.3	Mass budget	1202
52.7	Budget for λ mass in a λ_∞ -region	1202
52.7.1	Processes affecting the mass of λ -stuff	1202
52.7.2	Summary of the λ budget	1204
52.7.3	Processes leading to water mass transformation	1205
52.8	Surface water mass transformation	1206
52.8.1	Circulation driven by surface transformation	1207
52.8.2	Further study	1208
52.9	Buoyancy water mass transformation	1208
52.9.1	Material time changes to S and Θ	1208
52.9.2	Water mass transformation	1209
52.9.3	Summary of surface non-advective flux for S and Θ	1210
52.10	Tracer mass analysis	1210
52.10.1	General form of the mass budget	1210
52.10.2	Tracer processes	1211
52.10.3	Transport across an interior layer interface	1212
52.10.4	Transport across interior and surface boundaries	1213
52.10.5	The layer tracer budget	1214
52.10.6	Further study	1214
52.11	Regions bounded by a tracer contour/surface	1214
52.11.1	Closed region bounded by a tracer surface/contour	1214
52.11.2	Region with $C \geq \tilde{C}$	1216
52.11.3	Comments and further reading	1218

52.12 Mathematics of water mass configuration space	1218
52.12.1 Water mass configuration space	1218
52.12.2 Trajectories and velocity in \mathbf{q} -space	1219
52.12.3 Boundaries in \mathbf{q} -space	1220
52.12.4 Mass for an element of \mathbf{q} -space	1220
52.12.5 The mass transport differential form	1221
52.12.6 Mass continuity in \mathbf{q} -space	1222
52.12.7 Streamfunction for steady and source-free circulation	1223
52.12.8 Mass continuity when \mathbf{q} -space and \mathbf{x} -space are 1-to-1	1224
52.12.9 Water mass angular momentum	1226

52.1 Loose ends

- Nothing ATT.

52.2 Conceptual framework

Water mass analysis is a mathematical formalism supporting the study of budgets for seawater mass and tracer mass within layers or classes defined by properties such as Archimedean buoyancy (Chapter 25; shortened to “buoyancy” here), Conservative Temperature, salinity, or biogeochemical tracers. Water mass analysis is concerned with processes affecting the evolution of fluid within layers and in the characterization of circulation inferred from this evolution. We speak of a *water mass configuration space*, in which some or all of the coordinates of this space are determined by properties other than geographic/depth coordinates.¹ For example, Figure 52.1 provides an example configuration space described by (S, Θ, p) , whereby a fluid element has its position in water mass configuration space² determined by its Conservative Temperature, Θ , salinity, S , and pressure, p . We may also consider the water mass configuration space defined by binning the ocean according to potential density classes, or some other measure of buoyancy. We can also retain some degree of geographical information through binning according to latitude and buoyancy. Note that there is no implied constraint that the buoyancy surfaces remain stably stratified in the vertical, nor indeed that any of the water mass coordinates retain a 1-to-1 relation with geographic/depth space. We have more to say on this point in the following.

52.2.1 Water mass configuration space

We specify the position of a fluid element within water mass configuration space by specifying a list of properties and/or spatial information. That is, we can choose any number of properties or geographical coordinates to specify water mass configuration space. Consequently, water mass configuration space has an arbitrary number of dimensions. We furthermore observe that water mass configuration space generally has no metric³, particularly when none of the chosen coordinates are geographical (latitude or longitude) or depth. Hence, there is generally no notion of distance or angles between points in water mass configuration space. Indeed, what does it mean to be

¹We prefer the term “configuration space” over the alternative “phase space”, since phase space in Hamiltonian dynamics specifically refers to position and momentum coordinates. In contrast, configuration space, as used in our discussion of water masses, can be determined by most any property or geographic position.

²This water mass configuration space also serves as a thermodynamic configuration space when using the Gibbs potential discussed in Section 21.4.5.

³See Section 7.1 for a discussion of metric tensors, which are needed to measure distance on a manifold.

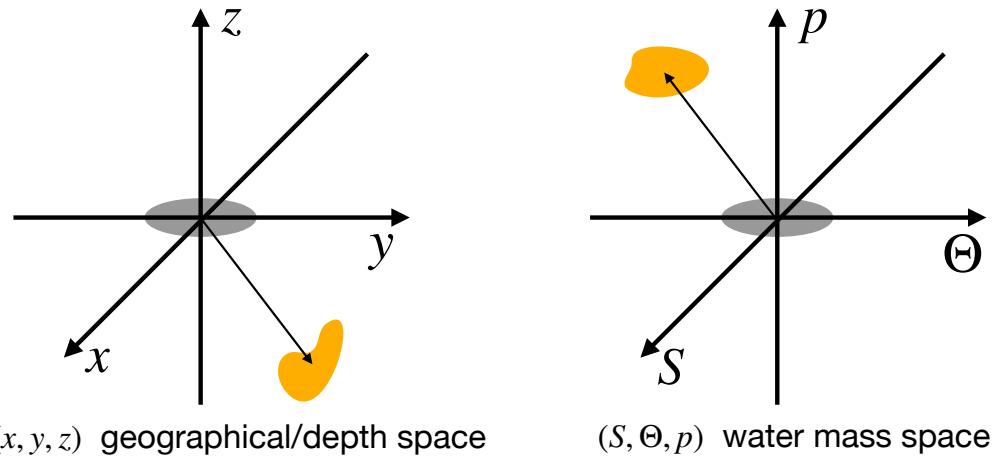


FIGURE 52.1: Left panel: a fluid element is positioned in geographical/depth space (x, y, z) according to its horizontal (x, y) (longitude, latitude) position and its vertical geopotential, z . Right panel: the same fluid element is positioned in a particular water mass configuration space, here defined by (S, Θ, p) with salinity, S , Conservative Temperature, Θ , and pressure, p . Mapping between the two spaces is generally not 1-to-1. Namely, a point in configuration space can be occupied by more than one point in geographical space since more than a single point in geographical space can have the same values for (S, Θ, p) . Although the coordinate axes are depicted as mutually orthogonal in the right panel, there is no objective means to determine angles in water mass configuration space since it contains no metric. That is, water mass configuration space, just as thermodynamic state space (Section 21.1.5), is a *topological manifold*.

orthogonal and temperature-salinity space or when studying the density-binned distribution of seawater?

The absence of a metric is something we have already seen when studying thermodynamic configuration space in Chapter 21, with mention of this point given in Section 21.1.5. Mathematically, we say that both thermodynamic configuration space and water mass configuration space are *topological manifolds*. Even so, one commonly sees a point in thermodynamic space depicted on a diagram with orthogonal axes (e.g., pressure-volume diagrams), or a position in water mass configuration space similarly depicted with orthogonal axes as in Figure 52.1. However, this depiction is arbitrary since there is no geometric structure afforded to such spaces since there is no metric tensor, which in turn means we cannot determine angles or orthogonality. Depictions with orthogonal axes merely satisfy a subjective desire for geometric structure when in fact there is none afforded to the manifold. Differential forms provide a suitable formalism for conducting calculus on a differential manifold sans a metric. We introduce the rudiments of differential forms in Chapter 9 and then make use of them in Section 52.12 in our study of generalized water mass analysis.

52.2.2 A novel kinematic lens

A water mass configuration space views the ocean circulation through the lens afforded by layers or classes. This perspective contrasts to the fixed points or regions of an Eulerian perspective, or the moving fluid particles or corresponding fluid regions of a Lagrangian perspective. Hence, the kinematic perspective of water mass analysis complements that from the Eulerian and Lagrangian kinematics encountered in other parts of this book (see Chapter 13 for a reminder of Eulerian and Lagrangian kinematics).

A point in water mass configuration space can be comprised of many points in geographic/depth space. Hence, there is no presumption of a 1-to-1 mapping between geographic/depth space and water mass configuration space. This situation is familiar to descriptive oceanographers who well know that there can be more than one fluid element in the ocean that has a particular (S, Θ, p)

value. Or more precisely, when classifying seawater according to binned values, $(\Theta \pm \Delta\Theta/2, S \pm \Delta S/2, p \pm \Delta p/2)$, fluid elements from any part of the ocean can fit into a particular bin so that a distribution of seawater in water mass configuration space is non-local in geographical/depth space. Even when classifying seawater according to latitude, longitude, and potential density, the possible case of a non-monotonic vertical potential density profile means that there is generally no 1-to-1 mapping between potential density and depth.

The lack of 1-to-1 mapping between water mass configuration space and geographic/depth space is a fundamental kinematic distinction from the 1-to-1 relation that holds between the Eulerian and Lagrangian descriptions of fluid motion (see Chapter 13). This property can be frustrating since circulation viewed in water mass configuration space generally has incomplete geographic/depth information, and yet oceanographers wish to know where on the planet something is happening.⁴ Conversely, abandoning the 1-to-1 relation is liberating since water mass configuration space provides a framework to infer ocean circulation, within the water mass space, even without measuring velocity of the fluid in geographical space. Correspondingly, viewing ocean circulation through a water mass lens can offer understanding that complements traditional Eulerian or Lagrangian views.

52.2.3 Transformation and formation

Water moves through water mass configuration space as it is modified or *transformed* by irreversible boundary and interior ocean processes.⁵ Such processes lead to material changes in fluid elements through mixing, solar radiation, and chemical reactions. The convergence (i.e., the local imbalance) of such transformation processes leads to the *formation* and destruction of water mass classes. As water moves through water mass configuration space we are afforded a distinct view of ocean circulation that has both direct and indirect connections to circulation in geographical/depth space. Notably, and quite trivially, we measure zero motion along a coordinate axis in water mass configuration space when the property defining that axis remains unchanged. For example, adiabatic and isohaline processes such as linear waves can render nontrivial motion in geographical/depth space whereas they lead to no motion in (S, Θ) space.

If one is interested in reversible processes then water mass configuration space generally offers an inappropriate kinematic perspective. Furthermore, given the possible non-local (x, y, z) aspects of water mass configuration space, it provides an unnatural venue to study forces and stresses acting between spatially adjacent fluid elements. Hence, momentum dynamics is better handled via Eulerian or Lagrangian kinematics. Where water mass configuration space shines is by revealing the dynamics associated with irreversible processes that affect properties defining the water mass classes. For example, a water mass perspective has found use in framing key questions of primary interest in the Anthropocene, such as ocean buoyancy and its transformation through interior and boundary mixing, ocean heat uptake and transport, the hydrological cycle, steric sea level rise, and irreversible changes to biogeochemical properties (see [Groeskamp et al. \(2019\)](#) for a review with many references).

⁴Auxiliary methods such as the water tagging method of [Groeskamp et al. \(2014\)](#) can be used to recover some geographical information.

⁵In many parts of this book the word “transformation” refers to coordinate transformations. Here, “transformation” refers to a process acting to irreversibly change seawater properties.

52.3 Buoyancy transformation and formation

Archimedean buoyancy is a common property used to distribute seawater, this buoyancy commonly approximated by potential density (Chapter 25). In this section we introduce the notions of *transformation* and *formation* when partitioning the ocean according to density (γ) classes that locally measure buoyancy.⁶ The ideas presented here extend to any scalar property used to bin the ocean fluid.

Water mass *transformation* measures the mass per time of water that moves across an isosurface in space, or moves from one bin to another within a distribution. By convention, the transformation is positive if water moves to larger density (more generally to a larger value for the property defining the layer or class) and negative if it enters a lighter density layer (smaller value for the property). Water mass *formation* refers to the difference in transformation across the surfaces bounding a layer, so that formation measures the change in mass of the layer. That is, *formation is the layer integrated convergence of transformation in water mass configuration space*.

Both transformation and formation have dimensions of mass per time (or volume per time when considering Boussinesq fluids) and are typically measured in Sverdrup units:

$$1 \text{ Sv} = 10^6 \text{ m}^3 \text{ s}^{-1} \quad \text{volume-Sverdrup} \quad (52.1a)$$

$$1 \text{ Sv} = 10^9 \text{ kg s}^{-1} \quad \text{mass-Sverdrup.} \quad (52.1b)$$

The volume-Sverdrup is commonly used in Boussinesq fluids with kinematics based on volume conservation (Chapter 24), whereas the mass-Sverdrup is commonly used in non-Boussinesq fluids whose kinematics is based on mass conservation. Even so, the mass-Sverdrup can be used for Boussinesq fluids merely by multiplying the volume-Sverdrup by the constant Boussinesq reference density, ρ_0 .

52.3.1 A three-layer thought experiment

To illustrate the concepts of transformation and formation, bin the World Ocean into density classes, thus ignoring all geographic and depth information. Furthermore, assume only three density classes bounded by four density interfaces:

$$\text{light density layer} = [\gamma - \delta\gamma/2, \gamma + \delta\gamma/2] \quad (52.2a)$$

$$\text{middle density layer} = [\gamma + \delta\gamma/2, \gamma + 3\delta\gamma/2] \quad (52.2b)$$

$$\text{heavy density layer} = [\gamma + 3\delta\gamma/2, \gamma + 5\delta\gamma/2], \quad (52.2c)$$

where $\delta\gamma > 0$ is the size for the density bins. Figure 52.2 depicts a sample mass distribution; i.e., the mass census for seawater binned into these three density layers.⁷ Now introduce a process that results in water leaving the middle density layer and entering both the light layer and the heavy layer. Let $G(\sigma)$ measure the mass per time that water crosses the density interface $\gamma = \sigma$; i.e., $G(\sigma)$ is the transformation. This particular thought experiment has the following transformations

⁶Many researchers make use of the *neutral density* coordinate defined by [Jackett and McDougall \(1997\)](#).

⁷A realistic ocean experiences boundary forcing that makes the maximum and minimum density a function of time. It is thus common to fix the lower density limit to be well below the lightest water in the ocean and the upper density limit well above the maximum density, thus ensuring that all seawater is contained by the chosen binning. We introduce such “infinity” bounds in Section 52.4.3.

across the various layer interfaces

$$G(\sigma) = \begin{cases} 0 & \sigma = \gamma - \delta\gamma/2 \\ < 0 & \sigma = \gamma + \delta\gamma/2 \\ > 0 & \sigma = \gamma + 3\delta\gamma/2 \\ 0 & \sigma = \gamma + 5\delta\gamma/2 \end{cases} \quad \begin{array}{l} \text{closed boundary} \\ \text{mass moves to light density from middle density} \\ \text{mass moves from middle density to heavy density} \\ \text{closed boundary.} \end{array} \quad (52.3)$$

The difference in the transformation across the interface boundaries of a particular layer determines the formation/destruction of water into that layer. Here, the convergence of water into the light and heavy layers means that there is a positive formation of water in these two density layers. In contrast, the divergence of water from the middle density layer means there is a negative formation or a destruction of some of its water. We write these layer formations mathematically as follows

$$\text{light-formation} = -[G(\gamma + \delta\gamma/2) - G(\gamma - \delta\gamma/2)] > 0 \quad (52.4a)$$

$$\text{middle-formation} = -[G(\gamma + 3\delta\gamma/2) - G(\gamma + \delta\gamma/2)] < 0 \quad (52.4b)$$

$$\text{heavy-formation} = -[G(\gamma + 5\delta\gamma/2) - G(\gamma + 3\delta\gamma/2)] > 0. \quad (52.4c)$$

The minus sign out front emphasizes that the formation is the layer integrated *convergence* of the transformation.

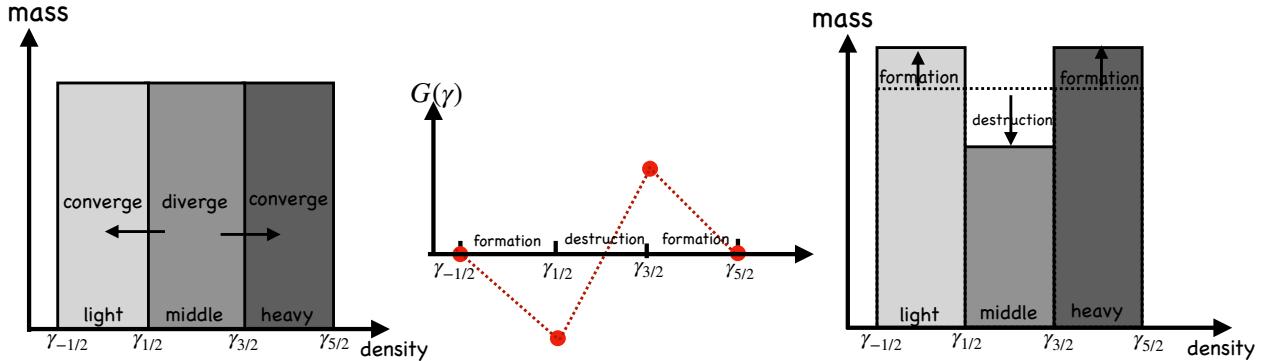


FIGURE 52.2: A sample mass distribution of the ocean binned into density layers (light, middle, heavy) bounded by the four density interfaces: $\gamma_{n/2} = \gamma + n\delta\gamma/2$ for $n = -1, 1, 3, 5$. The left panel depicts an ocean state with equal mass in each layer. Some process is then imagined to cause water to diverge from the middle layer and converge to the light and heavy layers. The right panel shows the mass distribution after the water has moved, so that the middle layer has experienced a negative *formation* (i.e., net loss of mass) whereas the light and heavy layers have experienced a positive *formation* (i.e., net mass gain). The middle panel depicts the *transformation*, G , which measures the mass per time moving across the layer boundaries. By convention, $G > 0$ for water moving into a heavier density layer and $G < 0$ for water entering a lighter density layer. The addition of more layers refines the picture (e.g., by smoothing the plot of G) but it does not modify the basic ideas illustrated in this thought experiment.

52.3.2 How processes lead to transformations

The central focus of water mass analysis is the movement of water between layers or classes, with this movement modifying the water mass distribution. Here we outline a few of the processes that affect this movement within water mass space, again focusing on buoyancy yet with an easy generalization to any other property that defines the water mass. Notably, we are not concerned with whether the fluid element moves, the boundary moves, or both, since it is only the relative motion that changes the water mass distribution. Indeed, we cannot determine motion of the fluid element without direct information about the velocity field. We return to this point when providing a mathematical expression for these ideas in Section 52.5.1.

Interior transformation from mixing

Mixing moves water across layer boundaries, with properties materially modified in the presence of mixing (so long as there are spatial gradients in the property). For example, recall our discussion in Chapter 16 where we saw that mixing causes tracers to move between fluid elements even as mixing does not alter the net mass of fluid elements (see the discussion of barycentric velocity in Section 16.1.2). Hence, in the presence of mixing, seawater fluid elements retain a fixed mass and yet the mass is redistributed among layers defined by property isosurfaces since the isosurfaces move in the presence of mixing.

Surface mass fluxes

Rain and evaporation alter the mass of the ocean. In turn, the layers where rain and evaporation occur; i.e., layers that outcrop, will have their mass altered. Additionally, if the buoyancy of the mass flux differs from that of the ocean layer that it enters/leaves, then the buoyancy of the ocean layer is modified.

Surface and bottom boundary transformation

Buoyancy surfaces that outcrop at the ocean surface or incrop at the ocean bottom are exposed to strong boundary fluxes that generally modify the buoyancy of the fluid within a layer. This modification in turn causes the layer boundaries to move so that the mass distributed within the layers is modified. A particularly striking example occurs in the upper ocean boundary layer where surface forcing leads to the seasonal migration of density outcrops. The associated lateral movement of density layers causes water to entrain and detrain from a layer since the layer boundaries can move faster than fluid elements. In so doing, the seasonal cycle of surface buoyancy forcing can inflate or deflate a buoyancy layer by moving the layer boundaries so that the layer entrains or detrains mass.

Penetrative shortwave radiation provides another means to modify water masses, with penetrative radiation a function of the optical properties of the fluid. This radiation provides a source of heating that can penetrate into the upper few tens of meters in the ocean, and can impact on the temperature and density structure of the ocean layers affected by radiation.

Finally, we note that layers that intersect the ocean bottom are exposed to geothermal heating as well as enhanced mixing from bottom boundary layer mixing. Each of these processes affects a transformation of the buoyancy, thus modifying the buoyancy layer interfaces and mass distribution within the layers.

Interior sources and sinks

When studying water masses defined by biogeochemical tracers (e.g., carbon, oxygen, nutrients), there are a variety of chemical reactions and biological processes that act to modify these properties. These processes generally cannot be represented mathematically as the convergence of a flux. They are thus sometimes referred to as “non-conservative” processes (see Section 22.11).

52.4 Mathematical framework

In this section we develop a suite of mathematical tools of use to quantify the conceptual ideas presented in Sections 52.2 and 52.3. In particular, we develop a formalism for integrating properties within a region bounded by isosurfaces of a scalar field, $\lambda = \lambda(\mathbf{x}, t)$. The formulation is given from

both a geometric perspective afforded by geographic/depth space, and a complementary distributional perspective afforded by binning seawater mass according to λ -classes. As in our study of Eulerian and Lagrangian kinematics elsewhere in this book, it is here useful to be adept at both the geographic/depth space perspective and the distributional perspective.

In Section 52.3 we assumed λ is the buoyancy field, $\lambda = \gamma$, whereas here we assume it is a generic scalar field, $\lambda(\mathbf{x}, t)$. In contrast to the case of a generalized vertical coordinate (Chapters 43 and 44), we make no assumption regarding the stratification of λ . Rather, λ -isosurfaces are free to overturn or even to be situated in spatially disconnected regions. This freedom is motivated by the behavior of most oceanographic scalar properties, which commonly exhibit vertically unstratified or negatively stratified profiles, particularly within boundary layers.

52.4.1 Seawater mass in an infinitesimal cylinder

Consider the calculation of seawater mass within an infinitesimal λ -layer bounded by two isosurfaces, $[\lambda - \delta\lambda/2, \lambda + \delta\lambda/2]$, as in Figure 52.3. The mass within a tiny cylinder extending from one interface to the other is given by the seawater density, ρ , multiplied by the volume of the cylinder,⁸

$$\delta M = \rho \delta V = \rho \delta h \delta \mathcal{S}, \quad (52.5)$$

where $\delta \mathcal{S}$ is the cross-sectional area element and δh is the layer thickness. The geometric thickness, δh , is related to the differential λ -increment separating the two interfaces according to

$$\delta \lambda = \nabla \lambda \cdot \delta \mathbf{x} = |\nabla \lambda| \hat{\mathbf{n}} \cdot \delta \mathbf{x} = |\nabla \lambda| \delta h \quad \text{with} \quad \hat{\mathbf{n}} = \nabla \lambda |\nabla \lambda|^{-1}, \quad (52.6)$$

where $\delta \mathbf{x}$ is a position vector connecting points on the two interfaces. We thus see that the layer thickness is given by

$$\delta h = \frac{\delta \lambda}{|\nabla \lambda|}, \quad (52.7)$$

which connects a geometric property of the layer, δh , to the λ -increment, $\delta \lambda > 0$. For a given λ -increment, the layer thickness is smaller with more tightly packed λ -isosurfaces as reflected by a larger $|\nabla \lambda|$. Furthermore, the geometric thickness is oriented according to the normal direction, $\hat{\mathbf{n}}$, so that δh measures the distance between the λ -interfaces in the direction of the normal direction. It follows that the seawater mass within the cylinder is given by

$$\delta M = \rho \delta V = \rho \delta h \delta \mathcal{S} = \frac{\rho \delta \lambda \delta \mathcal{S}}{|\nabla \lambda|}. \quad (52.8)$$

52.4.2 Seawater mass within a finite region

Making use of the infinitesimal cylinder mass (52.8) allows us to write the mass of seawater within the λ -region $\lambda_1 \leq \lambda \leq \lambda_2$

$$M(\lambda_1, \lambda_2) \equiv \int_{\Omega(\lambda_1 \leq \lambda \leq \lambda_2)} dM = \int_{\lambda_1}^{\lambda_2} \left[\int_{\partial\Omega(\lambda)} \frac{\rho d\mathcal{S}}{|\nabla \lambda|} \right] d\lambda. \quad (52.9)$$

In this equation, $\Omega(\lambda_1 \leq \lambda \leq \lambda_2)$ is the region in space bounded by the λ_1 -interface and λ_2 -interface, and $\partial\Omega(\lambda)$ is the surface defined by a λ -isosurface. The $\partial\Omega(\lambda)$ integral is taken over the area of the λ -isosurface, which is then integrated over the range, $\lambda_1 \leq \lambda \leq \lambda_2$, to thus accumulate the layer mass.

⁸Recall our notational convention is as follows: δ refers to an infinitesimal increment of a property measured within the fluid whereas d is a differential increment used for computing integrals. We made use of the same geometric analysis in Section 38.1.2 when studying potential vorticity.

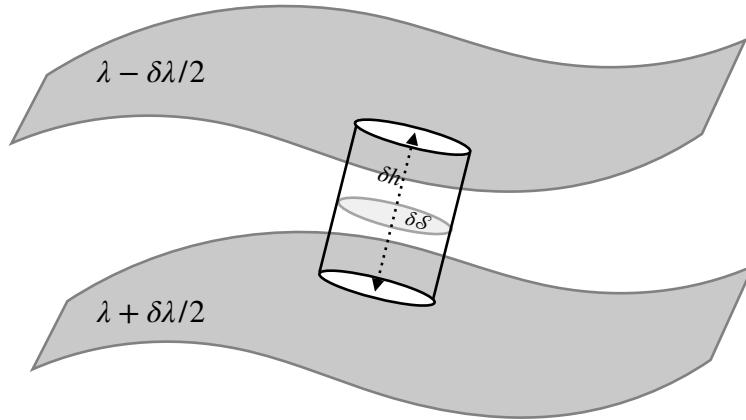


FIGURE 52.3: This schematic shows an infinitesimally thin λ -layer bounded by two interfaces $[\lambda - \delta\lambda/2, \lambda + \delta\lambda/2]$, with the λ -increment $\delta\lambda > 0$. The cylinder region extends between the two iso-surfaces and it has thickness $\delta h = \delta\lambda/|\nabla\lambda|$ and cross-sectional area δS . The cylinder is oriented according to the normal direction $\hat{n} = |\nabla\lambda|^{-1} \nabla\lambda$. We assume $|\nabla\lambda| \neq 0$, as required to define a normal direction. Indeed, if $\nabla\lambda = 0$ then we could not perform a binning according to λ classes, so the $|\nabla\lambda| \neq 0$ assumption is basic to the use of the scalar field, λ , for water mass analysis.

52.4.3 Seawater mass distribution/density function

The region bounded by the layer interfaces can have any shape in space and can even be spatially disconnected. This complexity motivates us to introduce the *mass distribution* or *mass density* function by integrating the mass over the surface, $\partial\Omega(\lambda)$

$$m(\lambda) \equiv \frac{dM}{d\lambda} = \int_{\partial\Omega(\lambda)} \frac{\rho dS}{|\nabla\lambda|}. \quad (52.10)$$

A mass distribution function is quite useful when the distribution is highly non-local in geographic/depth space. It is also effective when using multiple water mass coordinates such as Θ and S (Section 52.12). These points motivate leaving geographic/depth perspective to simply define the mass distribution function so that

$$dM = m(\lambda) d\lambda = \text{mass within the infinitesimal } \lambda\text{-layer } [\lambda - d\lambda/2, \lambda + d\lambda/2], \quad (52.11)$$

with an illustration given by Figure 52.4. The mass distribution function is the mass density within λ space; i.e., the mass per λ . Knowledge of the mass distribution function allows us to compute the seawater mass within a finite region, as in equation (52.9), according to

$$M(\lambda_1, \lambda_2) \equiv \int_{\Omega(\lambda_1 \leq \lambda \leq \lambda_2)} dM = \int_{\lambda_1}^{\lambda_2} m(\lambda) d\lambda. \quad (52.12)$$

52.4.4 Example regions

To help ground the previous expressions for mass, we here consider some example regions that are commonly considered in water mass analysis.

Δλ-layer defined by $[\lambda_1, \lambda_2] = [\lambda - \Delta\lambda/2, \lambda + \Delta\lambda/2]$

A $\Delta\lambda$ -layer is defined with the bounding interface values

$$\lambda_1 = \lambda - \Delta\lambda/2 \text{ and } \lambda_2 = \lambda + \Delta\lambda/2, \quad (52.13)$$

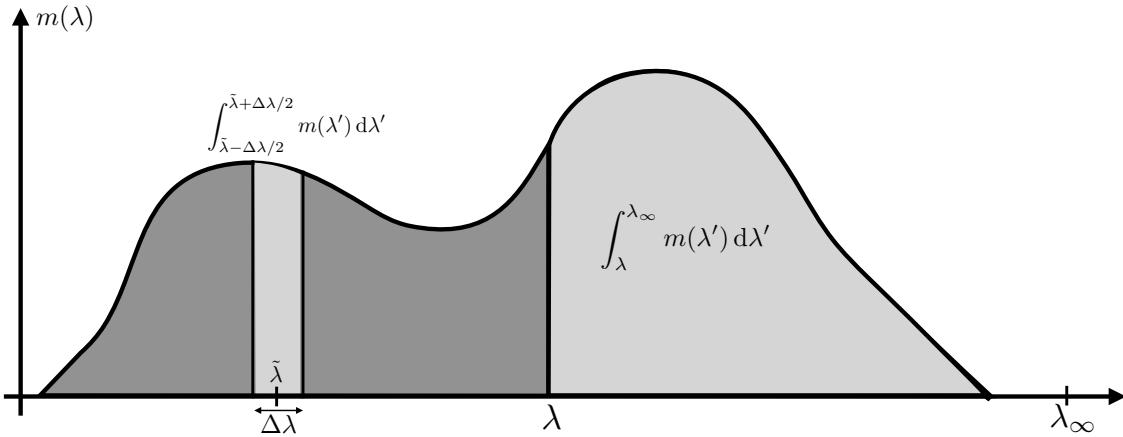


FIGURE 52.4: An example mass distribution function, $m(\lambda) = dM/d\lambda$, which measures the mass of seawater per λ -increment. Integration over a finite λ -region measures the seawater mass within that region. For example, the mass within a $\Delta\lambda$ -layer is given by $M(\tilde{\lambda} - \Delta\lambda/2, \tilde{\lambda} + \Delta\lambda/2) = \int_{\tilde{\lambda} - \Delta\lambda/2}^{\tilde{\lambda} + \Delta\lambda/2} m(\lambda') d\lambda'$ whereas the mass within the λ_∞ -region is $M(\lambda, \lambda_\infty) = \int_\lambda^{\lambda_\infty} m(\lambda') d\lambda'$, where we assume that λ_∞ is an arbitrary fixed value that is larger than any λ realized within the ocean.

for some finite difference increment $\Delta\lambda > 0$. In this case the layer mass is

$$M(\lambda - \Delta\lambda/2, \lambda + \Delta\lambda/2) = \int_{\lambda - \Delta\lambda/2}^{\lambda + \Delta\lambda/2} \left[\int_{\partial\Omega(\lambda')} \frac{\rho dS}{|\nabla \lambda'|} \right] d\lambda' = \int_{\lambda - \Delta\lambda/2}^{\lambda + \Delta\lambda/2} m(\lambda') d\lambda'. \quad (52.14)$$

Characterizing ocean properties according to their value of λ is generally performed by decomposing the ocean into $\Delta\lambda$ -bins and forming histograms to estimate the continuous distribution.

λ_∞ -region defined by $[\lambda_1, \lambda_2] = [\lambda, \lambda_\infty]$

A λ_∞ -region is defined with

$$\lambda_1 = \lambda \text{ and } \lambda_2 = \lambda_\infty, \quad (52.15)$$

where λ_∞ is an arbitrary fixed constant that is larger than any value of λ realized in the ocean. The region mass is thus given by

$$M(\lambda, \lambda_\infty) = \int_{\Omega(\lambda \leq \lambda_\infty)} dM = \int_\lambda^{\lambda_\infty} \left[\int_{\partial\Omega(\lambda')} \frac{\rho dS}{|\nabla \lambda'|} \right] d\lambda' = \int_\lambda^{\lambda_\infty} m(\lambda') d\lambda'. \quad (52.16)$$

An example λ_∞ -region is shown in Figure 52.5. The λ_∞ -region as so defined provides an expression for the differential mass increment

$$M(\lambda, \lambda_\infty) = \int_\lambda^{\lambda_\infty} m(\lambda') d\lambda' \implies dM(\lambda, \lambda_\infty) = -m(\lambda) d\lambda, \quad (52.17)$$

which follows since λ_∞ is a constant.

The value of the fixed constant, λ_∞ , is arbitrarily large, indeed it could be infinite. We can set it to an arbitrarily large constant value since there is no contribution to the integral from regions with λ' outside the range realized within the ocean, merely since there is no ocean mass in that region. As an example, let $\lambda = \Theta$, the Conservative Temperature, in which the region $\Theta \leq \Theta_\infty$ encompasses the ocean region where the Conservative Temperature is larger (warmer) than Θ .

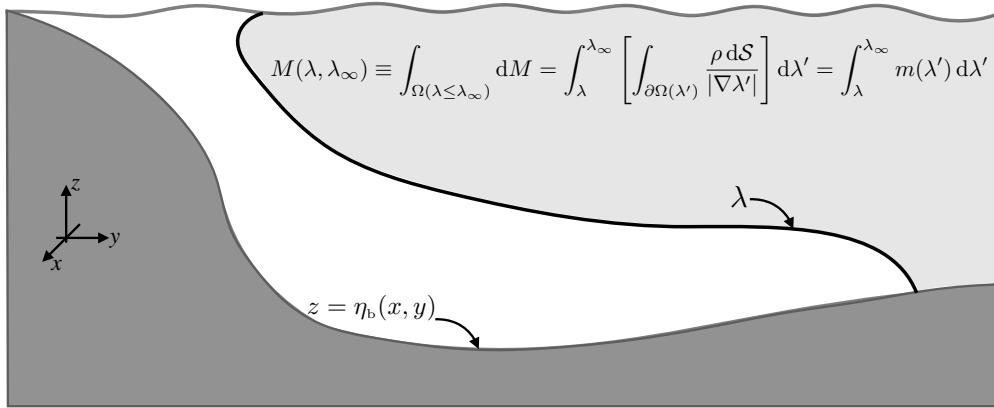


FIGURE 52.5: Depicting the mass of fluid within a λ_∞ -region, where $\lambda \leq \lambda_\infty$ with λ_∞ is an arbitrary constant that is larger than any value of λ in the ocean domain. This figure is oriented for the southern hemisphere with Antarctica on the left. An example of such a region is for $\lambda = \Theta$, whereby warmer waters are typically shallower and towards the equator.

$\lambda_{-\infty}$ -region defined by $[\lambda_1, \lambda_2] = [\lambda_{-\infty}, \lambda]$

A $\lambda_{-\infty}$ -region is defined with

$$\lambda_1 = \lambda_{-\infty} \text{ and } \lambda = \lambda, \quad (52.18)$$

where $\lambda_{-\infty}$ is an arbitrary constant that is smaller than any value of λ realized in the ocean. The region mass is thus given by

$$M(\lambda_{-\infty}, \lambda) \equiv \int_{\Omega(\lambda_{-\infty} \leq \lambda)} dM = \int_{\lambda_{-\infty}}^{\lambda} \left[\int_{\partial\Omega(\lambda')} \frac{\rho dS}{|\nabla \lambda'|} \right] d\lambda' = \int_{\lambda_{-\infty}}^{\lambda} m(\lambda') d\lambda'. \quad (52.19)$$

This mass is the complement of that contained in the λ_∞ -region. The $\lambda_{-\infty}$ -region mass implies a corresponding differential mass increment via

$$M(\lambda_{-\infty}, \lambda) = \int_{\lambda_{-\infty}}^{\lambda} m(\lambda') d\lambda' \implies dM(\lambda_{-\infty}, \lambda) = m(\lambda) d\lambda. \quad (52.20)$$

$\lambda_{\pm\infty}$ -region defined by $[\lambda_1, \lambda_2] = [\lambda_{-\infty}, \lambda_\infty]$

The full ocean is contained in the $\lambda_{\pm\infty}$ -region

$$\lambda_1 = \lambda_{-\infty} \text{ and } \lambda_2 = \lambda_\infty, \quad (52.21)$$

so that the full ocean mass is written

$$M(\lambda_{-\infty}, \lambda_\infty) = \int_{\Omega(\lambda_{-\infty} \leq \lambda \leq \lambda_\infty)} dM = \int_{\lambda_{-\infty}}^{\lambda_\infty} \left[\int_{\partial\Omega(\lambda)} \frac{\rho dS}{|\nabla \lambda|} \right] d\lambda = \int_{\lambda_{-\infty}}^{\lambda_\infty} m(\lambda') d\lambda'. \quad (52.22)$$

Difference of mass between two λ_∞ -regions

The difference in mass between two λ_∞ -regions is given by

$$M(\lambda_1, \lambda_\infty) - M(\lambda_2, \lambda_\infty) = \int_{\lambda_1}^{\lambda_\infty} \left[\int_{\partial\Omega(\lambda)} \frac{\rho dS}{|\nabla \lambda|} \right] d\lambda - \int_{\lambda_2}^{\lambda_\infty} \left[\int_{\partial\Omega(\lambda)} \frac{\rho dS}{|\nabla \lambda|} \right] d\lambda. \quad (52.23)$$

The arbitrary constant, λ_∞ , drops out when taking the difference so that we are left with the mass within the intersection of the two regions

$$M(\lambda_1, \lambda_2) = M(\lambda_1, \lambda_\infty) - M(\lambda_2, \lambda_\infty) = \int_{\lambda_1}^{\lambda_2} \left[\int_{\partial\Omega(\lambda)} \frac{\rho d\mathcal{S}}{|\nabla\lambda|} \right] d\lambda = \int_{\lambda_1}^{\lambda_2} m(\lambda') d\lambda'. \quad (52.24)$$

52.4.5 Integrals of arbitrary functions

We can extend the above formalism to integrals of an arbitrary function, $\mathcal{F}(\mathbf{x}, t)$, over a region defined by $\lambda_{1,2}$ -interfaces

$$\mathcal{G}_{\mathcal{F}}(\lambda_1, \lambda_2) \equiv \int_{\Omega(\lambda_1 \leq \lambda \leq \lambda_2)} \mathcal{F} dM = \int_{\lambda_1}^{\lambda_2} \left[\int_{\partial\Omega(\lambda)} \frac{\mathcal{F} \rho d\mathcal{S}}{|\nabla\lambda|} \right] d\lambda. \quad (52.25)$$

Performing the area integral amounts to binning the function according to λ -increments, in which case we define the distribution function

$$m_{\mathcal{F}}(\lambda) = \int_{\partial\Omega(\lambda)} \frac{\mathcal{F} \rho d\mathcal{S}}{|\nabla\lambda|}, \quad (52.26)$$

so that an integral over the distribution is given by

$$\mathcal{G}_{\mathcal{F}}(\lambda_1, \lambda_2) = \int_{\lambda_1}^{\lambda_2} m_{\mathcal{F}}(\lambda) d\lambda. \quad (52.27)$$

In particular, consider the integral over a λ_∞ -region

$$\mathcal{G}_{\mathcal{F}}(\lambda, \lambda_\infty) = \int_{\lambda}^{\lambda_\infty} \left[\int_{\partial\Omega(\lambda')} \frac{\mathcal{F} \rho d\mathcal{S}}{|\nabla\lambda'|} \right] d\lambda' = \int_{\lambda}^{\lambda_\infty} m_{\mathcal{F}}(\lambda') d\lambda', \quad (52.28)$$

which has the derivative

$$\frac{\partial \mathcal{G}_{\mathcal{F}}(\lambda, \lambda_\infty)}{\partial \lambda} = - \int_{\partial\Omega(\lambda)} \frac{\mathcal{F} \rho d\mathcal{S}}{|\nabla\lambda|} = -m_{\mathcal{F}}(\lambda), \quad (52.29)$$

as follows from the fundamental theorem of calculus. Note how the derivative removes the reference value, λ_∞ . Analogously, the integral over a $\lambda_{-\infty}$ -region has the derivative

$$\frac{\partial \mathcal{G}_{\mathcal{F}}(\lambda_{-\infty}, \lambda)}{\partial \lambda} = \int_{\partial\Omega(\lambda)} \frac{\mathcal{F} \rho d\mathcal{S}}{|\nabla\lambda|} = m_{\mathcal{F}}(\lambda). \quad (52.30)$$

52.4.6 Moments of λ

Setting $\mathcal{F} = \lambda$ in the integral (52.25) renders

$$\Lambda(\lambda_1, \lambda_2) \equiv \int_{\Omega(\lambda_1 \leq \lambda \leq \lambda_2)} \lambda dM = \int_{\Omega(\lambda_1 \leq \lambda \leq \lambda_2)} \lambda \rho dV = \int_{\lambda_1}^{\lambda_2} \left[\int_{\partial\Omega(\lambda)} \frac{\lambda \rho d\mathcal{S}}{|\nabla\lambda|} \right] d\lambda. \quad (52.31)$$

If λ is a tracer concentration (tracer mass per seawater mass), then $\Lambda(\lambda_1, \lambda_2)$ is the mass of tracer within the layer. Observe that λ can be pulled outside of the surface integral in equation (52.31) since λ is constant along $\partial\Omega(\lambda)$, thus rendering

$$\Lambda(\lambda_1, \lambda_2) = \int_{\lambda_1}^{\lambda_2} \left[\int_{\partial\Omega(\lambda)} \frac{\rho d\mathcal{S}}{|\nabla\lambda|} \right] \lambda d\lambda = \int_{\lambda_1}^{\lambda_2} m(\lambda) \lambda d\lambda. \quad (52.32)$$

The final expression reveals the first moment of λ as measured by the seawater mass distribution function. We can likewise define the n 'th moment as

$$\Lambda^{(n)}(\lambda_1, \lambda_2) \equiv \int_{\lambda_1}^{\lambda_2} \left[\int_{\partial\Omega(\lambda)} \frac{\rho dS}{|\nabla \lambda|} \right] \lambda^n d\lambda = \int_{\lambda_1}^{\lambda_2} m(\lambda) \lambda^n d\lambda = M(\lambda_1, \lambda_2) \langle \lambda^n \rangle. \quad (52.33)$$

The final equality introduced the mean value for the moment

$$\langle \lambda^n \rangle = \frac{\int_{\lambda_1}^{\lambda_2} m(\lambda) \lambda^n d\lambda}{\int_{\lambda_1}^{\lambda_2} m(\lambda) d\lambda} \quad (52.34)$$

as defined over the $[\lambda_1, \lambda_2]$ region.

52.4.7 Internal and external λ -moments

Now specify the region $[\lambda_1, \lambda_2] = [\tilde{\lambda}, \lambda_\infty]$ for the moment equation (52.33). Making use of the differential mass increment, $dM(\lambda, \lambda_\infty) = -m(\lambda) d\lambda$ as in equation (52.17) allows us to integrate the moment equation by parts

$$\Lambda^{(n)}(\tilde{\lambda}, \lambda_\infty) = \int_{\tilde{\lambda}}^{\lambda_\infty} \lambda^n m(\lambda) d\lambda \quad (52.35a)$$

$$= - \int_{\tilde{\lambda}}^{\lambda_\infty} \lambda^n dM \quad (52.35b)$$

$$= \int_{\tilde{\lambda}}^{\lambda_\infty} [-d(\lambda^n M) + n M \lambda^{n-1} d\lambda] \quad (52.35c)$$

$$= -\lambda_\infty^n M(\lambda_\infty, \lambda_\infty) + \lambda^n M(\tilde{\lambda}, \lambda_\infty) + n \int_{\tilde{\lambda}}^{\lambda_\infty} M(\lambda, \lambda_\infty) \lambda^{n-1} d\lambda \quad (52.35d)$$

$$= \lambda^n M(\tilde{\lambda}, \lambda_\infty) + n \int_{\tilde{\lambda}}^{\lambda_\infty} M(\lambda, \lambda_\infty) \lambda^{n-1} d\lambda, \quad (52.35e)$$

where the final equality follows since $M(\lambda_\infty, \lambda_\infty) = 0$. Making use of equation (52.33) thus leads to

$$M(\tilde{\lambda}, \lambda_\infty) \langle \lambda^n \rangle = \underbrace{M(\tilde{\lambda}, \lambda_\infty) \tilde{\lambda}^n}_{\text{external moment}} + \underbrace{n \int_{\tilde{\lambda}}^{\lambda_\infty} M(\lambda, \lambda_\infty) \lambda^{n-1} d\lambda}_{\text{internal moment}}. \quad (52.36)$$

We refer to the rightmost term as the *internal moment* since it is an integral over the region, whereas $M(\tilde{\lambda}, \lambda_\infty) \lambda^n$ is the *external moment*, which is the region mass times the boundary value, $\tilde{\lambda}^n$. We choose the moniker “external” since the external moment increases in direct proportion to the mass crossing the ocean layer boundaries, including the external boundaries. In Section 52.7 we develop a budget for the $n = 1$ moment, in which the internal moment from equation (52.36) takes the form

$$M(\tilde{\lambda}, \lambda_\infty) [\langle \lambda \rangle - \tilde{\lambda}] = \int_{\tilde{\lambda}}^{\lambda_\infty} M(\lambda, \lambda_\infty) d\lambda. \quad (52.37)$$

We return to the notion of internal and external moments in Section 52.11.2.

52.4.8 Further study

The formulation given here in terms of mass distribution functions follows that of [Waln \(1977\)](#) and [Waln \(1982\)](#). In these two papers, Waln pioneered the formalism of water mass analysis, which is sometimes referred to as *Waln analysis* in his honor. The concept of internal and external tracer moments follows the interior and exterior heat introduced by [Holmes et al. \(2019\)](#).

52.5 Water mass transformation across a λ -surface

We here develop the formalism to quantify transport of seawater crossing an interior λ -interface. This transport is referred to as the *water mass transformation* and is written as $G(\lambda)$. Figure 52.6 illustrates how this transformation appears in a mass budget for a $\Delta\lambda$ -layer.

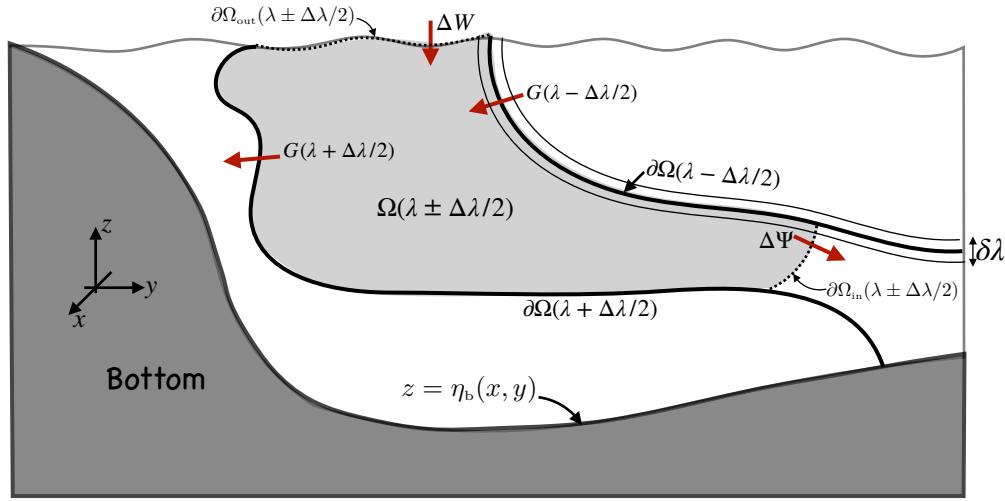


FIGURE 52.6: A layer of seawater with scalar property λ within the range $[\lambda - \Delta\lambda/2, \lambda + \Delta\lambda/2]$ and defined over a geographical domain $\Omega(\lambda \pm \Delta\lambda/2)$. In this example, λ increases to the south, towards Antarctica, with $\lambda = \gamma$ (Section 52.3) an example. The net seawater mass transport crossing the layer interfaces $\partial\Omega(\lambda \pm \Delta\lambda/2)$ is $G(\lambda \pm \Delta\lambda/2)$, with $G > 0$ for water moving to regions of larger λ . The seawater mass crossing the layer through the geographical bounds $\partial\Omega_{in}(\lambda \pm \Delta\lambda/2)$ is written $\Delta\Psi(\lambda \pm \Delta\lambda/2)$, with $\Delta\Psi(\lambda \pm \Delta\lambda/2) > 0$ for water leaving $\Omega(\lambda \pm \Delta\lambda/2)$. The boundary $\partial\Omega_{in}(\lambda \pm \Delta\lambda/2)$ is absent when the domain extends across a basin or the global ocean (e.g., see Figure 52.8). The mass crossing the sea surface, $\partial\Omega_{out}(\lambda \pm \Delta\lambda/2)$, through rain, evaporation, melt, or rivers is written $\Delta W(\lambda \pm \Delta\lambda/2)$, with $\Delta W(\lambda \pm \Delta\lambda/2) > 0$ for mass entering $\Omega(\lambda \pm \Delta\lambda/2)$. A layer interface can have an arbitrary stratification, such as the vertically non-monotonic profile depicted here for the $\lambda + \Delta\lambda/2$ interface. Additionally, the domain $\Omega(\lambda \pm \Delta\lambda/2)$ can generally be disconnected. The net domain boundaries are written $\partial\Omega_{in}(\lambda \pm \Delta\lambda/2) + \partial\Omega_{out}(\lambda \pm \Delta\lambda/2) + \partial\Omega(\lambda + \Delta\lambda/2) + \partial\Omega(\lambda - \Delta\lambda/2)$. The $\delta\lambda$ layer surrounding the $\partial\Omega(\lambda - \Delta\lambda/2)$ arises as part of the method detailed in Section 52.5.2 for computing G according to the λ -derivative of an integral over the $\delta\lambda$ layer.

52.5.1 Dia-surface flux and interior transformation

The object that measures the local water mass transformation is the dia-surface flux detailed in Section 44.4.7. This flux is given by

$$w^{\text{dia}} = \hat{\mathbf{n}} \cdot (\mathbf{v} - \mathbf{v}^{(\lambda)}) = \frac{\dot{\lambda}}{|\nabla\lambda|} \quad \text{with} \quad \hat{\mathbf{n}} = \frac{\nabla\lambda}{|\nabla\lambda|} \quad \text{and} \quad \dot{\lambda} = \frac{D\lambda}{Dt}, \quad (52.38)$$

with $w^{\text{dia}} > 0$ for water moving to regions of larger λ . It is computed as the projection of the relative velocity, $(\mathbf{v} - \mathbf{v}^{(\lambda)})$, onto the direction normal to the surface, with the relative velocity being the

difference between the fluid particle velocity, \mathbf{v} , and the velocity of a point on the λ -interface, $\mathbf{v}^{(\lambda)}$. The velocity of a point on the surface, $\mathbf{v}^{(\lambda)}$, satisfies the following kinematic constraint⁹

$$(\partial_t + \mathbf{v}^{(\lambda)} \cdot \nabla) \lambda = 0. \quad (52.39)$$

This constraint is based on assuming $\mathbf{v}^{(\lambda)}$ measures the velocity of a point fixed to the λ -surface. So in brief, the dia-surface flux, w^{dia} , locally measures the flux of seawater (volume per area per time) that penetrates a λ -interface in the direction of increasing λ .

The interior water mass transformation, $G(\lambda)$, is the area integral of ρw^{dia} over the full extent of the λ -surface

$$G(\lambda) \equiv \int_{\partial\Omega(\lambda)} \rho w^{\text{dia}} d\mathcal{S} = \int_{\partial\Omega(\lambda)} \rho \hat{\mathbf{n}} \cdot (\mathbf{v} - \mathbf{v}^{(\lambda)}) d\mathcal{S} = \int_{\partial\Omega(\lambda)} \frac{\rho \dot{\lambda}}{|\nabla \lambda|} d\mathcal{S}, \quad (52.40)$$

where $\partial\Omega(\lambda)$ is the surface occupied by the λ -interface (see Figure 52.6). Furthermore, the dimensions of $G(\lambda)$ are mass per time

$$G(\lambda) \quad [\equiv] \quad \text{M T}^{-1}, \quad (52.41)$$

thus measuring the mass per time crossing the λ -interface.

Based on the definition (52.40), we see that interior water mass transformation across a λ -interface occurs when there is a material change, $\dot{\lambda} \neq 0$, in the property defining the interface. Interior material changes arise from mixing, which generally causes irreversible changes to λ , thus driving seawater across the moving λ -interfaces. Material changes can also arise from sources and sinks, as when considering buoyancy surfaces in the presence of a nonlinear equation of state (Chapter 51). Sources and sinks are also commonly encountered by biogeochemical tracers.

52.5.2 Transformation as the derivative of an integral

Following the discussion from Section 52.4.5, we set $\mathcal{F} = \dot{\lambda}$ and consider the mass integral

$$\mathcal{I}_{\dot{\lambda}}(\lambda, \lambda_\infty) = \int_{\Omega(\lambda \leq \lambda_\infty)} \dot{\lambda}' dM = \int_{\lambda}^{\lambda_\infty} \left[\int_{\partial\Omega(\lambda')} \frac{\rho \dot{\lambda}'}{|\nabla \lambda'|} d\mathcal{S} \right] d\lambda' = \int_{\lambda}^{\lambda_\infty} G(\lambda') d\lambda'. \quad (52.42)$$

The fundamental theorem of calculus leads to the expression of the water mass transformation as the derivative

$$G(\lambda) = -\frac{\partial \mathcal{I}_{\dot{\lambda}}(\lambda, \lambda_\infty)}{\partial \lambda} \quad (52.43a)$$

$$= -\lim_{\delta\lambda \rightarrow 0} \frac{1}{\delta\lambda} \left[\int_{\lambda+\delta\lambda/2}^{\lambda_\infty} G(\lambda') d\lambda' - \int_{\lambda-\delta\lambda/2}^{\lambda_\infty} G(\lambda') d\lambda' \right] \quad (52.43b)$$

$$= \lim_{\delta\lambda \rightarrow 0} \frac{1}{\delta\lambda} \int_{\lambda-\delta\lambda/2}^{\lambda+\delta\lambda/2} G(\lambda') d\lambda' \quad (52.43c)$$

$$= \lim_{\delta\lambda \rightarrow 0} \frac{1}{\delta\lambda} \int_{\Omega(\lambda \pm \delta\lambda/2)} \rho \dot{\lambda}' dV. \quad (52.43d)$$

We conclude that calculation of interior water mass transformation requires the various contributions to the material time change, $\dot{\lambda}$, a weighting of the time changes according to the mass of the

⁹We encountered the relation (52.39) in Section 15.4.2 when deriving the kinematic boundary condition for a moving surface.

fluid element, $dM = \rho dV$, and a binning of $\rho \dot{\lambda} dV$ according to λ -classes. Note that in the final equality, equation (52.43d), the limit $\delta\lambda \rightarrow 0$ might appear to lead to a singularity. However, as seen by the form in equation (52.43c), the integration region volume also gets smaller as $\delta\lambda \rightarrow 0$ so that the limit is well defined.

52.5.3 Processes and kinematics

There are two equivalent and complementary methods to view interior water mass transformation: the process method and the kinematic method.

Process method

The *process method* focuses on physical processes leading to movement of fluid across the λ -interface, thus providing information concerning *why* interior transformation occurs. It does so by binning processes contributing to the right hand side of the equation

$$\rho \dot{\lambda} = -\nabla \cdot \mathbf{J} + \rho \dot{\Upsilon}, \quad (52.44)$$

where \mathbf{J} is a flux arising from non-advection processes such as diffusion, and $\dot{\Upsilon}$ is a source/sink term (dimensions of λ per time) that cannot be written as the convergence of a flux. The tracer equation (52.44) inserted into the transformation equation (52.43d) leads to

$$G(\lambda) = \lim_{\delta\lambda \rightarrow 0} \frac{1}{\delta\lambda} \int_{\Omega(\lambda \pm \delta\lambda/2)} [-\nabla \cdot \mathbf{J} + \rho \dot{\Upsilon}] dV \quad (52.45a)$$

$$= \lim_{\delta\lambda \rightarrow 0} \frac{1}{\delta\lambda} \oint_{\partial\Omega(\lambda \pm \delta\lambda/2)} [-\mathbf{J} \cdot \hat{\mathbf{n}}] d\mathcal{S} + \lim_{\delta\lambda \rightarrow 0} \frac{1}{\delta\lambda} \int_{\Omega(\lambda \pm \delta\lambda/2)} \rho \dot{\Upsilon} dV, \quad (52.45b)$$

The second equality made use of the divergence theorem to convert the volume integral into a surface integral. For the tracer sources, we assume they do not modify the seawater mass at a point so that there is no source in the seawater mass equation.

For many purposes it is useful to decompose the non-advection flux divergence into contributions from interior processes, such as ocean mixing, and boundary fluxes

$$\nabla \cdot \mathbf{J} = \nabla \cdot [\mathbf{J}_{\text{int}} + \mathbf{J}_{\text{out}} + \mathbf{J}_{\text{bot}}]. \quad (52.46)$$

By definition,

$$\mathbf{J}_{\text{int}} \cdot \hat{\mathbf{n}} = 0 \quad \text{on } \partial\Omega_{\text{out}}(\lambda \pm \delta\lambda/2) \text{ and } \partial\Omega_{\text{bot}}(\lambda \pm \delta\lambda/2), \quad (52.47)$$

whereas $\mathbf{J}_{\text{int}} \cdot \hat{\mathbf{n}}$ is generally nonzero on interior layer boundaries. In contrast, the boundary fluxes, $\mathbf{J}_{\text{out}} \cdot \hat{\mathbf{n}}$ and $\mathbf{J}_{\text{bot}} \cdot \hat{\mathbf{n}}$, are identically zero everywhere except on their respective boundaries. Correspondingly, it is convenient to bin the volume weighted divergence, $\nabla \cdot \mathbf{J}_{\text{int}} dV$, according to λ -classes, and to likewise bin the area weighted boundary fluxes, $\mathbf{J}_{\text{out}} \cdot \hat{\mathbf{n}} d\mathcal{S}$ and $\mathbf{J}_{\text{bot}} \cdot \hat{\mathbf{n}} d\mathcal{S}$. In this

way we write the non-advective contribution to water mass transformation in the form

$$\begin{aligned} G(\lambda)_{\text{non-adv}} = & \underbrace{- \lim_{\delta\lambda \rightarrow 0} \frac{1}{\delta\lambda} \int_{\Omega(\lambda \pm \delta\lambda/2)} \nabla \cdot \mathbf{J}_{\text{int}} \, dV}_{\text{interior transformation} = \text{volume integral of convergence}} \\ & - \underbrace{\lim_{\delta\lambda \rightarrow 0} \frac{1}{\delta\lambda} \int_{\partial\Omega_{\text{out}}(\lambda \pm \delta\lambda/2)} \mathbf{J}_{\text{out}} \cdot \hat{\mathbf{n}} \, d\mathcal{S}}_{\text{surface transformation} = \text{area integral of surface boundary fluxes}} \\ & - \underbrace{\lim_{\delta\lambda \rightarrow 0} \frac{1}{\delta\lambda} \int_{\partial\Omega_{\text{bot}}(\lambda \pm \delta\lambda/2)} \mathbf{J}_{\text{bot}} \cdot \hat{\mathbf{n}} \, d\mathcal{S}}_{\text{bottom transformation} = \text{area integral of bottom boundary fluxes}}. \end{aligned} \quad (52.48)$$

Again, this expression decomposes the contribution from interior processes, here represented as the volume integral of the interior flux convergence, from the surface and bottom contributions, here represented as the area integral of the boundary fluxes. This decomposition is further examined in Section 52.8 where we focus on the surface contribution to water mass transformation.

Since the boundary fluxes are, by definition, zero except on the boundaries, their divergence can be written in terms of a Dirac delta distribution¹⁰

$$\nabla \cdot [\mathbf{J}_{\text{out}} + \mathbf{J}_{\text{bot}}] = \mathbf{J}_{\text{out}} \cdot \hat{\mathbf{n}} \delta(z - \eta) + \mathbf{J}_{\text{bot}} \cdot \hat{\mathbf{n}} \delta(z - \eta_b). \quad (52.49)$$

Although this equation is more physically formal than mathematically rigorous, its use in the transformation equation (52.45a) correctly leads to the expression (52.48). Consequently, equation (52.49) proves to be a useful shorthand.

Kinematic method

The *kinematic method* focuses on the kinematic means for realizing dia-surface transport, thus providing information concerning *how* interior transformation occurs. It does so by binning processes contributing to the right hand side of

$$\rho \dot{\lambda} = \partial_t(\rho \lambda) + \nabla \cdot (\rho \lambda \mathbf{v}), \quad (52.50)$$

which arises from the local time tendency plus advection, so that

$$G(\lambda) = \lim_{\delta\lambda \rightarrow 0} \frac{1}{\delta\lambda} \int_{\Omega(\lambda \pm \delta\lambda/2)} [\partial_t(\rho \lambda) + \nabla \cdot (\rho \lambda \mathbf{v})] \, dV. \quad (52.51)$$

Comments

As we saw, equality of the process method and kinematic method follows simply because the two provide equivalent expressions for the material time derivative. However, in the analysis of numerical model output, it can be nontrivial to realize this equivalence due to the extreme care required to diagnose the terms appearing in the scalar budget equation.

¹⁰We detail properties of the Dirac delta distribution in Section 4.2. It has dimensions of inverse the dimensions of its argument, so that when the argument has dimensions of length then the Dirac delta has dimensions of inverse length. We thus see that equation (52.49) is dimensionally consistent.

52.5.4 Some details concerning interior transformation

We here consider some details of the water mass transformation arising just from interior processes such as diffusion.

A global integrated constraint on $G(\lambda)_{\text{int}}$

Consider the integrated water mass transformation given by equation (52.42), only now integrate over the full ocean domain

$$\mathcal{G}_{\dot{\lambda}}(\lambda_{-\infty}, \lambda_{\infty}) = \int_{\lambda_{-\infty}}^{\lambda_{\infty}} G(\lambda') d\lambda' = \int_{\Omega(\lambda_{-\infty}, \lambda_{\infty})} \rho \dot{\lambda}' dV. \quad (52.52)$$

This integral vanishes for water mass transformations arising just from interior processes

$$[\mathcal{G}_{\dot{\lambda}}(\lambda_{-\infty}, \lambda_{\infty})]_{\text{int}} = \int_{\lambda_{-\infty}}^{\lambda_{\infty}} G(\lambda')_{\text{int}} d\lambda' \quad (52.53a)$$

$$= - \int_{\Omega(\lambda_{-\infty}, \lambda_{\infty})} \nabla \cdot \mathbf{J}_{\text{int}} dV \quad (52.53b)$$

$$= - \oint_{\partial\Omega(\lambda_{-\infty}, \lambda_{\infty})} \mathbf{J}_{\text{int}} \cdot \hat{\mathbf{n}} d\mathcal{S} \quad (52.53c)$$

$$= 0, \quad (52.53d)$$

which follows since $\mathbf{J}_{\text{int}} \cdot \hat{\mathbf{n}} = 0$ on the ocean boundaries. Hence, there can be no net water mass transformation from interior processes. Rather, interior processes only lead to rearrangement of water within the λ -bins. This result follows since the interior processes conserve the total content of λ within the global domain. Furthermore, it provides a constraint on numerical realizations of water mass transformation diagnostics.

Interior transformation across constant λ surfaces

We now focus on the transformation occurring along surfaces of constant λ (Figure 52.7) so that

$$G(\lambda)_{\text{int}} = - \lim_{\delta\lambda \rightarrow 0} \frac{1}{\delta\lambda} \int_{\Omega(\lambda \pm \delta\lambda/2)} \nabla \cdot \mathbf{J}_{\text{int}} dV \quad (52.54a)$$

$$= - \lim_{\delta\lambda \rightarrow 0} \frac{1}{\delta\lambda} \left[\int_{\Omega(\lambda + \delta\lambda/2)} \hat{\mathbf{n}} \cdot \mathbf{J}_{\text{int}} d\mathcal{S} - \int_{\Omega(\lambda - \delta\lambda/2)} \hat{\mathbf{n}} \cdot \mathbf{J}_{\text{int}} d\mathcal{S} \right]. \quad (52.54b)$$

Transformation occurs if there is an imbalance between the diffusive transport across the two bounding surfaces, $\Omega(\lambda + \delta\lambda/2)$ and $\Omega(\lambda - \delta\lambda/2)$. As a special case, assume the ocean surface is a constant λ surface with $\lambda = \lambda_{\text{top}}$. Along this surface we have $\hat{\mathbf{n}} \cdot \mathbf{J}_{\text{int}} = 0$, so that the transformation at $\lambda_{\text{top}} - \delta\lambda/2$ has a contribution just from the flux crossing the $\Omega(\lambda_{\text{top}} - \delta\lambda)$ surface

$$\lim_{\delta\lambda \rightarrow 0} G(\lambda_{\text{top}} - \delta\lambda/2)_{\text{int}} = \lim_{\delta\lambda \rightarrow 0} \frac{1}{\delta\lambda} \int_{\Omega(\lambda_{\text{top}} - \delta\lambda)} \hat{\mathbf{n}} \cdot \mathbf{J}_{\text{int}} d\mathcal{S}. \quad (52.55)$$

Likewise, along the top surface we have

$$G(\lambda_{\text{top}})_{\text{int}} = \lim_{\delta\lambda \rightarrow 0} \frac{1}{\delta\lambda} \int_{\Omega(\lambda_{\text{top}} - \delta\lambda/2)} \hat{\mathbf{n}} \cdot \mathbf{J}_{\text{int}} d\mathcal{S}, \quad (52.56)$$

where we set $\hat{\mathbf{n}} \cdot \mathbf{J}_{\text{int}} = 0$ for the surface $\Omega(\lambda_{\text{top}} + \delta\lambda/2)$, since this surface exists outside of the ocean.

The results (52.55) and (52.56) make it appear that $G(\lambda_{\text{top}} - \delta\lambda/2)_{\text{int}}$ and $G(\lambda_{\text{top}})_{\text{int}}$ are unbounded as $\delta\lambda \rightarrow 0$, so long as there is a nonzero diffusive transport through $\Omega(\lambda_{\text{top}} - \delta\lambda)$ or $\Omega(\lambda_{\text{top}} - \delta\lambda/2)$. However, this unbounded water mass transformation is not realized since the diffusive flux gets smaller in magnitude when approaching the ocean surface, and it does so in order to satisfy the no-flux surface boundary condition (52.47) satisfied by interior processes¹¹

$$\hat{\mathbf{n}} \cdot \mathbf{J}_{\text{int}} = 0 \quad \text{at } z = \eta. \quad (52.57)$$

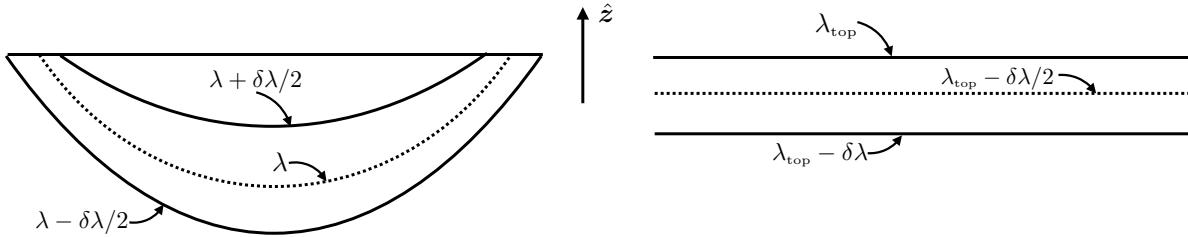


FIGURE 52.7: Example λ surfaces for studying interior transformation due to diffusion. The left panel shows λ surfaces that outcrop to the ocean surface, with the dotted surface the chosen λ surface across which we compute the water mass transformation. The right panel assumes the λ surfaces are flat and with $\lambda = \lambda_{\text{top}}$ the value along the ocean surface.

To further our understanding of the transformation in the region near the ocean surface, let the λ surfaces be flat near the ocean surface so that $\hat{\mathbf{n}} = -\hat{z}$ along $\Omega(\lambda_{\text{top}} - \delta\lambda)$. Setting the diffusive flux to

$$\mathbf{J} = -\rho_0 \kappa \nabla \lambda \quad (52.58)$$

as per a Boussinesq fluid with reference density ρ_0 and diffusivity $\kappa > 0$, leads to the transformation

$$G(\lambda_{\text{top}} - \delta\lambda/2)_{\text{int}} = \frac{1}{\delta\lambda} \int_{\Omega(\lambda_{\text{top}} - \delta\lambda)} (\rho_0 \kappa \partial_z \lambda) d\mathcal{S}. \quad (52.59)$$

With $\kappa \partial_z \lambda > 0$, the transformation is positive, $G(\lambda_{\text{top}} - \delta\lambda/2)_{\text{int}} > 0$, so that water is entrained into the layer bounded by $\lambda_{\text{top}} - \delta\lambda$ and λ_{top} . Note that in the absence of boundary fluxes, the value of λ_{top} is reduced due to diffusive mixing with interior waters since these waters have lower values of λ . Furthermore, as the surface is approached, $\partial_z \lambda$ reduces in magnitude to satisfy the no-flux condition, $\kappa \partial_z \lambda = 0$, at the ocean surface. In this manner, the transformation remains bounded even as $\delta\lambda \rightarrow 0$.

52.6 Budget for seawater mass in a $\Delta\lambda$ -layer

In this section we construct the seawater mass budget for a $\Delta\lambda$ -layer, making reference to Figure 52.6 for the notation. As a shorthand, we write the layer mass as

$$\Delta M(\lambda \pm \Delta\lambda/2) \equiv M(\lambda - \Delta\lambda/2, \lambda + \Delta\lambda/2), \quad (52.60)$$

along with a similar notation for other contributions to the layer mass budget.

¹¹As discussed in Sections 16.5.3, 51.3, and 52.9.3, when water is transported across the ocean surface the diffusive flux picks up a nonzero boundary contribution. That contribution is assumed to be part of the surface transformation in equation (52.48) so that the interior diffusive flux still satisfies the no-flux boundary condition (52.57).

52.6.1 Transport crossing interior open boundaries

As depicted in Figure 52.6, the layer region has an open boundary that is within the interior of the ocean. The mass transport leaving the layer through this interior open boundary is written

$$\Delta\Psi(\lambda \pm \Delta\lambda/2) = \int_{\partial\Omega_{in}(\lambda \pm \Delta\lambda/2)} \rho (\mathbf{v} - \mathbf{v}^{(b)}) \cdot \hat{\mathbf{n}} d\mathcal{S}, \quad (52.61)$$

where $\mathbf{v}^{(b)}$ is the velocity for a point on the boundary and $\hat{\mathbf{n}}$ is the outward normal along the boundary. Introducing the mass distribution for this transport renders the equivalent expression

$$\Delta\Psi(\lambda \pm \Delta\lambda/2) = \int_{\lambda - \Delta\lambda/2}^{\lambda + \Delta\lambda/2} \dot{m}_\Psi(\lambda') d\lambda', \quad (52.62)$$

where

$$\dot{m}_\Psi(\lambda) d\lambda = \text{mass per time crossing } \partial\Omega_{in} \text{ within the increment } [\lambda - d\lambda/2, \lambda + d\lambda/2]. \quad (52.63)$$

We make particular use of \dot{m}_Ψ in Section 52.7 when studying the λ -budget in a $\Delta\lambda$ -layer. One common example for an open interior boundary is when choosing a particular latitude, in which case $\mathbf{v}^{(b)} = 0$ and $\hat{\mathbf{n}} = \hat{\mathbf{y}}$ so that

$$\Delta\Psi(\lambda \pm \Delta\lambda/2) = \int_{\partial\Omega_{in}(\lambda \pm \Delta\lambda/2)} \rho v dx dz. \quad (52.64)$$

In this case, $\partial\Omega_{in}(\lambda \pm \Delta\lambda/2)$ specifies the depth and longitude range for the layer at its intersection along the constant latitude boundary.

52.6.2 Mass transport crossing the ocean surface

The mass transport crossing the ocean free surface is written

$$\Delta W(\lambda \pm \Delta\lambda/2) = - \int_{\partial\Omega_{out}(\lambda \pm \Delta\lambda/2)} \rho (\mathbf{v} - \mathbf{v}^{(\eta)}) \cdot \hat{\mathbf{n}} d\mathcal{S} = \int_{\partial\Omega_{out}(\lambda \pm \Delta\lambda/2)} \mathcal{Q}_m d\mathcal{S}, \quad (52.65)$$

where we made use of the surface kinematic boundary condition (15.63) to write

$$\rho (\mathbf{v} - \mathbf{v}^{(b)}) \cdot \hat{\mathbf{n}} \equiv -\mathcal{Q}_m, \quad (52.66)$$

where $\mathcal{Q}_m d\mathcal{S}$ is the mass transport of water crossing the free surface ($\mathcal{Q}_m > 0$ for water entering the ocean). Introducing the mass distribution leads to the equivalent expression

$$\Delta W(\lambda \pm \Delta\lambda/2) = \int_{\lambda - \Delta\lambda/2}^{\lambda + \Delta\lambda/2} \dot{m}_w(\lambda') d\lambda', \quad (52.67)$$

where

$$\dot{m}_w(\lambda) d\lambda = \text{mass per time crossing } \partial\Omega_{out} \text{ within the increment } [\lambda - d\lambda/2, \lambda + d\lambda/2]. \quad (52.68)$$

We make particular use of \dot{m}_w in Section 52.7 when studying the λ -budget in a $\Delta\lambda$ -layer.

52.6.3 Mass budget

Bringing the above pieces together leads to the layer mass budget

$$\frac{d\Delta M}{dt} = -\Delta\Psi + \Delta W - [G(\lambda + \Delta\lambda/2) - G(\lambda - \Delta\lambda/2)], \quad (52.69)$$

where for brevity we dropped $\lambda \pm \Delta\lambda/2$ arguments for ΔM , $\Delta\Psi$, and ΔW . It is common to define the layer mass *formation* as the mass accumulation within the layer

$$\underbrace{\frac{d\Delta M}{dt} + \Delta\Psi}_{\text{storage + outflow}} = \underbrace{\Delta W - [G(\lambda + \Delta\lambda/2) - G(\lambda - \Delta\lambda/2)]}_{\text{formation into layer } \Omega_{(\lambda \pm \Delta\lambda/2)}}. \quad (52.70)$$

This equality defines water mass formation as the time change for the mass within the layer (sometimes referred to as the *storage term*), plus the net mass leaving the interior open boundary. Formation into a layer occurs if there is mass converging through transformation across interior layer interfaces, plus mass entering through the surface boundary outcrop region.

We arrive at a differential equation for the mass budget (52.70) by dividing through by the layer increment, $\Delta\lambda$, and taking the limit as this increment tends to zero

$$\frac{\partial}{\partial\lambda} \left[\frac{dM}{dt} + \Psi - W + G \right] = 0. \quad (52.71)$$

Integrating from a reference value $\lambda_{-\infty}$ to λ leads to

$$\int_{\lambda_{-\infty}}^{\lambda} \frac{\partial\Psi}{\partial\lambda} d\lambda = \int_{\lambda_{-\infty}}^{\lambda} \frac{\partial}{\partial\lambda} \left[-\frac{dM}{dt} + W - G \right] d\lambda \implies \Psi = -\frac{dM}{dt} + W - G. \quad (52.72)$$

We dropped the contribution from the constant reference value, $\lambda_{-\infty}$, since it sits outside of the ocean domain. The differential water mass equation (52.72) is a continuous version of the discrete equation (52.70). We wrote this equation as an expression for Ψ given that an accumulation from the bottom up leads to a transport streamfunction in the steady limit.

52.7 Budget for λ mass in a λ_∞ -region

We build from our understanding of the seawater mass budget in Section 52.6 to develop a budget for the mass of λ within the λ_∞ -region of Section 52.4.4 and as illustrated in Figure 52.8. Part of our aim is to further develop the formalism while also offering added insights into the causes for water mass transformation, $G(\lambda)$.

We here choose to be specific by considering λ to be an intensive property such as a material tracer concentration, in which case $\lambda \rho dV$ has dimensions of tracer mass.¹² For non-material scalar fields, such as Conservative Temperature or buoyancy, the dimensions are modified accordingly.

52.7.1 Processes affecting the mass of λ -stuff

Our starting point is the Leibniz-Reynolds budget for a scalar field derived in Section 16.3.4, here including the possibility of scalar sources

$$\frac{d}{dt} \left[\int_{\Omega(\lambda \leq \lambda_\infty)} \rho \lambda dV \right] = \int_{\Omega(\lambda \leq \lambda_\infty)} \rho \dot{\Upsilon} dV - \int_{\partial\Omega(\lambda \leq \lambda_\infty)} [\rho \lambda (\mathbf{v} - \mathbf{v}^{(b)}) + \mathbf{J}] \cdot \hat{\mathbf{n}} d\mathcal{S}. \quad (52.73)$$

The right hand side of this budget equation can be decomposed into the following processes illustrated in Figure 52.8.

¹²See Section 16.3.1 for more on intensive and extensive fluid properties.

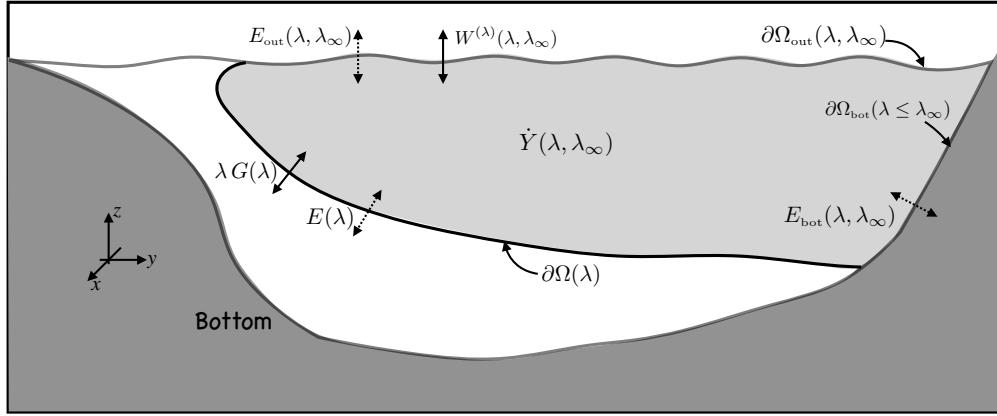


FIGURE 52.8: A λ_∞ -region for studying the λ budget, with the region bounded by the solid-earth bottom, $\partial\Omega_{\text{bot}}(\lambda, \lambda_\infty)$, the ocean surface boundary, $\partial\Omega_{\text{out}}(\lambda, \lambda_\infty)$, and the λ -interface, $\partial\Omega(\lambda)$. Hence, in contrast to the $\Delta\lambda$ -region shown in Figure 52.6, the λ_∞ -region has no interior open boundary. Along the surface boundary, the λ budget is affected by the non-advectional transport, $E_{\text{out}}(\lambda, \lambda_\infty)$, arising from processes such as diffusion, plus advective transport, $W^{(\lambda)}(\lambda, \lambda_\infty)$, arising from mass transported across the surface that can carry a non-zero amount of λ . Along the bottom boundary, the λ budget is affected by non-advectional transport, $E_{\text{bot}}(\lambda, \lambda_\infty)$, arising from processes such as geothermal heating. There is no corresponding advective transport along the bottom since we assume there is no mass crossing the ocean bottom. Along the interior boundary, $\partial\Omega(\lambda)$, the budget is affected by non-advectional transport, $E(\lambda)$, arising from diffusion, as well as advective transport through $\lambda G(\lambda)$, with $G(\lambda)$ the interior water mass transformation from Section 52.5. Finally, there is the possibility for an interior volume source, $\dot{Y}(\lambda \leq \lambda_\infty)$, particularly for buoyancy in the presence of a nonlinear equation of state, and for biogeochemical tracers.

Non-conservative sources and sinks

As noted in Section 52.5.3, the source term $\rho \dot{Y}$ accounts for processes that cannot be represented as the convergence of a flux. We write its region integrated contribution using the shorthand

$$\dot{Y}(\lambda \leq \lambda_\infty) \equiv \int_{\Omega(\lambda \leq \lambda_\infty)} \rho \dot{Y} dV \equiv \int_\lambda^{\lambda_\infty} \dot{m}_{\dot{Y}}(\lambda') d\lambda'. \quad (52.74)$$

The final equality introduced the distribution function for the source, in which

$$\dot{m}_{\dot{Y}}(\lambda) d\lambda = \text{mass per time of } \lambda \text{ created within the region } [\lambda - d\lambda/2, \lambda + d\lambda/2]. \quad (52.75)$$

Transport from non-advectional processes

The contribution from boundary area integrated non-advectional fluxes appears in the term

$$-\int_{\partial\Omega(\lambda \leq \lambda_\infty)} \mathbf{J} \cdot \hat{\mathbf{n}} d\mathcal{S} = -\int_{\partial\Omega_{\text{out}}(\lambda \leq \lambda_\infty)} \mathbf{J} \cdot \hat{\mathbf{n}} d\mathcal{S} - \int_{\partial\Omega_{\text{bot}}(\lambda \leq \lambda_\infty)} \mathbf{J} \cdot \hat{\mathbf{n}} d\mathcal{S} - \int_{\partial\Omega(\lambda)} \mathbf{J} \cdot \hat{\mathbf{n}} d\mathcal{S}. \quad (52.76)$$

Recall the minus signs arise since a non-advectional flux increases the λ content of the region if the flux is oriented into the region, whereas $\hat{\mathbf{n}}$ is the region outward normal. The surface $\partial\Omega_{\text{out}}(\lambda \leq \lambda_\infty)$ extends along the upper ocean boundary and supports non-advectional surface boundary fluxes. Likewise, the boundary $\partial\Omega_{\text{bot}}(\lambda \leq \lambda_\infty)$ intersects the ocean bottom and generally experiences bottom boundary fluxes such as geothermal heating. Finally, the surface $\partial\Omega(\lambda)$ has non-advectional fluxes that cross the λ -interface, with diffusive fluxes the canonical example. The boundary area integrated non-advectional fluxes give rise to non-advectional transports, with dimensions of mass of λ per time, that are written using the shorthand

$$-\int_{\partial\Omega(\lambda \leq \lambda_\infty)} \mathbf{J} \cdot \hat{\mathbf{n}} d\mathcal{S} = E_{\text{out}}(\lambda, \lambda_\infty) + E_{\text{bot}}(\lambda, \lambda_\infty) + E(\lambda), \quad (52.77)$$

with a term having a positive value if it increases the λ mass of the region. We furthermore find it useful to introduce the distribution functions according to

$$E_{\text{out}}(\lambda, \lambda_\infty) + E_{\text{bot}}(\lambda, \lambda_\infty) = \int_\lambda^{\lambda_\infty} [\dot{m}_{\text{out}}^E(\lambda') + \dot{m}_{\text{bot}}^E(\lambda')] d\lambda', \quad (52.78)$$

where

$$\dot{m}_{\text{out}}^E(\lambda) d\lambda = \text{mass of } \lambda \text{ per time from } \partial\Omega_{\text{out}} \text{ transport in range } [\lambda - d\lambda/2, \lambda + d\lambda/2] \quad (52.79a)$$

$$\dot{m}_{\text{bot}}^E(\lambda) d\lambda = \text{mass of } \lambda \text{ per time from } \partial\Omega_{\text{bot}} \text{ transport in range } [\lambda - d\lambda/2, \lambda + d\lambda/2]. \quad (52.79b)$$

λ transported with interior boundary mass fluxes

We next consider the contribution to the budget equation (52.73) arising from the transport of λ with mass that crosses the interior boundary, $\partial\Omega(\lambda)$, whereby

$$-\int_{\partial\Omega(\lambda)} \rho \lambda (\mathbf{v} - \mathbf{v}^{(\lambda)}) \cdot \hat{\mathbf{n}} d\mathcal{S} = -\lambda \int_{\partial\Omega(\lambda)} \rho (\mathbf{v} - \mathbf{v}^{(\lambda)}) \cdot \hat{\mathbf{n}} d\mathcal{S} = \lambda G(\lambda), \quad (52.80)$$

To reach this result we noted that λ can be pulled outside of the $\partial\Omega(\lambda)$ integral since it is constant along this surface, thus allowing for the introduction of the water mass transformation, $G(\lambda)$, given by equation (52.40).

Surface boundary mass fluxes

The final term contributing to the right hand side of the λ budget equation (52.73) arises from the surface mass transport along the boundary $\partial\Omega_{\text{out}}(\lambda \leq \lambda_\infty)$

$$-\int_{\partial\Omega(\lambda \leq \lambda_\infty)} \rho \lambda (\mathbf{v} - \mathbf{v}^{(b)}) \cdot \hat{\mathbf{n}} d\mathcal{S} = \int_{\partial\Omega_{\text{out}}(\lambda \leq \lambda_\infty)} \lambda \mathcal{Q}_m d\mathcal{S} \equiv W^{(\lambda)}(\lambda, \lambda_\infty). \quad (52.81)$$

To reach the first equality we followed the steps in Section 52.6.2 by using the kinematic boundary condition (15.63) to introduce the surface mass transport, $\mathcal{Q}_m d\mathcal{S}$. The final equality introduced a shorthand that corresponds to the $W(\lambda, \lambda_\infty)$ from Section 52.6.2. In the following, we find it useful to introduce the mass distribution function $\dot{m}_w(\lambda)$ from equation (52.68), thus rendering

$$W(\lambda, \lambda_\infty) = \int_\lambda^{\lambda_\infty} \dot{m}_w(\lambda') d\lambda' \quad \text{and} \quad W^{(\lambda)}(\lambda, \lambda_\infty) = \int_\lambda^{\lambda_\infty} \lambda \dot{m}_w(\lambda') d\lambda'. \quad (52.82)$$

Following the discussion in Sections 51.3.2, we have not assumed a relation between λ along the interface, $\partial\Omega_{\text{out}}(\lambda \leq \lambda_\infty)$, and the concentration, λ_m , contained in the entering mass. We prefer to keep the discussion general for now, providing a relation only when necessary.

52.7.2 Summary of the λ budget

Bringing terms together leads to the λ mass budget equation (52.73) in the form

$$\begin{aligned} \frac{d}{dt} \left[\int_{\Omega(\lambda \leq \lambda_\infty)} \rho \lambda dV \right] \\ = \dot{Y}(\lambda \leq \lambda_\infty) + E_{\text{out}}(\lambda, \lambda_\infty) + E_{\text{bot}}(\lambda, \lambda_\infty) + E(\lambda) + W^{(\lambda)}(\lambda, \lambda_\infty) + \lambda G(\lambda). \end{aligned} \quad (52.83)$$

This budget takes the equivalent form in terms of the distribution functions

$$\frac{d}{dt} \int_{\lambda}^{\lambda_\infty} \lambda' m(\lambda') d\lambda' = \int_{\lambda}^{\lambda_\infty} [\dot{m}_{\dot{Y}}(\lambda') + \dot{m}_{\text{out}}^E(\lambda') + \dot{m}_{\text{bot}}^E(\lambda') + \lambda \dot{m}_W(\lambda')] d\lambda' + E(\lambda) + \lambda G(\lambda). \quad (52.84)$$

Setting λ to a global constant and assuming there are no seawater mass sources leads to the seawater mass budget for the λ_∞ -region

$$\frac{d}{dt} \left[\int_{\Omega(\lambda \leq \lambda_\infty)} \rho dV \right] = W(\lambda, \lambda_\infty) + G(\lambda), \quad (52.85)$$

which takes on the following form in terms of distributions

$$\frac{d}{dt} \int_{\lambda}^{\lambda_\infty} m(\lambda') d\lambda' = \int_{\lambda}^{\lambda_\infty} \dot{m}_W(\lambda') d\lambda' + G(\lambda). \quad (52.86)$$

52.7.3 Processes leading to water mass transformation

We now focus on the left hand side of equation (52.85) for the purpose of identifying terms leading to water mass transformation by making use of the moment equation (52.37) to write

$$\int_{\Omega(\lambda \leq \lambda_\infty)} \rho \lambda dV = M(\lambda, \lambda_\infty) \langle \lambda \rangle = M(\lambda, \lambda_\infty) \lambda + \int_{\lambda}^{\lambda_\infty} M(\lambda', \lambda_\infty) d\lambda', \quad (52.87)$$

which then leads to

$$\frac{d[M(\lambda, \lambda_\infty) \langle \lambda \rangle]}{dt} = \lambda \frac{dM(\lambda, \lambda_\infty)}{dt} + \int_{\lambda}^{\lambda_\infty} \frac{dM(\lambda', \lambda_\infty)}{dt} d\lambda'. \quad (52.88)$$

Integrated water mass transformation over the λ_∞ -region

Use of the λ budget equation (52.85) for the left hand side of equation (52.88), and the mass budget equation (52.85) for the right hand side renders

$$\begin{aligned} \dot{Y}(\lambda \leq \lambda_\infty) + E_{\text{out}}(\lambda, \lambda_\infty) + E_{\text{bot}}(\lambda, \lambda_\infty) + E(\lambda) + W^{(\lambda)}(\lambda, \lambda_\infty) + \lambda G(\lambda) \\ = \lambda [W(\lambda, \lambda_\infty) + G(\lambda)] + \int_{\lambda}^{\lambda_\infty} [W(\lambda', \lambda_\infty) + G(\lambda')] d\lambda'. \end{aligned} \quad (52.89)$$

Contributions from the surface boundary mass fluxes cancel as per the following identity

$$W^{(\lambda)}(\lambda, \lambda_\infty) - \lambda W(\lambda, \lambda_\infty) = \int_{\lambda}^{\lambda_\infty} (\lambda' - \lambda) \dot{m}_W(\lambda') d\lambda' \quad (52.90a)$$

$$= \int_{\lambda}^{\lambda_\infty} \left[\int_{\lambda'}^{\lambda_\infty} \dot{m}_W(\lambda'') d\lambda'' \right] d\lambda' \quad (52.90b)$$

$$= \int_{\lambda}^{\lambda_\infty} W(\lambda', \lambda_\infty) d\lambda', \quad (52.90c)$$

where the second equality follows from the double integral formula (3.97). To understand the physical reason we see no water mass transformation from surface mass fluxes, recall the discussion in Section 51.3.2. Namely, mixing and internal sources provide the only means for irreversible changes

to water masses and thus to water mass transformation. In contrast, boundary mass transport contributes to transformation only if the mass participates in mixing. That is, the mass associated with boundary mass transport is incorporated into the ocean (or leaves the ocean) only in the presence of mixing. It is thus reassuring that the formalism leads to this same conclusion. Furthermore, this result is consistent with the expression (52.45b), whereby the water mass transformation is, again, determined solely in terms of the non-advection fluxes at the region boundaries plus the interior source term.

Cancelling the mass transport terms thus leads to the integrated water mass transformation

$$\int_{\lambda}^{\lambda_{\infty}} G(\lambda') d\lambda' = \dot{Y}(\lambda \leq \lambda_{\infty}) + E_{\text{out}}(\lambda, \lambda_{\infty}) + E_{\text{bot}}(\lambda, \lambda_{\infty}) + E(\lambda), \quad (52.91)$$

which is, again, just another way to write equation (52.45b). We thus see that the accumulated effects from sources within the interior plus non-advection fluxes along the surface and interior boundaries lead to an integrated interior water mass transformation. Note that each term in this equation has dimensions of λ -mass per time.

Water mass transformation across the λ -interface

We can develop an expression for the water mass transformation across the λ -interface by taking $\partial/\partial\lambda$ of equation (52.91) to reveal

$$G(\lambda) = -\frac{\partial}{\partial\lambda} \left[\dot{Y}(\lambda \leq \lambda_{\infty}) + E_{\text{out}}(\lambda, \lambda_{\infty}) + E_{\text{bot}}(\lambda, \lambda_{\infty}) + E(\lambda) \right], \quad (52.92)$$

thus revealing that the water mass transformation across a λ -surface is the λ -convergence of the mixing processes plus the interior sources. Finally, this equation takes on the following distributional form

$$G(\lambda) = \dot{m}_{\dot{Y}}(\lambda) + \dot{m}_{\text{bot}}^{\text{E}}(\lambda) + \dot{m}_{\text{out}}^{\text{E}}(\lambda) - \frac{\partial E(\lambda)}{\partial\lambda}. \quad (52.93)$$

52.8 Surface water mass transformation

We have articulated all the terms needed to form the $\Delta\lambda$ -layer mass budget according to Figure 52.6 as well as the λ_{∞} -region mass budget according to Figure 52.8. In this section we focus on contributions to transformation from surface processes in the transformation equation (52.93)

$$G_{\text{out}}(\lambda) \equiv -\frac{\partial E_{\text{out}}(\lambda, \lambda_{\infty})}{\partial\lambda} = \dot{m}_{\text{out}}^{\text{E}}(\lambda) = -\lim_{\delta\lambda \rightarrow 0} \frac{1}{\delta\lambda} \int_{\partial\Omega_{\text{out}}(\lambda \pm \delta\lambda/2)} \mathbf{J} \cdot \hat{\mathbf{n}} d\mathcal{S}, \quad (52.94)$$

where the final equality made use of equation (52.45b). Such *surface transformation* forms the focus of many studies of water mass transformation because it only requires surface boundary information, which is generally more accessible than information from interior ocean mixing processes or bottom geothermal processes. Furthermore, much of the transformation of water occurs in surface regions since this region is home to large contributions from surface boundary fluxes and associated ocean mixing. The basic equation we use is the non-advection flux equation (16.72), rewritten here for our scalar field λ

$$-\mathbf{J} \cdot \hat{\mathbf{n}} = \mathcal{Q}_{\lambda} - \lambda \mathcal{Q}_{\text{m}} = \mathcal{Q}_{\lambda}^{\text{non-adv}} + (\lambda_{\text{m}} - \lambda) \mathcal{Q}_{\text{m}}. \quad (52.95)$$

52.8.1 Circulation driven by surface transformation

The layer mass budget (52.70) and its continuous expression (52.72) provide the basis for inferences about circulation and transformation. As an illustration, consider the continuous expression (52.72) and integrate from the reference value up to λ

$$\int_{\lambda_{-\infty}}^{\lambda} \Psi d\lambda' = \int_{\lambda_{-\infty}}^{\lambda} \left[-\frac{\partial M}{\partial t} + W - G \right] d\lambda'. \quad (52.96)$$

The left hand side is an expression for the circulation in λ -space at the specified interior open boundary. The right hand side means that a nonzero circulation is driven by mass through the ocean surface, convergence of mass transformed across the λ -interfaces, and/or time changes to the mass within the domain. Correspondingly, in the absence of surface boundary mass fluxes, a steady circulation is driven just by water mass transformation

$$\int_{\lambda_{-\infty}}^{\lambda} \Psi d\lambda' = - \int_{\lambda_{-\infty}}^{\lambda} G d\lambda'. \quad (52.97)$$

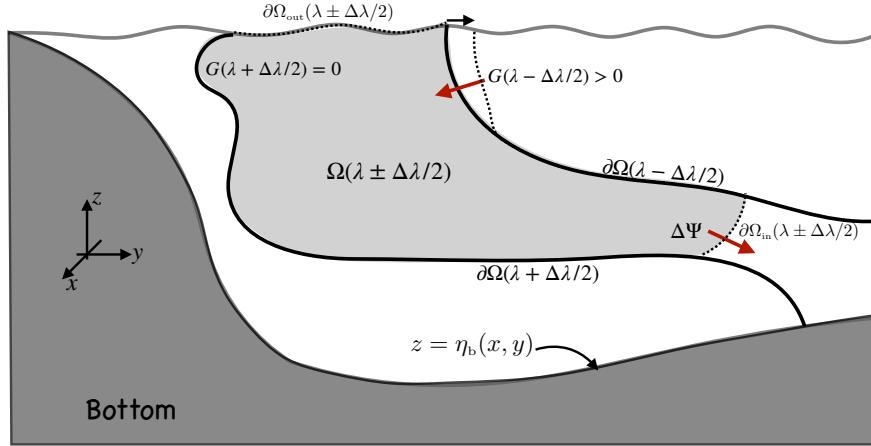


FIGURE 52.9: An example of a surface transformation driven circulation oriented according to the Southern Ocean with Antarctica to the left. Here we depict a layer that is exposed to some form of air-sea interaction that causes the $\lambda - \Delta\lambda/2$ -interface to move meridionally. For example, if $\lambda = \gamma$, then an air-sea buoyancy flux generally causes γ outcrops to move laterally. Movement of the near-surface portion of the $\partial\Omega(\lambda - \Delta\lambda/2)$ interface causes seawater to entrain into the layer and thus contribute to the surface water mass transformation $G(\lambda - \Delta\lambda/2) > 0$ (red arrow near the surface directed to the south). In turn, the boundary $\partial\Omega_{in}(\lambda \pm \Delta\lambda/2)$ expands as the near-surface portion of the interface $\partial\Omega(\lambda - \Delta\lambda/2)$ moves to the south as a result of the entrained new water (black arrow moving to the north). If there is a net convergence of water mass into the layer (e.g., by having less mass move across the other layer interface at $\partial\Omega(\lambda + \Delta\lambda/2)$), then mass accumulates within the layer $[\lambda - \Delta\lambda/2, \lambda + \Delta\lambda/2]$. There is a steady state (i.e., layer mass is constant) only if the same amount of mass that converges into the layer via surface or interior water mass transformation leaves the layer through circulation at the boundary $\partial\Omega_{in}(\lambda \pm \Delta\lambda/2)$.

We depict an example in Figure 52.9 where the surface outcrop of the layer is exposed to air-sea interactions that lead to a meridional movement of the $\lambda - \Delta\lambda/2$ -interface. This movement laterally entrains mass into the layer. If there is a net convergence of mass into the layer, then the layer mass increases. A steady state mass budget for the layer is reached if the amount of mass entrained through surface transformation is reflected in the same mass leaving through the circulation at the open boundary $\partial\Omega_{in}(\lambda \pm \Delta\lambda/2)$. We depict another case in Figure 52.10, here focusing on how a meridional gradient in the surface buoyancy loss causes entrainment into buoyancy layers. These and other statements related to the water mass budget are rather routine mathematically. Yet

since the mass budget is formulated over layers, the mass budget offers the means to make very general statements about the circulation even without a direct measurement of the flow. This is a key power of water mass analysis.

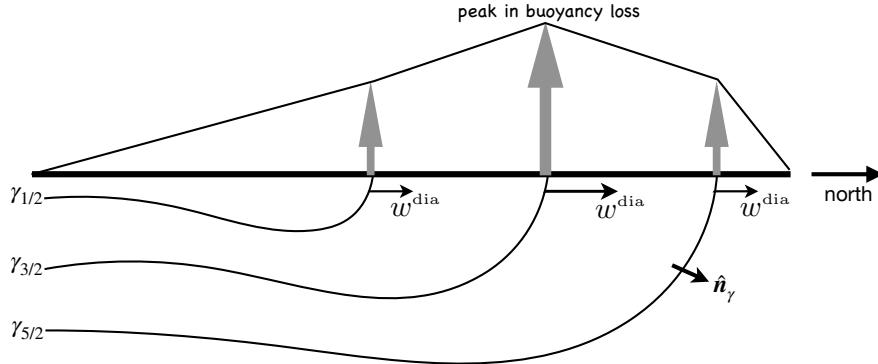


FIGURE 52.10: An example of surface water mass transformation, here illustrating the effects of transformation due to a meridional gradient in the surface buoyancy loss (we here assume that λ is buoyancy as measured by the neutral density γ). The example is oriented for the northern hemisphere with increasing latitudes to the north/right. Buoyancy loss is denoted by the blue vertical arrows indicating the removal of buoyancy from the ocean, thus causing surface water to loose buoyancy. The surface buoyancy loss causes γ interfaces to migrate to the south, which in turn causes dianeutral mass flux to move from lighter layers to denser layers (black vectors pointed to the north, w^{dia}). With a peak in the buoyancy loss at a particular latitude, more entrainment is driven into the layer to the north of the peak (water converges to the layer $\gamma_{3/2} \leq \gamma \leq \gamma_{5/2}$) and less entrainment into the layer to the south (water diverges from the layer $\gamma_{1/2} \leq \gamma \leq \gamma_{3/2}$).

52.8.2 Further study

Much of the formalism in this section follows that reviewed by [Groeskamp et al. \(2019\)](#). This paper offers specific examples of water mass transformation analysis as well as citations to numerous research papers.

52.9 Buoyancy water mass transformation

In Section 52.3 we considered the transformation of water masses as defined by buoyancy classes, with $\lambda = \gamma$ where γ is a field whose isosurfaces approximate constant buoyancy directions; i.e., the neutral directions from Section 25.5. We here fill in further details for such buoyancy water mass analysis.

52.9.1 Material time changes to S and Θ

The material time derivative of γ can be written as the sum of contributions from salinity and Conservative Temperature

$$\rho \dot{\gamma} = \frac{\partial \gamma}{\partial S} \rho \dot{S} + \frac{\partial \gamma}{\partial \Theta} \rho \dot{\Theta}. \quad (52.98)$$

Following the decomposition of the water mass transformation in Section 52.5.3, for a general tracer, we here write the material time derivatives in the form

$$\rho \dot{S} = -\nabla \cdot \mathbf{J}_{\text{int}}^{(S)} - \mathbf{J}_{\text{out}}^{(S)} \cdot \hat{\mathbf{n}} \delta(z - \eta) - \mathbf{J}_{\text{bot}}^{(S)} \cdot \hat{\mathbf{n}} \delta(z - \eta_b) \quad (52.99a)$$

$$\rho \dot{\Theta} = -\nabla \cdot \mathbf{J}_{\text{int}}^{(\Theta)} - \mathbf{J}_{\text{out}}^{(\Theta)} \cdot \hat{\mathbf{n}} \delta(z - \eta) - \mathbf{J}_{\text{bot}}^{(\Theta)} \cdot \hat{\mathbf{n}} \delta(z - \eta_b), \quad (52.99b)$$

where we assumed there are no interior sources of S or Θ . The surface and bottom boundary contributions are weighted by Dirac delta distributions and projected into the normal direction along the two respective boundary surfaces.

52.9.2 Water mass transformation

The decomposition of water mass transformation given by equation (52.48) takes on the form for buoyancy

$$\begin{aligned}
 G(\gamma) = & - \underbrace{\lim_{\delta\gamma \rightarrow 0} \frac{1}{\delta\gamma} \int_{\Omega(\gamma \pm \delta\gamma/2)} \left(\frac{\partial\gamma}{\partial S} \nabla \cdot \mathbf{J}_{\text{int}}^{(S)} + \frac{\partial\gamma}{\partial \Theta} \nabla \cdot \mathbf{J}_{\text{int}}^{(\Theta)} \right) dV }_{\text{int buoyancy transformation} = \text{volume integral of convergence}} \\
 & - \underbrace{\lim_{\delta\gamma \rightarrow 0} \frac{1}{\delta\gamma} \int_{\partial\Omega_{\text{out}}(\gamma \pm \delta\gamma/2)} \left(\frac{\partial\gamma}{\partial S} \mathbf{J}_{\text{out}}^{(S)} + \frac{\partial\gamma}{\partial \Theta} \mathbf{J}_{\text{out}}^{(\Theta)} \right) \cdot \hat{\mathbf{n}} d\mathcal{S} }_{\text{surface buoyancy transformation} = \text{area integral of surface boundary fluxes}} \\
 & - \underbrace{\lim_{\delta\gamma \rightarrow 0} \frac{1}{\delta\gamma} \int_{\partial\Omega_{\text{bot}}(\gamma \pm \delta\gamma/2)} \left(\frac{\partial\gamma}{\partial S} \mathbf{J}_{\text{bot}}^{(S)} + \frac{\partial\gamma}{\partial \Theta} \mathbf{J}_{\text{bot}}^{(\Theta)} \right) \cdot \hat{\mathbf{n}} d\mathcal{S} }_{\text{bottom buoyancy transformation} = \text{area integral of bottom boundary fluxes}}. \tag{52.100}
 \end{aligned}$$

Note that contributions from cabbeling and thermobaricity (Section 50.4) arise from the interior transformation. To touch base with the commonly employed surface buoyancy transformation, insert the surface fluxes into equation (52.100) as per Section 52.9.3 to write

$$\begin{aligned}
 G(\gamma)^{\text{surface}} = & \\
 & \lim_{\delta\gamma \rightarrow 0} \frac{1}{\delta\gamma} \int_{\partial\Omega_{\text{out}}(\gamma \pm \delta\gamma/2)} (\beta_S [\mathcal{Q}_S^{\text{non-adv}} + (S_m - S) \mathcal{Q}_m] - \alpha_\Theta [\mathcal{Q}_\Theta^{\text{non-adv}} + (\Theta_m - \Theta) \mathcal{Q}_m]) \gamma d\mathcal{S}, \tag{52.101}
 \end{aligned}$$

where we introduced the thermal expansion and saline contraction coefficients, here defined according to γ ,

$$\alpha_\Theta = -\frac{1}{\gamma} \frac{\partial\gamma}{\partial \Theta} \quad \text{and} \quad \beta_S = \frac{1}{\gamma} \frac{\partial\gamma}{\partial S}. \tag{52.102}$$

The integrand to equation (52.101) corresponds to minus the surface buoyancy flux derived in Section 51.4.3. The only difference is that we here make use of the surface element, $d\mathcal{S}$, and the corresponding fluxes $\mathcal{Q}_\Theta^{\text{non-adv}}$, $\mathcal{Q}_S^{\text{non-adv}}$, and \mathcal{Q}_m . However, if the ocean surface has no overturns, we can write its vertical position as $z = \eta(x, y, t)$ and can also define the horizontal projection of the area element as (see equation (5.30))

$$d\mathcal{S} = \sqrt{1 + \nabla(\eta)^2} dA. \tag{52.103}$$

In this case we can introduce the fluxes $\mathcal{Q}_\Theta^{\text{non-adv}}$, $\mathcal{Q}_S^{\text{non-adv}}$, and \mathcal{Q}_m used in Section 51.4.3 via

$$\mathcal{Q}_\Theta^{\text{non-adv}} d\mathcal{S} = Q_\Theta^{\text{non-adv}} dA \quad \text{and} \quad \mathcal{Q}_S^{\text{non-adv}} d\mathcal{S} = Q_S^{\text{non-adv}} dA \quad \text{and} \quad \mathcal{Q}_m d\mathcal{S} = Q_m dA. \tag{52.104}$$

Recall that $G(\gamma) > 0$ occurs when water is transformed into regions with larger γ . For example, net surface cooling in the presence of a positive thermal expansion coefficient ($\alpha_\Theta > 0$) leads to $\mathcal{Q}_\Theta^{\text{non-adv}} + (\Theta_m - \Theta) \mathcal{Q}_m < 0$. Such cooling then leads to a positive contribution to $G(\gamma)^{\text{surface}}$ as water is transformed from light to heavy γ -classes. Likewise, a positive net salt transport into the upper ocean, $\mathcal{Q}_S^{\text{non-adv}} + (S_m - S) \mathcal{Q}_m > 0$, leads to a positive contribution to $G(\gamma)^{\text{surface}}$.

52.9.3 Summary of surface non-advection flux for S and Θ

For completeness, we review a few of the distinct characteristics of surface non-advection fluxes of S and Θ as detailed in Section 51.3.2, here working with salinity, S , rather than salt concentration, $S = S/1000$.

Non-advection salt flux

The non-advection surface boundary flux for salt is given by equation (51.33), here written as

$$-\mathbf{J}^{(S)} \cdot \hat{\mathbf{n}} = \mathcal{Q}_S^{\text{non-adv}} + (S_m - S) \mathcal{Q}_m, \quad (52.105)$$

where $\mathcal{Q}_S^{\text{non-adv}}$ is a non-advection salt flux, such as might arise from parameterized turbulent transfer. For the salt concentration of water crossing the ocean surface, we generally take $S_m = 0$ for precipitation, evaporation, and river runoff, whereas $S_m \neq 0$ for ice melt and formation. Furthermore, the boundary term, $S = S(z = \eta)$, is commonly approximated by the bulk salt concentration in the upper ocean.

Non-advection flux for Conservative Temperature

For Conservative Temperature we follow the discussion in Section 51.3.3, whereby the non-advection flux is given by equation

$$-\mathbf{J}^{(\Theta)} \cdot \hat{\mathbf{n}} = \mathcal{Q}_\Theta^{\text{non-adv}} + (\Theta_m - \Theta) \mathcal{Q}_m. \quad (52.106)$$

It is common to approximate the difference $\Theta_m - \Theta(z = \eta) = 0$, in which case the non-advection flux is just due to turbulent and radiative heat fluxes

$$-\mathbf{J}^{(\Theta)} \cdot \hat{\mathbf{n}} = \mathcal{Q}_\Theta^{\text{non-adv}} \text{ if } \Theta_m - \Theta(z = \eta) = 0. \quad (52.107)$$

52.10 Tracer mass analysis

In Sections 52.6 and 52.7 we developed the budgets for λ within layers defined λ . Here we extend that analysis to develop budgets for a tracer concentration, C , localized in a region within a layer of buoyancy, γ , as depicted in Figure 52.11. The upper panels to this figure illustrate a tracer patch in geographic/depth space along with isolines of buoyancy, whereas the lower panels show the tracer distribution (histogram) binned within the buoyancy classes. If the tracer is mixed within a layer, such as via the neutral diffusion process of Section 50.3, then the tracer patch is spread laterally within the buoyancy layer and yet the distribution (lower panel) is unchanged. In contrast, if the tracer is mixed across layer interfaces then the tracer distribution is spread within buoyancy space.

Another means to alter the tracer distribution is to modify the buoyancy field. This situation is especially common for tracer near the surface, where boundary buoyancy forcing can act to move the layers thus causing tracer to move between layers even if the tracer patch is stationary in geographical space. That is, if the tracer moves at a velocity distinct from the buoyancy field, then its distribution within buoyancy classes will change.

52.10.1 General form of the mass budget

As depicted in Figure 52.11 for buoyancy layers, and Figure 52.12 for generic layers, there are two general processes whereby a tracer distribution within layers can be modified: (i) the tracer can mix between layers and (ii) the layers can move relative to the tracer. These ideas transcend buoyancy

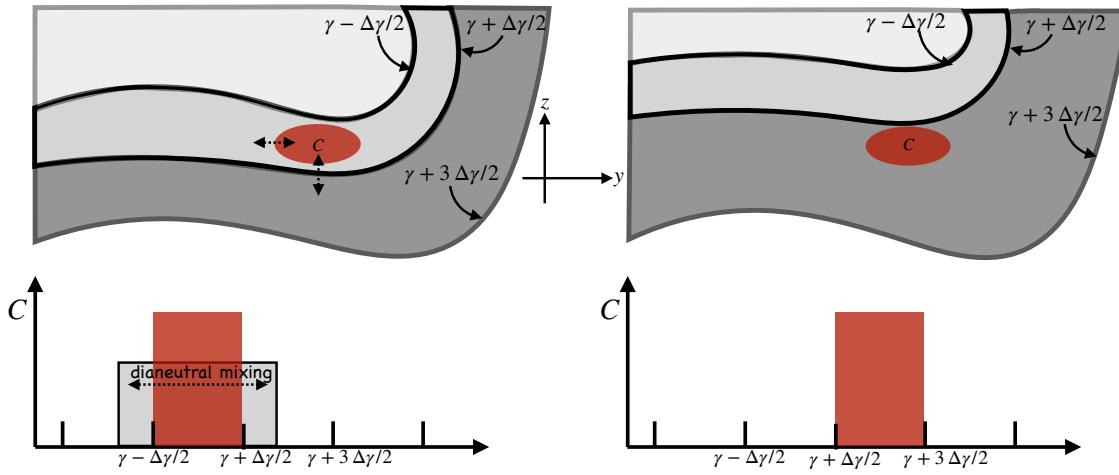


FIGURE 52.11: Depicting a tracer patch within the buoyancy layer bounded by the interface values $[\gamma - \Delta\gamma/2, \gamma + \Delta\gamma/2]$ (left panel) and $[\gamma + \Delta\gamma/2, \gamma + 3\Delta\gamma/2]$ (right panel). The upper panels show the tracer and buoyancy in geographic/depth space whereas the bottom panels show the tracer distribution (histogram) binned according to buoyancy. There are two general means to modify the distribution of tracer within the buoyancy classes. The first occurs via dianeutral mixing that spreads the tracer distribution to other buoyancy layers as depicted by the vertical arrow in the upper left panel and the horizontal arrows in the lower left panel. The lateral arrow in the upper left panel depicts neutral diffusion, which laterally spreads the tracer within a layer but does not alter the distribution across layers (see Section 50.3). The second means to alter the distribution occurs when the buoyancy surfaces move relative to the tracer. This scenario is depicted in the lower right panel whereby the tracer patch originally in buoyancy layer $[\gamma - \Delta\gamma/2, \gamma + \Delta\gamma/2]$ now finds itself in the layer $[\gamma + \Delta\gamma/2, \gamma + 3\Delta\gamma/2]$. This depiction is not realistic, since motion of interior buoyancy surfaces generally occurs along with mixing of tracer patches. Nonetheless, this example emphasizes that motion of the buoyancy surfaces need not coincide with motion of the tracer patch.

and thus can be applied to any scalar field, λ , used to classify water masses. We quantify these two processes by writing the time change of tracer content within a λ -layer, which is arrived at by applying the Leibniz-Reynolds transport theorem from Section 16.3.4 to a λ -layer

$$\frac{d}{dt} \Delta M_C(\lambda \pm \Delta\lambda/2) = \int_{\Omega(\lambda \pm \Delta\lambda/2)} \rho \dot{C} dV - \oint_{\partial\Omega(\lambda \pm \Delta\lambda/2)} \rho C (\mathbf{v} - \mathbf{v}^{(b)}) \cdot \hat{\mathbf{n}} dS, \quad (52.108)$$

where

$$\Delta M_C(\lambda \pm \Delta\lambda/2) = \int_{\Omega(\lambda \pm \Delta\lambda/2)} \rho C dV \quad (52.109)$$

is the mass of tracer within the layer. The volume integral on the right hand side of equation (52.108) arises from material time changes to the tracer within the layer, whereas the surface integral arises from dia-surface transport across the layer boundary.

52.10.2 Tracer processes

We determine the material time changes for a “conservative” tracer according to the convergence of a flux

$$\rho \dot{C} = \rho \frac{DC}{Dt} = -\nabla \cdot \mathbf{J}. \quad (52.110)$$

Many biogeochemical tracers have additional “non-conservative” source terms beyond the flux convergence considered here. These effects can be readily incorporated into the following by adding a source term that acts throughout a layer and not just at the layer boundaries.

Gauss' divergence theorem converts the convergence, $-\nabla \cdot \mathbf{J}$, into the area integral of fluxes over the layer boundaries, including interior layer interfaces as well as intersections with the surface and bottom boundaries. For the interior interfaces it is typically simpler diagnostically to bin the volume integrated material time changes within the λ -classes. In contrast, the surface and bottom boundary contributions are fed into the budget via Neumann boundary conditions applied to the flux \mathbf{J}

$$\mathbf{J} \cdot \hat{\mathbf{n}} d\mathcal{S} = \text{boundary tracer transport.} \quad (52.111)$$

Note that when there is an advective/skew diffusive component to the subgrid scale flux (Chapters 49 and 50), then it adds to the resolved advective component to render a residual mean material time operator

$$\rho \frac{D^\dagger C}{Dt} = -\nabla \cdot \mathbf{J}_{\text{non-advect}}, \quad (52.112)$$

where

$$\frac{D^\dagger}{Dt} = \frac{\partial}{\partial t} + (\mathbf{v} + \mathbf{v}^*) \cdot \nabla, \quad (52.113)$$

with \mathbf{v}^* an eddy-induced velocity (see Section 50.1). For the purposes of water mass transformation analysis, we write

$$\dot{C} = \frac{D^\dagger C}{Dt}, \quad (52.114)$$

thus incorporating the eddy-induced stirring into the kinematic expression for the material time derivative.

There are many interior and boundary processes that contribute to \dot{C} within a layer. As a general expression for these contributions to the layer budget we write

$$\Delta E_C(\lambda \pm \Delta\lambda/2) = \int_{\Omega(\lambda \pm \Delta\lambda/2)} \rho \dot{C} dV = - \int_{\Omega(\lambda \pm \Delta\lambda/2)} \nabla \cdot \mathbf{J} dV \quad (52.115)$$

which is sometimes usefully decomposed into interior and surface boundary processes

$$\Delta E_C^{\text{int}}(\lambda \pm \Delta\lambda/2) = \int_{\Omega(\lambda \pm \Delta\lambda/2)} \rho \dot{C}^{\text{int}} dV \quad (52.116a)$$

$$\Delta E_C^{\text{out bdy}}(\lambda \pm \Delta\lambda/2) = - \int_{\partial\Omega_{\text{out}}(\lambda \pm \Delta\lambda/2)} \mathbf{J} \cdot \hat{\mathbf{n}} d\mathcal{S}. \quad (52.116b)$$

If the region boundary intersects the ocean bottom along $\partial\Omega_{\text{bot}}(\lambda \pm \Delta\lambda/2)$, then there is an additional bottom boundary contribution in the form

$$\Delta E_C^{\text{bot bdy}}(\lambda \pm \Delta\lambda/2) = - \int_{\partial\Omega_{\text{bot}}(\lambda \pm \Delta\lambda/2)} \mathbf{J} \cdot \hat{\mathbf{n}} d\mathcal{S}. \quad (52.117)$$

52.10.3 Transport across an interior layer interface

The surface integral in the budget (52.108) involves transport across the layer interfaces, with this transport requiring motion of the interface relative to a fluid particle. The same formalism introduced earlier can be used to compute this transport. That is, we can generalize the transformation equation (52.43d) to write

$$G_C(\lambda) = \int_{\partial\Omega(\lambda)} \rho C (\mathbf{v} - \mathbf{v}^{(b)}) \cdot \hat{\mathbf{n}} d\mathcal{S} = \frac{\partial}{\partial \lambda} \int_{\Omega(\lambda_0 \leq \lambda)} \rho \dot{\lambda} C dV. \quad (52.118)$$

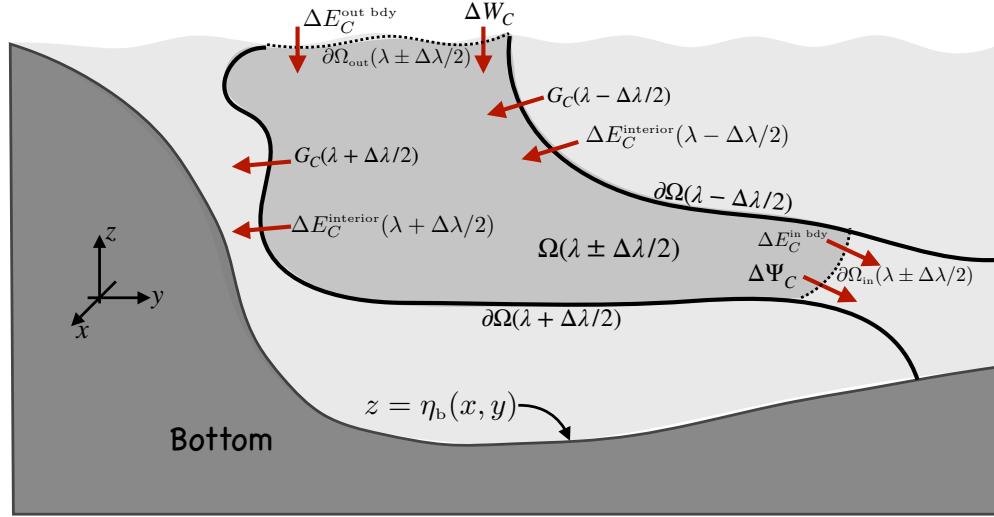


FIGURE 52.12: As for the schematic of a layer seawater mass budget depicted in Figure 52.6, we here illustrate the tracer budget within a layer of seawater with scalar property λ within the range $[\lambda - \Delta\lambda/2, \lambda + \Delta\lambda/2]$ and defined over a geographical/depth domain $\partial\Omega_{\text{in}}(\lambda \pm \Delta\lambda/2) + \partial\Omega_{\text{out}}(\lambda \pm \Delta\lambda/2) + \partial\Omega(\lambda + \Delta\lambda/2) + \partial\Omega(\lambda - \Delta\lambda/2)$. The budget for a tracer, C , over this layer is affected by the transport of tracer substance across the variety of layer boundaries. Transport processes include those determined by mixing and/or radiation across interior and surface boundaries, ΔE_C (equation (52.116b)). This term has no associated transfer of seawater mass and thus is absent from the water mass budget in Figure 52.6. Tracer budgets are also affected by processes that move seawater mass across layer boundaries: water mass transformation processes giving rise to $G_C(\lambda \pm \Delta\lambda/2)$ (equation (52.118)); transport across the surface domain boundary, ΔW_C , arising from precipitation, evaporation, runoff, and melt (equation (52.119)); and transport within the circulation crossing an interior domain boundary, $\Delta\Psi_C$ (equation (52.120)).

As a check, note that for the special case where the tracer concentration is a constant along the layer interface, then $G_C(\lambda) = C G(\lambda)$. We consider this special case in Section 52.11 when studying budgets over regions bounded by a tracer isosurface.

52.10.4 Transport across interior and surface boundaries

We now consider the impact on layer tracer mass budgets due to boundary transport. The budget contribution from mass fluxes crossing the ocean surface boundary is determined by making use of the surface kinematic boundary condition (16.71)

$$\Delta W_C = \int_{\partial\Omega_{\text{out}}(\lambda \pm \Delta\lambda/2)} \rho C (\mathbf{v} - \mathbf{v}^{(b)}) \cdot \hat{\mathbf{n}} dS = \int_{\partial\Omega_{\text{out}}(\lambda \pm \Delta\lambda/2)} Q_m C_m dA, \quad (52.119)$$

where C_m is the tracer concentration within the mass transported across the boundary.¹³ As a check, note that in the special case of a constant tracer concentration in the mass transported across the boundary, then $\Delta W_C = C_m \Delta W$, where ΔW is the water mass transported across the ocean free surface as given by equation (52.65).

For the interior open boundary the contribution is written in the generic manner

$$\Delta\Psi_C = \int_{\partial\Omega_{\text{in}}(\lambda \pm \Delta\lambda/2)} C \rho (\mathbf{v} - \mathbf{v}^{(b)}) \cdot \hat{\mathbf{n}} dS. \quad (52.120)$$

Again, in the special case where the tracer concentration is a constant along the interior boundary, C_b , then $\Delta\Psi_C = C_b \Delta\Psi$, where $\Delta\Psi$ is the seawater mass transport given by equation (52.61).

¹³Note that equation (26) in [Groeskamp et al. \(2019\)](#) incorrectly writes the integrand in equation (52.119) as $Q_m (C_m - C)$.

52.10.5 The layer tracer budget

Bringing all terms together leads to the layer tracer mass budget

$$\frac{d\Delta M_C}{dt} + \Delta\Psi_C = \Delta E_C + \Delta W_C - [G_C(\lambda + \Delta\lambda/2) - G_C(\lambda - \Delta\lambda/2)], \quad (52.121)$$

which is directly analogous to the seawater layer mass budget (52.69), with the added term ΔE_C arising from material tracer changes. As for the seawater mass budget discussed in Section 52.6.3, the layer tracer budget (52.121) provides the framework for rather general inferences about tracer transport within λ -classes.

52.10.6 Further study

Much in this section follows the treatment given by *Groeskamp et al.* (2019). This paper offers specific examples of tracer mass analysis, which is an area seeing many new applications within the ocean and atmospheric communities.

52.11 Regions bounded by a tracer contour/surface

In Section 52.10 we developed equations for a layer tracer budget where the scalar field, λ , that defines the layer is generally distinct from the tracer, C , whose budget we are studying. In this section we specialize to the case where we set $\lambda = C$ so that the region boundaries are determined by the tracer whose budget is under study. These budgets were introduced in Sections 52.6 and 52.7, and here we derive some rather useful simplifications that arise as a result of setting $\lambda = C$.

As in Section 52.10, our starting point is the Leibniz-Reynolds budgets from Section 16.3.4 for seawater mass and tracer mass computed over an arbitrary region, \mathcal{R}

$$\frac{d}{dt} \left[\int_{\mathcal{R}} \rho C dV \right] = - \int_{\partial\mathcal{R}} [\rho C (\mathbf{v} - \mathbf{v}^{(b)}) + \mathbf{J}] \cdot \hat{\mathbf{n}} dS \quad (52.122a)$$

$$\frac{d}{dt} \left[\int_{\mathcal{R}} \rho dV \right] = - \int_{\partial\mathcal{R}} [\rho (\mathbf{v} - \mathbf{v}^{(b)})] \cdot \hat{\mathbf{n}} dS. \quad (52.122b)$$

The region \mathcal{R} is rather arbitrary, and can in general be disconnected. Throughout this section we make use of the following shorthand notation for region-integrated quantities

$$M = \int_{\mathcal{R}} \rho dV \quad \text{region seawater mass} \quad (52.123a)$$

$$M_C = \int_{\mathcal{R}} C \rho dV \quad \text{region tracer mass} \quad (52.123b)$$

$$\langle C \rangle = \frac{1}{M} \int_{\mathcal{R}} C \rho dV = \frac{M_C}{M} \quad \text{region averaged tracer concentration.} \quad (52.123c)$$

52.11.1 Closed region bounded by a tracer surface/contour

Consider a closed region, $\tilde{\mathcal{R}}$, bounded by a surface of constant tracer concentration, $C = \tilde{C}$, such as depicted in Figure 52.13. The tracer budget (52.122a) for this region is given by

$$\frac{d(M \langle C \rangle)}{dt} = -\tilde{C} \int_{\partial\tilde{\mathcal{R}}} \rho (\mathbf{v} - \mathbf{v}^{(b)}) \cdot \hat{\mathbf{n}} dS - \int_{\partial\tilde{\mathcal{R}}} \mathbf{J} \cdot \hat{\mathbf{n}} dS \quad (52.124)$$

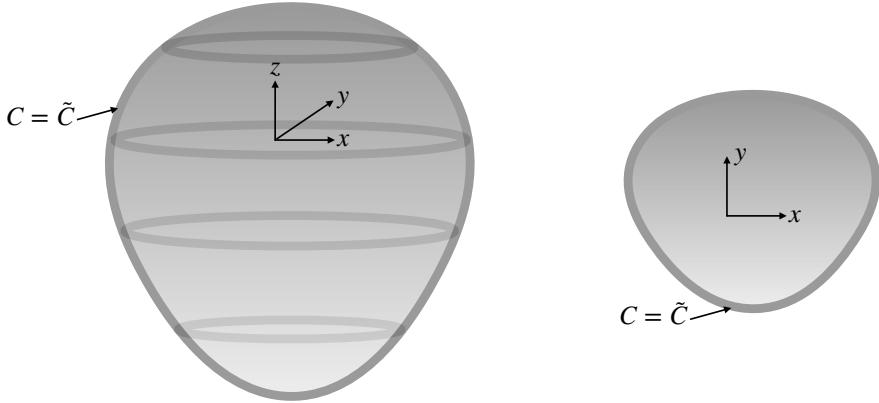


FIGURE 52.13: Left panel: a closed three-dimensional region, $\tilde{\mathcal{R}}$, with its boundary, $\partial\tilde{\mathcal{R}}$, defined by a surface of constant tracer concentration, $C = \tilde{C}$. Right panel: the analog closed two-dimensional region with its boundary defined by a contour of constant tracer concentration, $C = \tilde{C}$.

where we pulled the tracer concentration outside of the surface integral since, by construction, it is constant on the boundary, $\partial\tilde{\mathcal{R}}$. Use of the mass budget (52.122b) then leads to the rather tidy result

$$\frac{d[M(\langle C \rangle - \tilde{C})]}{dt} = - \int_{\partial\tilde{\mathcal{R}}} \mathbf{J} \cdot \hat{\mathbf{n}} dS. \quad (52.125)$$

The left hand side is the time change of the mass-weighted difference between the region averaged tracer concentration, $\langle C \rangle$, and the value of the tracer concentration defining the region boundary, \tilde{C} . These time changes are driven by a nonzero diffusive tracer transport bringing tracer mass across the region boundary. A nonzero diffusive flux on the region boundary arises only when there is a gradient of tracer concentration across that boundary. In the special case of a zero net diffusive tracer transport across the region boundary, the budget equation (52.125) reaches a steady state whereby

$$\frac{d}{dt} [M(\langle C \rangle - \tilde{C})] = 0 \iff \int_{\partial\tilde{\mathcal{R}}} \mathbf{J} \cdot \hat{\mathbf{n}} dS = 0. \quad (52.126)$$

A 3D region bounded by a constant tracer concentration is not commonly encountered in large-scale ocean and atmospheric fluids. In contrast, we often encounter quasi-2D regions as depicted in Figure 52.14, in which one may find 2D regions bounded by a closed contour of constant tracer concentration. For example, in many parts of the ocean and atmosphere transport occurs predominantly along 2D surfaces defined by a constant buoyancy. We may thus find closed contours of tracer concentrations along constant buoyancy surfaces.

To help illustrate a necessary condition to reach a steady state, consider the particular example depicted in Figure 52.14. In this figure, the tracer contour defining the region boundary is a thick shell defined by a uniform concentration $C = \tilde{C}$. The diffusive flux vanishes at each point within the boundary shell since the tracer concentration is uniform. Hence, the steady budget (52.126) leads to

$$(\langle C \rangle - \tilde{C}) \frac{dM}{dt} + M \frac{d\langle C \rangle}{dt} = 0. \quad (52.127)$$

If the total seawater mass within the region is constant, then the averaged tracer concentration is also constant, so that both terms in this steady budget vanish individually. Even so, this configuration does not reach a steady state at each point throughout the domain interior. The reason is that diffusion in the interior causes tracer to move from regions of high concentration to low concentration. Consequently, at any particular point within the domain there is an evolving

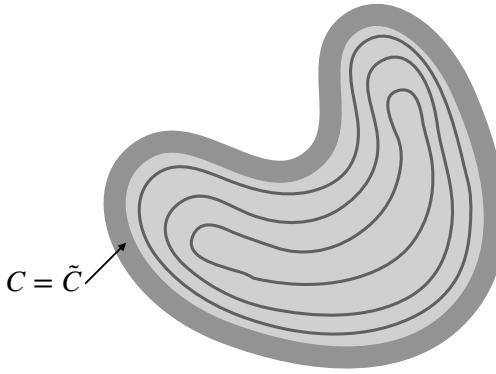


FIGURE 52.14: A two-dimensional region bounded by a finite-thick shell with constant tracer concentration, $C = \tilde{C}$. Inside the shell region the tracer concentration is not uniform.

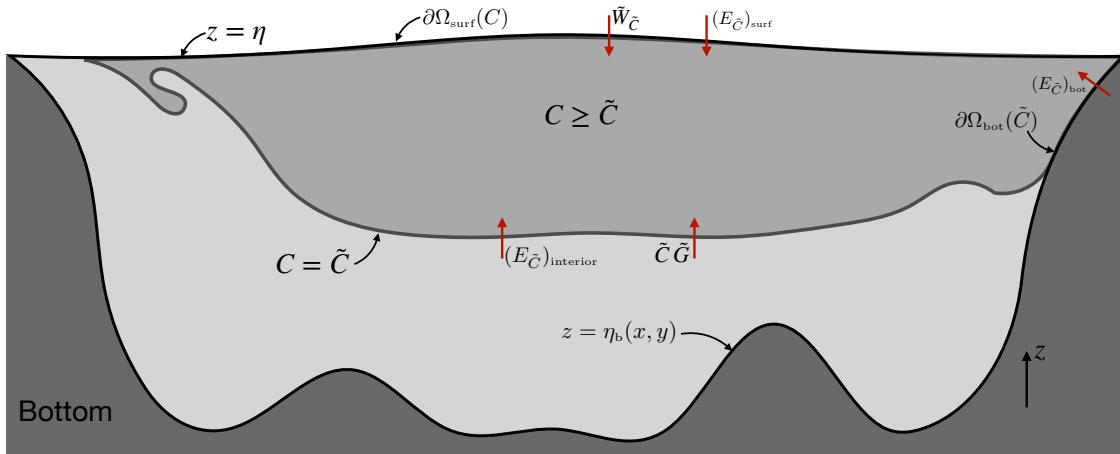


FIGURE 52.15: An ocean region where the tracer concentration is greater than a nominal value, $C \geq \tilde{C}$. A specific example is with $C = \Theta$, the Conservative Temperature, in which we are concerned with the ocean with temperature greater than $\tilde{\Theta}$. Here we depict a case where the tracer concentration generally increases upward (as with $C = \Theta$), and yet with vertical stratification not everywhere monotonic, such as for $C = \Theta$ in the high latitudes where salinity effects on density stratification become dominant. Transport processes affecting the budget of C within this region arise from mixing at the interior boundary and surface boundary, $(E_{\tilde{C}})_{\text{int}}$ and $(E_{\tilde{C}})_{\text{surf}}$, advection at the surface, $\tilde{W}_{\tilde{C}}$, and the tracer weighted water mass transformation across the interior layer boundary, $\tilde{C} \tilde{G}$.

tracer concentration. The only way for each point to reach a steady state within a region bounded by a tracer contour is for the tracer concentration to be a uniform constant throughout the region interior

$$C = \tilde{C} \quad \text{steady state tracer throughout a closed tracer region.} \quad (52.128)$$

Diffusion thus expells tracer gradients from steady state regions bounded by closed tracer contours, thus leaving a homogenous interior.

52.11.2 Region with $C \geq \tilde{C}$

As a second example of the formalism, consider the tracer budget for a region where the tracer concentration is greater than or equal to a particular tracer value as depicted in Figure 52.15. In contrast to the domain in Figure 52.12, here there is no inner boundary. We introduce the following

terms in developing both the seawater mass budget and the tracer substance budget

$$\widetilde{M} = \int_{C \geq \tilde{C}} \rho dV \quad \text{seawater mass in the region } C \geq \tilde{C} \quad (52.129\text{a})$$

$$\widetilde{M}_{\tilde{C}} = \int_{C \geq \tilde{C}} C \rho dV = \widetilde{M} \langle C \rangle \quad \text{tracer mass in the region } C \geq \tilde{C} \quad (52.129\text{b})$$

$$\widetilde{G} = - \int_{C=\tilde{C}} \rho (\mathbf{v} - \mathbf{v}^{(b)}) \cdot \hat{\mathbf{n}} d\mathcal{S} = - \int_{C=\tilde{C}} \rho w^{\text{dia}} d\mathcal{S} \quad \text{transformation across } \tilde{C} \quad (52.129\text{c})$$

$$\widetilde{G}_{\tilde{C}} = - \int_{C=\tilde{C}} C \rho (\mathbf{v} - \mathbf{v}^{(b)}) \cdot \hat{\mathbf{n}} d\mathcal{S} = - \int_{C=\tilde{C}} C \rho w^{\text{dia}} d\mathcal{S} \quad \text{tracer advected across } \tilde{C} \quad (52.129\text{d})$$

$$\widetilde{W} = - \int_{\partial\Omega_{\text{surf}}(\tilde{C})} \rho (\mathbf{v} - \mathbf{v}^{(b)}) \cdot \hat{\mathbf{n}} d\mathcal{S} = \int_{\partial\Omega_{\text{surf}}(\tilde{C})} Q_m dA \quad \text{surface mass transport} \quad (52.129\text{e})$$

$$\widetilde{W}_{\tilde{C}} = - \int_{\partial\Omega_{\text{surf}}(\tilde{C})} C \rho (\mathbf{v} - \mathbf{v}^{(b)}) \cdot \hat{\mathbf{n}} d\mathcal{S} = \int_{\partial\Omega_{\text{surf}}(\tilde{C})} C Q_m dA \quad \text{surface mass } C\text{-transport} \quad (52.129\text{f})$$

$$(E_{\tilde{C}})_{\text{int}} = - \int_{C=\tilde{C}} \mathbf{J} \cdot \hat{\mathbf{n}} d\mathcal{S} \quad \text{subgrid } C\text{-transport across } \tilde{C} \quad (52.129\text{g})$$

$$(E_{\tilde{C}})_{\text{surf}} = - \int_{\partial\Omega_{\text{surf}}(\tilde{C})} \mathbf{J} \cdot \hat{\mathbf{n}} d\mathcal{S} \quad \text{subgrid } C\text{-transport across surface} \quad (52.129\text{h})$$

$$(E_{\tilde{C}})_{\text{bot}} = - \int_{\partial\Omega_{\text{bot}}(\tilde{C})} \mathbf{J} \cdot \hat{\mathbf{n}} d\mathcal{S} \quad \text{subgrid } C\text{-transport across bottom} \quad (52.129\text{i})$$

Recall that $\hat{\mathbf{n}}$ is the *outward* normal on a boundary so that positive values for the above transports increase the tracer mass within the region. For equations (52.129c) and (52.129d), we introduced the dia-surface transport velocity according to equation (52.38) for flow across the $C = \tilde{C}$ layer interface. Likewise, for equations (52.129e) and (52.129f) we made use of the surface kinematic boundary condition (15.73c)

$$\rho (\mathbf{v} - \mathbf{v}^{(s)}) \cdot \hat{\mathbf{n}} d\mathcal{S} = -Q_m dA, \quad (52.130)$$

where Q_m is the mass transport across the free surface, with $Q_m > 0$ adding mass to the ocean, and dA is the horizontal projection of the surface area element. By inspection of Figure 52.15, the seawater mass and tracer mass budgets for this region are given by

$$\frac{d\widetilde{M}}{dt} = \widetilde{G} + \widetilde{W} \quad (52.131\text{a})$$

$$\frac{d[\widetilde{M} \langle C \rangle]}{dt} = \widetilde{C} \widetilde{G} + \widetilde{W}_{\tilde{C}} + (E_{\tilde{C}})_{\text{surf}} + (E_{\tilde{C}})_{\text{bot}} + (E_{\tilde{C}})_{\text{int}}, \quad (52.131\text{b})$$

where we assumed that no mass crosses through the solid earth. Furthermore, along the \tilde{C} -boundary we pulled the tracer concentration outside of the surface integral to write $\widetilde{G}_{\tilde{C}} = \widetilde{C} \widetilde{G}$.

Just as we did in Section 52.11.1, the tracer budget (52.131b) can be simplified by making use of the seawater mass budget (52.131a) to eliminate the water mass transformation contribution \widetilde{G} , thus rendering

$$\frac{d\widetilde{M}_{I\tilde{C}}}{dt} = [\widetilde{W}_{\tilde{C}} - \widetilde{W} \widetilde{C}] + (E_{\tilde{C}})_{\text{surf}} + (E_{\tilde{C}})_{\text{bot}} + (E_{\tilde{C}})_{\text{int}}. \quad (52.132)$$

In this equation we introduced the *internal tracer mass* according to

$$\widetilde{M}_{I\tilde{C}} \equiv \widetilde{M} (\langle C \rangle - \widetilde{C}) = \int_{C \geq \tilde{C}} (C - \widetilde{C}) \rho dV. \quad (52.133)$$

For completeness we express the internal tracer mass budget (52.132) in its integral form

$$\frac{d\widetilde{M}_{I\tilde{C}}}{dt} = \int_{\partial\Omega_{\text{out}}} \left[Q_m (C - \tilde{C}) dA - \mathbf{J} \cdot \hat{\mathbf{n}} d\mathcal{S} \right] - \int_{\partial\Omega_{\text{bot}}} \mathbf{J} \cdot \hat{\mathbf{n}} d\mathcal{S} - \int_{C=\tilde{C}} \mathbf{J} \cdot \hat{\mathbf{n}} d\mathcal{S}. \quad (52.134)$$

52.11.3 Comments and further reading

Elimination of the watermass transformation, \tilde{G} , from the internal tracer mass budget equations (52.132) and (52.134) offers a practical advantage since \tilde{G} can be rather noisy in applications. Furthermore, for some applications (e.g., see [Holmes et al. \(2019\)](#)) it is sufficient to consider the simpler budget (52.132) for internal tracer mass, rather than the budget (52.131b) for the total tracer mass.

The two-dimensional version of the result (52.128) was studied by [Rhines and Young \(1982\)](#) when studying the mechanics of isopycnal quasi-geostrophic potential vorticity in ocean gyres. We further consider their result in Exercise 48.10. It is satisfying to see their result follow from the present formalism based on Leibniz-Reynolds.

52.12 Mathematics of water mass configuration space

At the start of this chapter we introduced the notion of water mass configuration space and spoke in general terms of the properties of this space. In this section we tidy up some loose ends by presenting a mathematical structure for water mass configuration space. The presentation makes use of the *differential forms* presented in Chapter 9, so it is useful, though not essential, to be familiar with that material before reading this section.

As motivation for this section, we observe that there are occasions when we are presented with a suite of water mass properties, such as material tracer concentrations (Absolute Salinity, carbon concentration), dynamical tracers (potential vorticity), thermodynamic properties (Conservative Temperature, pressure), geographical position (latitude, longitude), geopotential, and others. Our aim is to develop a mathematical formalism enabling the study of water masses and their transformations using any combination and any number of properties. Examples with two or three water mass coordinates will be used to exemplify the formalism. Differential forms from Chapter 9 provide an important mathematical technology enabling the use of calculus on an arbitrary smooth manifold without a metric. Furthermore, we retain the space plus time split used in this book for our study of classical mechanics, so that our use of differential forms is restricted to just the spatial coordinates.¹⁴

52.12.1 Water mass configuration space

We organize the continuous valued water mass properties into an array

$$\mathbf{q} = (q^1, q^2, q^3, \dots, q^N) = q^{\bar{a}}, \quad (52.135)$$

¹⁴One can make use of unified spacetime differential forms as commonly done in relativistic physics. However, we choose the split into space plus a universal time that is far more common for Newtonian physics and that also follows our development of tensor analysis in this book (though see chapter 12 of [Misner et al. \(1973\)](#) for a discussion of a rarely used spacetime covariant Newtonian mechanics). Hence, we make use just of spatial differential forms (though with a brief exception in Section 52.12.8 when considering Lagrangian coordinates). Correspondingly, the exterior derivative operator, d (Section 9.3), only acts on spatial coordinates in this section. Hence, we can only use the Poincaré Lemma from Section 9.3 (i.e., the exterior derivative acting twice is the null operator) when the flow has no time dependence.

with $N \geq 1$ the number of properties. The superscripts signify elements of the array (they are not powers), with $q^{\bar{a}}$ an arbitrary array element. The properties, $q^{\bar{a}}$, define coordinates of a point within water mass configuration space, with the number of properties, N , determining the dimension of the space. Vectors and tensors are representable using either \mathbf{x} -space coordinates (i.e., geographical and geopotential) that are labelled x^a with $a = 1, 2, 3$, or \mathbf{q} -space coordinates that are labelled $q^{\bar{a}}$ with $\bar{a} = 1, N$. It is important to maintain a distinction between $q^{\bar{a}}$ as a coordinate in water mass configuration space versus $q^{\bar{a}}(\mathbf{x}, t)$ as a field in \mathbf{x} -space that is also a function of time.

The continuous coordinates of water mass configuration space (\mathbf{q} -space) define a smooth and oriented manifold, thus allowing us to use differential and integral calculus to study mass budgets and circulation in \mathbf{q} -space. Like all smooth manifolds, those defined by water mass configuration space locally appear like familiar Euclidean space, though globally they are generally quite different. As such, many of the formulae derived in this section have a Cartesian form reminiscent of Euclidean space. However, it is important to keep in mind that there is no metric structure imposed here, so that the associated tools from Riemannian differential geometry are missing (e.g., covariant derivatives, curvature, distance, angles).

Water mass coordinates are not generally in a 1-to-1 relation to \mathbf{x} -space. Hence, we cannot map a fluid element in \mathbf{x} -space to a unique region of \mathbf{q} -space. And as stated above, since there is generally no metric on water mass space, we cannot measure distance or angles in \mathbf{q} -space, nor can we use many facets of Cartesian tensor analysis (Chapter 1) nor those of general tensor analysis of Chapter 7, both of which assume a metric.¹⁵ Yet even without a metric, we can orient surfaces in \mathbf{q} -space through use of the anti-symmetry property of the wedge product (Section 9.2), with orientation key to deriving mass budgets in water mass configuration space.

52.12.2 Trajectories and velocity in \mathbf{q} -space

Motion of fluid elements within \mathbf{x} -space may or may not correspond to motion in \mathbf{q} -space. For example, if \mathbf{q} -space is defined by the coordinates $(q^1, q^2) = (S, \Theta)$, and if there are no boundary fluxes nor interior mixing processes, then such adiabatic and isohaline motion of a fluid element in \mathbf{x} -space leads to no motion in \mathbf{q} -space since all fluid elements retain their initial values of S and Θ . Conversely, the heating or cooling of a fluid element in a stagnant fluid may lead to zero motion in \mathbf{x} -space and yet will alter the position in \mathbf{q} -space.

We formalize the notion of a trajectory in \mathbf{q} -space following that used for \mathbf{x} -space in Section 13.3 when studying Lagrangian and Eulerian reference frames. Let $\mathbf{Q}(\mathbf{q}_0, t)$ represent the position at time, t , of an infinitesimal fluid element in \mathbf{q} -space that is labelled by its initial \mathbf{q} -space position,

$$\mathbf{Q}(\mathbf{q}_0, t = t_0) = \mathbf{q}_0. \quad (52.136)$$

The fluid element moves through \mathbf{q} -space at a rate determined by the time derivative of the trajectory. When this time derivative is evaluated at the \mathbf{q} -space point, then we have the time rate of change at that point

$$\dot{\mathbf{q}} \equiv \frac{\partial \mathbf{Q}(\mathbf{q}_0, t)}{\partial t} \quad \text{for } \mathbf{Q}(\mathbf{q}_0, t) = \mathbf{q}. \quad (52.137)$$

So in summary, $\dot{\mathbf{q}}$ measures the rate that the coordinate vector \mathbf{q} moves through water mass configuration space; i.e., $\dot{\mathbf{q}}$ is the velocity for motion in \mathbf{q} -space.

¹⁵We can *impose* a metric structure on \mathbf{q} -space *if* there is a 1-to-1 mapping between each point in \mathbf{q} -space to each point in \mathbf{x} -space. However, we aim for a mathematical description of \mathbf{q} -space that is independent of \mathbf{x} -space. Although independence may not be needed for all applications, it is a useful baseline that allows for the generalization sought in this section.

52.12.3 Boundaries in \mathbf{q} -space

Boundaries in \mathbf{x} -space are specified by the geometry of the domain containing the fluid; i.e., by the ocean bottom at $z = \eta_b(x, y)$ and surface at $z = \eta(x, y, t)$. However, water mass space has no direct information about the ocean geometry. Hence, if presented just with \mathbf{q} -space information, and if the water mass coordinates are not geometric coordinates, then we do not know if an elemental region of \mathbf{q} -space touches the ocean bottom or surface. Boundaries in \mathbf{q} -space are quite distinct from those in \mathbf{x} -space.

If a water mass coordinate is specified by a tracer concentration, then the physical value of the coordinate can range between zero and unity. Likewise, the range for temperature extends from the seawater freezing point to the boiling point, both of which depend on pressure and salinity. We can conceive of setting the water mass boundaries according to physically realized values for the water mass coordinates, with the respective boundaries generally a function of time. However, we here argue for a simpler approach by taking a lesson from an earlier approach in this chapter.

Namely, in Section 52.4 we integrated the water mass density function over a domain bounded by fixed values set well outside the physical range available to the coordinate (see in particular Section 52.4.4). We follow that approach for specifying the boundaries of non-geometric \mathbf{q} -space coordinates, thus ensuring that the seawater mass density, m , is zero at the \mathbf{q} -space boundary. This approach is analogous to setting the \mathbf{x} -space boundary to infinity, which in practice means the geometric boundary is arbitrarily far away so as to play no role in the physics of interest. Note that if a particular \mathbf{q} -space coordinate is geometric, then the boundary condition accords with the usual no-normal flux of mass at material boundaries.

52.12.4 Mass for an element of \mathbf{q} -space

We introduced the mass density function, m , in Section 52.4.3. This density specifies the seawater mass per unit “volume” for an elemental region of \mathbf{q} -space. The corresponding differential mass N -form for this region is given by

$$dM = m dV, \quad (52.138)$$

where dV is the differential volume N -form in \mathbf{q} -space

$$dV \equiv dq^1 \wedge dq^2 \wedge \dots \wedge dq^N, \quad (52.139)$$

and \wedge is the wedge product introduced in Section 9.2. For example, let $N = 3$ and $\mathbf{q} = (S, \Theta, p)$, in which case the differential mass 3-form and differential volume 3-form are given by

$$dM = m(S, \Theta, p) dV \quad (52.140a)$$

$$dV = dS \wedge d\Theta \wedge dp. \quad (52.140b)$$

To ground the above notions in something more familiar, consider the special case of a 1-to-1 mapping between \mathbf{q} -space and \mathbf{x} -space so that the mass of a particular fluid element can be written using either \mathbf{x} -space or \mathbf{q} -space coordinates

$$dM = m(\mathbf{q}) dV = \rho(\mathbf{x}) dV. \quad (52.141)$$

The 1-to-1 relation requires \mathbf{q} -space to have precisely three dimensions so that the differential volume 3-forms are given by

$$dV = dq^1 \wedge dq^2 \wedge dq^3 \quad \text{and} \quad dV = dx \wedge dy \wedge dz. \quad (52.142)$$

The volume elements are related through a coordinate transformation via

$$dV = dq^1 \wedge dq^2 \wedge dq^3 \quad (52.143a)$$

$$= \frac{\partial q^1}{\partial x^a} dx^a \wedge \frac{\partial q^2}{\partial x^b} dx^b \wedge \frac{\partial q^3}{\partial x^c} dx^c \quad (52.143b)$$

$$= \mathcal{J} dx \wedge dy \wedge dz \quad (52.143c)$$

$$= \mathcal{J} dV, \quad (52.143d)$$

where \mathcal{J} is the Jacobian of transformation (see Section 7.1.4). It thus follows that the identity (52.141) means that m and ρ are related through the Jacobian

$$m(\mathbf{q}) = \rho(\mathbf{x}) \frac{\partial(x, y, z)}{\partial(q^1, q^2, q^3)} = \rho(\mathbf{x}) \frac{\partial \mathbf{x}}{\partial \mathbf{q}} = \rho(\mathbf{x}) \mathcal{J} \quad (52.144)$$

where the second and third equalities introduced common notations for the Jacobian.

Again, the identity (52.144) holds only if there is a 1-to-1 mapping between \mathbf{q} -space and \mathbf{x} -space. In this case, the two coordinate representations refer to the same infinitesimal fluid element. More generally, a seawater element in \mathbf{q} -space does not correspond to the same seawater element in \mathbf{x} -space, in which case the Jacobian of transformation is singular. For example, consider a homogeneous ocean with $\mathbf{q} = (S, \Theta)$ where S and Θ are spatially uniform. This ocean is spread throughout \mathbf{x} -space yet is mapped to a single point in \mathbf{q} -space, so that clearly there is no 1-to-1 relation in this case. Another notion to keep in mind is that even if the fluid flow in \mathbf{x} -space is incompressible, as in a Boussinesq fluid, it is generally compressible in \mathbf{q} -space. Correspondingly, even if there are no mass sources in \mathbf{x} -space, there can be mass sources in \mathbf{q} -space.

52.12.5 The mass transport differential form

To develop a mass budget for a region fixed in \mathbf{q} -space, we introduce the mass transport differential form that measures the oriented mass transport through an $N-1$ dimensional surface. We build an understanding by considering some examples in which we encounter the generality provided by the differential forms to work in arbitrary dimensioned \mathbf{q} -space. In particular, the ability of the mass transport differential form to orient the mass transport, without making use of a normal direction, allows us to study mass transport in \mathbf{q} -space without a metric and with the same formalism holding for arbitrary dimension.

Mass transport differential 3-form

For an $N = 4$ dimensional water mass configuration space, the mass transport differential 3-form is given by

$$\begin{aligned} \mathcal{I} = m (\dot{q}^1 dq^2 \wedge dq^3 \wedge dq^4 + \dot{q}^2 dq^3 \wedge dq^4 \wedge dq^1 \\ + \dot{q}^3 dq^4 \wedge dq^1 \wedge dq^2 + \dot{q}^4 dq^1 \wedge dq^2 \wedge dq^3). \end{aligned} \quad (52.145)$$

As noted in Section 52.12.2, \dot{q}^a is the rate of change of \mathbf{q} -space coordinates, thus measuring motion of a fluid element through \mathbf{q} -space. For example, $m \dot{q}^1 dq^2 \wedge dq^3 \wedge dq^4$ measures the mass transport (mass per time) penetrating the volume \mathbf{q} -space defined by $dq^2 \wedge dq^3 \wedge dq^4$. We have a difficult time visualizing this transport given its high dimensionality. Nonetheless, the formalism is general enough to account for any number of dimensions.

Mass transport differential 2-form

Now consider $N = 3$ dimensional water mass configuration space, where the mass transport differential 2-form is given by

$$\mathcal{T} = m (\dot{q}^1 dq^2 \wedge dq^3 + \dot{q}^2 dq^3 \wedge dq^1 + \dot{q}^3 dq^1 \wedge dq^2). \quad (52.146)$$

For example, $m \dot{q}^1 dq^2 \wedge dq^3$ measures the mass transport (mass per time) penetrating the surface defined by $dq^2 \wedge dq^3$. Furthermore, if we set $\mathbf{q} = \mathbf{x}$ and invoke Cartesian vector calculus, then $\mathcal{T} = \rho \dot{x} dy \wedge dz$, which is the mass transport through the $y-z$ surface. If we assume the right hand rule then $\rho \dot{x} dy \wedge dz$ is oriented in the positive $\hat{\mathbf{x}}$ -direction.

Mass transport differential 1-form

Now consider the case of $N = 2$, in which case the mass transport differential 1-form is

$$\mathcal{T} = m (\dot{q}^1 dq^2 - \dot{q}^2 dq^1) \quad (52.147)$$

The particular case of $(q^1, q^2) = (S, \Theta)$ water mass analysis leads to the mass transport

$$\mathcal{T} = m (\dot{S} d\Theta - \dot{\Theta} dS). \quad (52.148)$$

52.12.6 Mass continuity in \mathbf{q} -space

We here develop the budget for mass contained in an elemental region fixed in \mathbf{q} -space. That is, we want to determine what affects the time derivative

$$\partial_t(dM) = \partial_t[m dV] = (\partial_t m) dV, \quad (52.149)$$

where the time derivative is computed holding the \mathbf{q} -coordinates fixed so that $\partial_t(dV) = 0$. As when developing the Eulerian mass budget for \mathbf{x} -space in Section 15.1, we presume that the mass of an elemental region of \mathbf{q} -space is affected by the convergence of mass transport into the volume of \mathbf{q} -space. As noted earlier, we do not assume a 1-to-1 mapping between fluid elements in \mathbf{q} -space and \mathbf{x} -space, so there can be mass sources in \mathbf{q} -space even if there are none in \mathbf{x} -space. Likewise, flow in \mathbf{q} -space can be compressible even if incompressible in \mathbf{x} -space.

These considerations lead us to formulate the mass budget for an elemental volume in \mathbf{q} -space in the form

$$(\partial_t m) dV = -d\mathcal{T} + M dV \quad (52.150)$$

with the \mathbf{q} -space mass source (mass per time) given by $M dV$, and with $d\mathcal{T}$ the spatial exterior derivative of the mass transport differential form.

To help understand the mass continuity equation (52.150), consider the case of $N = 2$, in which case the spatial exterior derivative of \mathcal{T} takes the form

$$d\mathcal{T} = d [m \dot{q}^1 dq^2 - m \dot{q}^2 dq^1] \quad (52.151a)$$

$$= \partial_{q^1}(m \dot{q}^1) dq^1 \wedge dq^2 - \partial_{q^2}(m \dot{q}^2) dq^2 \wedge dq^1 \quad (52.151b)$$

$$= [\partial_{q^1}(m \dot{q}^1) + \partial_{q^2}(m \dot{q}^2)] dq^1 \wedge dq^2, \quad (52.151c)$$

where we made use of the rules for computing an spatial exterior derivative as summarized in Section 9.3. The \mathbf{q} -space mass conservation equation (52.150) thus takes the form

$$\partial_t m = -\partial_{q^1}(m \dot{q}^1) - \partial_{q^2}(m \dot{q}^2) + M. \quad (52.152)$$

A few more lines of algebra leads to the result for $N = 3$

$$d\mathcal{T} = d[m \dot{q}^1 dq^2 \wedge dq^3 + m \dot{q}^2 dq^3 \wedge dq^1 + m \dot{q}^3 dq^1 \wedge dq^2] \quad (52.153a)$$

$$= [\partial_{q^1}(m \dot{q}^1) + \partial_{q^2}(m \dot{q}^2) + \partial_{q^3}(m \dot{q}^3)] dV, \quad (52.153b)$$

so that

$$\partial_t m = -\partial_{q^1}(m \dot{q}^1) - \partial_{q^2}(m \dot{q}^2) - \partial_{q^3}(m \dot{q}^3) + m. \quad (52.154)$$

These two results can be readily generalized to arbitrary dimensionality of \mathbf{q} -space thus yielding the mass continuity equation

$$\partial_t m = -\partial_{\bar{a}}(m \dot{q}^{\bar{a}}) + m = -\nabla_{\mathbf{q}} \cdot (m \dot{\mathbf{q}}) + m, \quad (52.155)$$

where we introduced the shorthand for the partial derivative operator

$$\partial_{\bar{a}} = \frac{\partial}{\partial q^{\bar{a}}}. \quad (52.156)$$

The second equality in equation (52.155) introduced the operator $\nabla_{\mathbf{q}}$ as a shorthand for the partial derivatives. This operator is a convenient shorthand but it should not be confused with the covariant divergence (e.g., equation (7.118)).¹⁶

The mass continuity equation (52.155) says that the mass density, m , changes in time within a fixed \mathbf{q} -space elemental region according to the \mathbf{q} -space convergence of the mass flux plus any mass sources. It is the natural generalization of the Cartesian coordinate continuity equation from Section 15.1, which is recovered by setting $\mathbf{q} = \mathbf{x}$, $m = \rho$, and $M = 0$ so that

$$\partial_t \rho = -\partial_x(\rho \dot{x}) - \partial_y(\rho \dot{y}) - \partial_z(\rho \dot{z}) = -\nabla \cdot (\rho \mathbf{v}). \quad (52.157)$$

We emphasize that derivation of the \mathbf{q} -space mass continuity equation (52.155) made no use of \mathbf{x} -space nor any metric structure. So although it is intuitive to conceive of the transport as moving through tiny hypercubes formed by the \mathbf{q} -space coordinate increments, that picture presumes a metric (to define lengths and angles) and yet \mathbf{q} -space generally has no metric.

52.12.7 Streamfunction for steady and source-free circulation

In the absence of \mathbf{q} -space mass sources and for a steady state, the continuity equation (52.150) says that the mass transport differential form has zero spatial exterior derivative

$$d\mathcal{T} = 0. \quad (52.158)$$

The Poincaré Lemma from Section 9.3 then means that we can introduce a streamfunction so that

$$\mathcal{T} = d\psi. \quad (52.159)$$

Exposing tensor indices reveals that the steady and source free mass continuity equation (52.155) means that the \mathbf{q} -space mass flux satisfies

$$\nabla_{\mathbf{q}} \cdot (m \dot{\mathbf{q}}) = \partial_{\bar{a}}(m \dot{q}^{\bar{a}}) = 0. \quad (52.160)$$

¹⁶It is only for Euclidean space with Cartesian coordinates that the covariant divergence simplifies to the partial derivative operator.

The corresponding \mathbf{q} -space streamfunction has an expression that depends on the dimension of \mathbf{q} -space. For example, the cases of $N = 2$ and $N = 3$ take the form

$$N = 2 : m \dot{q}^{\bar{a}} = \epsilon^{\bar{a}\bar{b}} \partial_{\bar{b}} \psi \quad (52.161a)$$

$$N = 3 : m \dot{q}^{\bar{a}} = \epsilon^{\bar{a}\bar{b}\bar{c}} \partial_{\bar{b}} \psi_{\bar{c}}, \quad (52.161b)$$

where $\epsilon^{\bar{a}\bar{b}}$ and $\epsilon^{\bar{a}\bar{b}\bar{c}}$ are the totally anti-symmetric permutation symbols for $N = 2$ and $N = 3$ dimensions.¹⁷ The physical dimensions of the streamfunctions are

$$N = 2 : \psi \equiv M/T \quad (52.162a)$$

$$N = 3 : \psi_{\bar{a}} \equiv M/(T q^{\bar{a}}), \quad (52.162b)$$

where we note that the dimension for the $N = 3$ streamfunction depends on the dimension of the water mass coordinates.

52.12.8 Mass continuity when \mathbf{q} -space and \mathbf{x} -space are 1-to-1

We now consider mass continuity in the special case of a 1-to-1 relation between \mathbf{q} -space and \mathbf{x} -space in which a fluid element in \mathbf{q} -space refers to the same fluid element in \mathbf{x} -space. Correspondingly, both spaces have $N = 3$ dimensions.

With a 1-to-1 relation between \mathbf{q} -space and \mathbf{x} -space, the mass density functions are related by equation (52.144) so that

$$m = \mathcal{J} \rho \quad (52.163)$$

with $\mathcal{J} = \partial \mathbf{x} / \partial \mathbf{q}$ the nonsingular Jacobian of transformation from equation (52.144). Analogously, the mass source in \mathbf{q} -space is related to an \mathbf{x} -space mass source via

$$m dV = \mathfrak{m} dV \implies m = \mathcal{J} \mathfrak{m}, \quad (52.164)$$

with \mathfrak{m} the mass source per time per \mathbf{x} -space volume. The mass conservation equation (52.155) thus takes the form

$$\partial_t(\mathcal{J} \rho) = -\partial_{\bar{a}}(\mathcal{J} \rho \dot{q}^{\bar{a}}) + \mathcal{J} \mathfrak{m}. \quad (52.165)$$

Lagrangian \mathbf{q} -coordinates¹⁸

Assume the 1-to-1 relation is between Eulerian (\mathbf{x} -space) and Lagrangian (\mathbf{q} -space) coordinates. In this case, the velocity of a fluid element at a point in \mathbf{x} -space is given by the time derivative of a trajectory, $\mathbf{X}(\mathbf{q}, t)$, holding $q^{\bar{a}}$ fixed and evaluated at $\mathbf{X} = \mathbf{x}$

$$v^a = \frac{\partial X^a(\mathbf{q}, t)}{\partial t} \quad \text{for } \mathbf{X} = \mathbf{x}, \quad (52.166)$$

so that we can write

$$dx^a - v^a dt = \frac{\partial x^a}{\partial t} + \frac{\partial x^a}{\partial q^{\bar{a}}} dq^{\bar{a}} - v^a dt = \frac{\partial x^a}{\partial q^{\bar{a}}} dq^{\bar{a}}. \quad (52.167)$$

¹⁷See Sections 1.4.1 and 7.6 for a discussion of the permutation symbol. In particular, the numerical values of the permutation symbol are the same whether indices are raised or lowered.

¹⁸The Lagrangian coordinate presentation given here largely follows Section 10.6 of *Flanders (1989)*.

Here we introduced the time coordinate 1-form, dt , though only as an intermediate part of the derivation. Evaluating the trajectories at a point in \mathbf{x} -space leads to the expression for the volume differential 3-form

$$(dx^1 - v^1 dt) \wedge (dx^2 - v^2 dt) \wedge (dx^3 - v^3 dt) = \frac{\partial x^a}{\partial q^{\bar{a}}} \frac{\partial x^b}{\partial q^{\bar{b}}} \frac{\partial x^c}{\partial q^{\bar{c}}} dq^{\bar{a}} \wedge dq^{\bar{b}} \wedge dq^{\bar{c}} \quad (52.168a)$$

$$= \mathcal{J} dq^1 \wedge dq^2 \wedge dq^3 \quad (52.168b)$$

$$= dV, \quad (52.168c)$$

where we made use of equation (52.143d) that relates the volume differential 3-form in \mathbf{q} -space to that in \mathbf{x} -space. Assuming no mass sources, the mass of a fluid element

$$dM = \rho dV = \rho \mathcal{J} dV \quad (52.169)$$

remains constant following a trajectory. Additionally, the Lagrangian volume element, dV , remains constant following a trajectory. We thus see that in Lagrangian coordinates the \mathbf{q} -space mass density is constant

$$\partial_t(\rho \mathcal{J}) = \partial_t m = 0. \quad (52.170)$$

We understand this result by noting that the \mathbf{q} -coordinates are Lagrangian so that both the mass and the \mathbf{q} -space volume of the fluid element are fixed along a trajectory. Additionally, this result is consistent with the general continuity equation (52.155) with $\dot{q}^{\bar{a}} = 0$ and $m = 0$.

Tracer \mathbf{q} -coordinates

Now assume the \mathbf{q} -space coordinates are each tracers that maintain a 1-to-1 relation with \mathbf{x} -space. Hence, the material time evolution for the water mass coordinates is given by the tracer equation

$$\rho \dot{q}^{\bar{a}} = -\mathcal{J}^{-1} \partial_{\bar{b}}(\mathcal{J} J^{\bar{a}\bar{b}}) + \mathcal{S}^{\bar{a}} \implies m \dot{q}^{\bar{a}} = -\partial_{\bar{b}}(\mathcal{J} J^{\bar{a}\bar{b}}) + \mathcal{J} \mathcal{S}^{\bar{a}}, \quad (52.171)$$

with $\mathcal{S}^{\bar{a}}$ a source for $q^{\bar{a}}$. The first term on the right hand side is the covariant convergence (see equation (7.118)) of the tracer flux, $\mathbf{J}^{\bar{a}}$. Furthermore, the tracer flux has \mathbf{q} -space components, $J^{\bar{a}\bar{b}}$, related to \mathbf{x} -space components via¹⁹

$$J^{\bar{a}\bar{b}} = \frac{\partial q^{\bar{b}}}{\partial x^a} J^{\bar{a}a}. \quad (52.172)$$

Following the discussion of subgrid scale tracer mixing in Section 50.1, we assume all tracer fluxes in \mathbf{x} -space are determined by the *same* symmetric mixing tensor²⁰

$$\mathbf{J}^{\bar{a}a} = -\mathbb{K}^{ac} \frac{\partial q^{\bar{a}}}{\partial x^c}, \quad (52.173)$$

in which case the \mathbf{q} -space components to the tracer flux take the form

$$J^{\bar{a}\bar{b}} = -\frac{\partial q^{\bar{b}}}{\partial x^a} \mathbb{K}^{ac} \frac{\partial q^{\bar{a}}}{\partial x^c} = -\mathbb{K}^{\bar{b}\bar{a}} = -\mathbb{K}^{\bar{a}\bar{b}} = J^{\bar{b}\bar{a}}. \quad (52.174)$$

¹⁹See Section 7.1 for details of coordinate transformations.

²⁰Double-diffusion leads to a distinct diffusion tensor for Θ and S . Otherwise, it is common to assume the same diffusion tensor for all tracers.

The second equality expresses the components to the diffusion tensor in terms of tracer coordinates, and the following equalities follow from symmetry of the tracer diffusion tensor as expressed in either \mathbf{x} -space coordinates or \mathbf{q} -space coordinates

$$\mathbb{K}^{ac} = \mathbb{K}^{ca} \quad \text{and} \quad \mathbb{K}^{\bar{a}\bar{b}} = \mathbb{K}^{\bar{b}\bar{a}}. \quad (52.175)$$

We thus see that the tracer flux when written in tracer coordinates is minus the diffusion tensor as written in tracer coordinates.

Making use of these results brings to the continuity equation (52.165) into the tracer space form

$$\partial_t(\mathcal{J}\rho) = \partial_{\bar{a}} \left[\partial_{\bar{b}}(\mathcal{J} J^{\bar{a}\bar{b}}) - \mathcal{J} S^{\bar{a}} \right] + \mathcal{J} m, \quad (52.176)$$

which can be written in the equivalent manner

$$\partial_t m = -\partial_{\bar{a}} \left[\partial_{\bar{b}}(\mathcal{J} K^{\bar{a}\bar{b}}) + \mathcal{J} S^{\bar{a}} \right] + m. \quad (52.177)$$

This equation explicitly connects the evolution of mass within tracer space to the mixing of tracers, which moves mass across the tracer contours, plus any sources of tracer and mass.

Mass continuity (52.177) in the steady state with zero tracer sources and zero mass sources is given by

$$\partial_{\bar{a}\bar{b}}(\mathcal{J} K^{\bar{a}\bar{b}}) = -\partial_{\bar{a}\bar{b}}(\mathcal{J} J^{\bar{a}\bar{b}}) = 0. \quad (52.178)$$

Connecting to the $N = 3$ streamfunction in Section 52.12.7 leads to

$$m \dot{q}^{\bar{a}} = \epsilon^{\bar{a}\bar{b}\bar{c}} \partial_{\bar{b}} \psi_{\bar{c}} = \partial_{\bar{b}}(\mathcal{J} K^{\bar{a}\bar{b}}) \iff m \dot{\mathbf{q}} = \nabla_{\mathbf{q}} \wedge \boldsymbol{\psi} = \nabla_{\mathbf{q}} \cdot (\mathcal{J} \bar{\mathbb{K}}), \quad (52.179)$$

with the second equality introducing standard vector notion and is to be read as a shorthand for the first equality. The overline on the diffusion tensor indicates that it is represented using tracer coordinates as defined by the first equality in equation (52.179). This equation means that streamfunction for steady circulation in tracer space is directly related to the tracer diffusion tensor. This relation follows since tracer mixing generates circulation in tracer space, and mixing is here parameterized by a symmetric diffusion tensor.

52.12.9 Water mass angular momentum

Circulation in water mass configuration space commonly exhibits some form of rotational motion, particularly if it reaches a steady state whereby the circulation closes. We are thus motivated to introduce a measure of this circulation by studying the *water mass angular momentum* differential form,²¹ with our focus on water mass configuration spaces with $N = 2$ and $N = 3$

$$N = 2 : \quad L = \epsilon_{\bar{b}\bar{c}} q^{\bar{b}} \dot{q}^{\bar{c}} m \, dV \quad (52.180a)$$

$$N = 3 : \quad L_{\bar{a}} = \epsilon_{\bar{a}\bar{b}\bar{c}} q^{\bar{b}} \dot{q}^{\bar{c}} m \, dV. \quad (52.180b)$$

These definitions of angular momentum satisfy the following properties.

- The physical dimensions of the water mass angular momentum depend on the dimensions of the water mass coordinates, and the different $L_{\bar{a}}$ components for the $N = 3$ case generally having distinct dimensions.

²¹In equations (52.180a) and (52.180b) we see permutation symbols with indices downstairs. However, as noted in Section 7.6, the permutation symbol has the same value when its tensor indices are raised or lowered: $\epsilon_{\bar{a}\bar{b}\bar{c}} = \epsilon^{\bar{a}\bar{b}\bar{c}}$.

- For $N = 3$ the angular momentum satisfies

$$q^{\bar{a}} L_{\bar{a}} = \epsilon_{\bar{a}\bar{b}\bar{c}} q^{\bar{a}} q^{\bar{b}} \dot{q}^{\bar{c}} m dV = 0, \quad (52.181)$$

which also holds for $\mathbf{q} = \mathbf{x}$, in which case $L_{\bar{a}}$ equals to the Newtonian angular momentum from Section 19.3.

- Just as for Newtonian angular momentum, a shift in the definition of a water mass coordinate, such as for Absolute Temperature shifted from Kelvin to Celsius, shifts the value of angular momentum.

As an example, consider the coordinates $(q^1, q^2, q^3) = (S, \Theta, p)$, in which case the water mass angular momentum has the following components

$$L_1 = (\Theta \dot{p} - p \dot{\Theta}) m dS \wedge d\Theta \wedge dp \quad (52.182a)$$

$$L_2 = (p \dot{S} - S \dot{p}) m dS \wedge d\Theta \wedge dp \quad (52.182b)$$

$$L_3 = (S \dot{\Theta} - \Theta \dot{S}) m dS \wedge d\Theta \wedge dp. \quad (52.182c)$$

Likewise, the $N = 2$ configuration space with $(q^1, q^2) = (S, \Theta)$ has the angular momentum

$$L = (S \dot{\Theta} - \Theta \dot{S}) m dS \wedge d\Theta. \quad (52.183)$$

Steady and source-free water mass angular momentum with $N = 2$

As seen in Section 52.12.7, we can relate $m \dot{q}^{\bar{c}}$ to a water mass space streamfunction when the flow is steady and there are no \mathbf{q} -space mass sources. For $N = 2$ we have $m \dot{q}^{\bar{c}} = \epsilon^{\bar{c}\bar{d}} \partial_{\bar{d}} \psi$, so that the steady state angular momentum is given by

$$L = \epsilon_{\bar{b}\bar{c}} q^{\bar{b}} \dot{q}^{\bar{c}} m dV \quad (52.184a)$$

$$= \epsilon_{\bar{b}\bar{c}} q^{\bar{b}} \epsilon^{\bar{c}\bar{d}} \partial_{\bar{d}} \psi dV \quad (52.184b)$$

$$= -(\epsilon_{1\bar{b}} \epsilon^{1\bar{d}} + \epsilon_{\bar{b}2} \epsilon^{\bar{d}2}) q^{\bar{b}} \partial_{\bar{d}} \psi dV \quad (52.184c)$$

$$= -(q^1 \partial \psi / \partial q^1 + q^2 \partial \psi / \partial q^2) dV \quad (52.184d)$$

$$= -\mathbf{q} \cdot \nabla_{\mathbf{q}} \psi dV \quad (52.184e)$$

$$= -\nabla_{\mathbf{q}} \cdot (\mathbf{q} \psi) dV + 2\psi dV, \quad (52.184f)$$

where in the final step we used $\nabla_{\mathbf{q}} \cdot \mathbf{q} = \partial q^1 / \partial q^1 + \partial q^2 / \partial q^2 = 2$. Recall from Section 52.12.3 that the boundaries of water mass configuration space are assumed to be outside the range where seawater exists. Hence, if we integrate the angular momentum (52.184f) over all of water mass space, then the total derivative term drops out since the streamfunction vanishes at the boundaries. We are thus led to

$$\int L = 2 \int \psi dV = 2 \int \psi dq^1 dq^2, \quad (52.185)$$

so that the integrated steady state water mass angular momentum is given by twice the integrated circulation streamfunction. To within an arbitrary sign, this result accords with the angular momentum computed for a non-divergent barotropic fluid as considered in Exercise 34.4.

Steady and source-free water mass angular momentum with $N = 3$

For $N = 3$ the water mass streamfunction is determined by $m \dot{q}^{\bar{c}} = \epsilon^{\bar{c}\bar{d}\bar{e}} \partial_{\bar{d}} \psi_{\bar{e}}$, so that the steady and source-free water mass angular momentum is²²

$$L_{\bar{a}} = \epsilon_{\bar{a}\bar{b}\bar{c}} q^{\bar{b}} \dot{q}^{\bar{c}} m dV \quad (52.186a)$$

$$= \epsilon_{\bar{c}\bar{a}\bar{b}} q^{\bar{b}} \epsilon^{\bar{c}\bar{d}\bar{e}} \partial_{\bar{d}} \psi_{\bar{e}} dV \quad (52.186b)$$

$$= (\delta_{\bar{a}}^{\bar{d}} \delta_{\bar{b}}^{\bar{e}} - \delta_{\bar{a}}^{\bar{e}} \delta_{\bar{b}}^{\bar{d}}) q^{\bar{b}} \partial_{\bar{d}} \psi_{\bar{e}} dV \quad (52.186c)$$

$$= (q^{\bar{e}} \partial_{\bar{a}} \psi_{\bar{e}} - q^{\bar{d}} \partial_{\bar{d}} \psi_{\bar{a}}) dV, \quad (52.186d)$$

$$= q^{\bar{b}} (\partial_{\bar{a}} \psi_{\bar{b}} - \partial_{\bar{b}} \psi_{\bar{a}}) dV, \quad (52.186e)$$

where the third equality made use of the identity (1.38) to relate the contraction of permutation symbols to the Kronecker deltas.

Tracer space angular momentum

We now study the case where \mathbf{q} -space (i) has $N = 3$ dimensions, (ii) is 1-to-1 related to \mathbf{x} -space, and (iii) all of the \mathbf{q} -coordinates are tracers. In this case we make use of the tracer equation (52.171) to write

$$L_{\bar{a}} = \epsilon_{\bar{a}\bar{b}\bar{c}} q^{\bar{b}} \left[-\partial_{\bar{d}} (\mathcal{J} J^{\bar{c}\bar{d}}) + \mathcal{J} S^{\bar{c}} \right] dV \quad (52.187a)$$

$$= \epsilon_{\bar{a}\bar{b}\bar{c}} \left[-\partial_{\bar{d}} (\mathcal{J} J^{\bar{c}\bar{d}} q^{\bar{b}}) + \partial_{\bar{d}} q^{\bar{b}} \mathcal{J} J^{\bar{c}\bar{d}} + \mathcal{J} S^{\bar{c}} q^{\bar{b}} \right] dV \quad (52.187b)$$

$$= \epsilon_{\bar{a}\bar{b}\bar{c}} \left[-\partial_{\bar{d}} (\mathcal{J} J^{\bar{c}\bar{d}} q^{\bar{b}}) + \mathcal{J} S^{\bar{c}} q^{\bar{b}} \right] dV. \quad (52.187c)$$

For the final step we made use of the identity

$$\epsilon_{\bar{a}\bar{b}\bar{c}} \partial_{\bar{d}} q^{\bar{b}} J^{\bar{c}\bar{d}} = \epsilon_{\bar{a}\bar{b}\bar{c}} \delta_{\bar{d}}^{\bar{b}} J^{\bar{c}\bar{d}} = \epsilon_{\bar{a}\bar{b}\bar{c}} J^{\bar{c}\bar{b}} = 0, \quad (52.188)$$

which follows from anti-symmetry of $\epsilon_{\bar{a}\bar{b}\bar{c}}$ along with symmetry of $J^{\bar{c}\bar{d}}$ as per equation (52.174). We thus conclude that each component of the tracer angular momentum is determined by the tracer-space flux convergence plus a source

$$L_{\bar{a}} = \epsilon_{\bar{a}\bar{b}\bar{c}} \left[-\partial_{\bar{d}} (\mathcal{J} J^{\bar{c}\bar{d}} q^{\bar{b}}) + \mathcal{J} S^{\bar{c}} q^{\bar{b}} \right] dV. \quad (52.189)$$

Integrating over all of tracer space, and assuming zero fluxes at the tracer boundaries (see Section 52.12.3), we find the tracer integral of the tracer angular momentum is determined only by the tracer sources.

$$\int L_{\bar{a}} = \epsilon_{\bar{a}\bar{b}\bar{c}} \int \mathcal{J} S^{\bar{c}} q^{\bar{b}} dV. \quad (52.190)$$

As tracer sources commonly vanish, we conclude that the global tracer space integral for each of the three components to the tracer angular momentum is identically zero when the source for the complementary tracers vanish

$$\int L_{\bar{a}} = 0 \quad \text{if } S^{\bar{c}} = 0 \text{ for } \bar{c} \neq \bar{a}. \quad (52.191)$$

²²Signed messed up!

It is remarkable that the integrated tracer angular momentum remains zero in the absence of tracer sources. Again, this result holds so long as the diffusion tensor is symmetric and the same diffusion tensor used by each of the three tracers. There is some precedent for this result from our study of \mathbf{x} -space angular momentum in Section 19.3. In that section we argued that there are no internal torques within a Newtonian fluid, which then constrains the stress tensor to be symmetric. So a symmetric stress tensor cannot alter the angular momentum within a fluid; it can only do so through boundary effects. Likewise, a symmetric diffusion tensor cannot alter the tracer angular momentum of a fluid. So if the fluid starts with a zero angular momentum then it remains zero.

Considering the global invariance of tracer $L_{\bar{a}}$ using \mathbf{x} -space

Consider the above results by working with the \mathbf{x} -space expression of the tracer equation. In this case the tracer angular momentum is given by

$$L_{\bar{a}} = \epsilon_{\bar{a}\bar{b}\bar{c}} q^{\bar{b}} (-\nabla \cdot \mathbf{J}^{\bar{c}} + \mathcal{S}^{\bar{c}}) dV \quad (52.192a)$$

$$= \epsilon_{\bar{a}\bar{b}\bar{c}} \left[-\nabla \cdot (q^{\bar{b}} \mathbf{J}^{\bar{c}}) + \nabla q^{\bar{b}} \cdot \mathbf{J}^{\bar{c}} + q^{\bar{b}} \mathcal{S}^{\bar{c}} \right] dV \quad (52.192b)$$

$$= \epsilon_{\bar{a}\bar{b}\bar{c}} \left[-\nabla \cdot (q^{\bar{b}} \mathbf{J}^{\bar{c}}) + q^{\bar{b}} \mathcal{S}^{\bar{c}} \right] dV, \quad (52.192c)$$

where we set

$$\epsilon_{\bar{a}\bar{b}\bar{c}} \nabla q^{\bar{b}} \cdot \mathbf{J}^{\bar{c}} = -\epsilon_{\bar{a}\bar{b}\bar{c}} \nabla q^{\bar{b}} \cdot \mathbb{K} \cdot \nabla q^{\bar{c}} = 0. \quad (52.193)$$

We again see that the global integral of the tracer angular momentum reduces to contributions from sources plus boundary fluxes.



Bibliography

- Abernathay, R., D. Ferreira, and A. Klocker, Diagnostics of isopycnal mixing in a circumpolar channel, *Ocean Modelling*, 72(0), 1 – 16, doi:10.1016/j.ocemod.2013.07.004, 2013. 1155, 1159, 1168
- Acheson, D., *Elementary Fluid Dynamics*, Oxford Applied Mathematics and Computing Science Series, Oxford, Oxford, 1990. iv, 28, 833
- Adcroft, A., R. Hallberg, and M. Harrison, A finite volume discretization of the pressure gradient force using analytic integration, *Ocean Modelling*, 22, 106–113, doi:10.1016/j.ocemod.2008.02.001, 2008. 1031, 1032, 1034
- Aiki, H., T. Jacobson, and T. Yamagata, Parameterizing ocean eddy transports from surface to bottom, *Journal of Geophysical Research*, 31, L19 302, doi:10.1029/2004GL020703, 2004. 1151
- Anderson, J., Ludwig Prandtl's boundary layer, *Physics Today*, 58, 42–48, doi:10.1063/1.2169443, 2005. 395, 397
- Andresen, B., R. Berry, R. Gilmore, E. Ihrig, and P. Salamon, Thermodynamic geometry and the metrics of Weinhold and Gilmore, *Physical Review A*, 37(3), 845–848, doi:10.1103/PhysRevA.37.845, 1988. 427
- Andrews, D., and M. McIntyre, An exact theory of nonlinear waves on a Lagrangian-mean flow, *Journal of Fluid Mechanics*, 89, 609–646, 1978a. 1115, 1124
- Andrews, D. G., A finite-amplitude Eliassen-Palm theorem in isentropic coordinates, *Journal of Atmospheric Sciences*, 40, 1877–1883, doi:10.1175/1520-0469(1983)040<1877:AFAEPT>2.0.CO;2, 1983. 539
- Andrews, D. G., and M. E. McIntyre, On wave action and its relatives, *Journal of Fluid Mechanics*, 89, 647–664, 1978b. 1115, 1124
- Anstey, J., and L. Zanna, A deformation-based parametrization of ocean mesoscale eddy Reynolds stresses, *Ocean Modelling*, 112, 99–111, 2017. 391
- Apel, J. R., *Principles of Ocean Physics*, International Geophysics Series, vol. 38, Academic Press, London, 1987. iv, 194, 366, 621, 643

- Arakawa, A., and V. R. Lamb, The UCLA general circulation model, in *Methods in Computational Physics: General Circulation Models of the Atmosphere*, vol. 17, edited by J. Chang, pp. 174–265, Academic Press, 1977. 758
- Arbic, B., S. T. Garner, R. W. Hallberg, and H. L. Simmons, The accuracy of surface elevations in forward global barotropic and baroclinic tide models, *Deep Sea Research*, 51, 3069—3101, 2004. 644
- Aris, R., *Vectors, Tensors and the Basic Equations of Fluid Mechanics*, Dover Publishing, New York, 1962. iv, 5, 129, 241, 265, 267, 278, 286, 311, 317, 357, 366, 376, 386, 391
- Badin, G., and F. Crisciani, *Variational formulation of fluid and geophysical fluid dynamics*, Springer International Publishing, Switzerland, 218 pages, 2018. 259
- Barthel, A., A. Hogg, S. Waterman, and S. Keating, Jet-topography interactions affect energy pathways to the deep Southern Ocean, *Journal of Physical Oceanography*, 47(1799–1816), doi: 10.1175/JPO-D-16-0220.1, 2017. 714
- Batchelor, G. K., *An Introduction to Fluid Dynamics*, Cambridge University Press, Cambridge, England, 1967. iv, 234, 240, 269, 297, 376, 403, 470
- Becker, E., Frictional heating in global climate models, *Monthly Weather Review*, 131, 508–520, doi:10.1175/1520-0493(2003)131<0508:FHIGCM>2.0.CO;2, 2003. 472
- Beckers, J.-M., H. Burchard, J.-M. Campin, E. Deleersnijder, and P. P. Mathieu, Another reason why simple discretizations of rotated diffusion operators cause problems in ocean models: Comments on isoneutral diffusion in a z -coordinate ocean model, *Journal of Physical Oceanography*, 28, 1552–1559, 1998. 1157
- Beckers, J.-M., H. Burchard, E. Deleersnijder, and P. P. Mathieu, Numerical discretization of rotated diffusion operators in ocean models, *Monthly Weather Review*, 128, 2711–2733, 2000. 1157
- Bennett, A., The geometry of neutral paths, *Journal of Physical Oceanography*, 49, 3037–3044, doi:10.1175/JPO-D-19-0047.1, 2019. 569
- Bishop, S., R. Small, and F. Bryan, The global sink of available potential energy by mesoscale air-sea interaction, *Journal of Advances in Modeling Earth Systems*, in review, 2020. 544
- Bleck, R., Finite difference equations in generalized vertical coordinates. Part I: Total energy conservation, *Contributions to Atmospheric Physics*, 51, 360–372, 1978. 982
- Böning, C. W., W. R. Holland, F. O. Bryan, G. Danabasoglu, and J. C. McWilliams, An overlooked problem in model simulations of the thermohaline circulation and heat transport in the Atlantic Ocean, *Journal of Physical Oceanography*, 8, 515–523, 1995. 1143
- Bretherton, C., and C. Schär, Flux of potential vorticity substance: a simple derivation and uniqueness property, *Journal of the Atmospheric Sciences*, 50, 1834–1836, doi:10.1175/1520-0469(1993)050<1834:FOPVSA>2.0.CO;2, 1993. 901, 904
- Bretherton, F. P., Critical layer instability in baroclinic flows, *Quarterly Journal of the Royal Meteorological Society*, 92, 325–334, 1966. 90, 539

- Brown, E., *Waves, tides, and shallow-water processes*, 227 pp., The Open University, Milton Keys, UK, 1999. [643](#)
- Bryan, K., A numerical method for the study of the circulation of the world ocean, *Journal of Computational Physics*, 4, 347–376, 1969. [775](#), [956](#)
- Bühler, O., *Waves and mean flows*, 2nd ed., Cambridge University Press, Cambridge, UK, doi: 10.1017/CBO9781107478701, 2014a. [41](#), [663](#), [1115](#), [1124](#)
- Bühler, O., A gentle stroll through EP flux theory, *European Journal of Mechanics B*, 47, 12–15, doi:10.1016/j.euromechflu.2014.01.010, 2014b. [1051](#)
- Bühler, O., J. Callies, and R. Ferrari, Wave–vortex decomposition of one-dimensional ship-track data, *Journal of Fluid Mechanics*, 756, 1007–1026, doi:10.1017/jfm.2014.488, 2014. [546](#)
- Callen, H. B., *Thermodynamics and an Introduction to Thermostatics*, John Wiley and Sons, New York, 493 + xvi pp, 1985. [423](#), [426](#), [427](#), [472](#), [478](#), [480](#), [990](#)
- Cane, M., V. Kamenkovich, and A. Krupitsky, On the utility and disutility of JEBAR, *Journal of Physical Oceanography*, 28, 519–526, doi:10.1175/1520-0485(1998)028<0519:OTUADO>2.0.CO;2, 1998. [957](#)
- Chaikin, P. M., and T. C. Lubensky, *Principles of Condensed Matter Physics*, Cambridge University Press, Cambridge, United Kingdom, 1995. [109](#)
- Chandrasekhar, S., *Hydrodynamic and Hydromagnetic Stability*, Dover Publications, New York, 654 pp, 1961. [519](#)
- Charney, J., R. Fjörtoft, and J. von Neumann, Numerical integration of the barotropic vorticity equation, *Tellus*, 2, 237–254, doi:10.3402/tellusa.v2i4.8607, 1950. [761](#), [769](#)
- Chatwin, P. C., The vorticity equation as an angular momentum equation, *Mathematical Proceedings of the Cambridge Philosophical Society*, 74, 365–367, doi:10.1017/S0305004100048131, 1973. [877](#)
- Cole, S. T., C. Wortham, E. Kunze, and W. B. Owens, Eddy stirring and horizontal diffusivity from argo float observations: Geographic and depth variability, *Geophysical Research Letters*, 42(10), 3989–3997, doi:10.1002/2015GL063827, 2015GL063827, 2015. [1168](#)
- Courant, R., and D. Hilbert, *Methods of Mathematical Physics Volume I*, Wiley-Interscience, New York, 1953. [1075](#)
- Courant, R., and D. Hilbert, *Methods of Mathematical Physics Volume II: Partial Differential Equations*, Wiley-Interscience, 1962. [1075](#)
- Cox, M. D., Isopycnal diffusion in a z -coordinate ocean model, *Ocean Modelling*, 74, 1–5, 1987. [1157](#), [1158](#)
- Crocco, L., Eine neue stromfunktion fur die erforschung der bewegung der gas emit rotation, *Zamm-Zeitschrift Fur Angewandte Mathematik Und Mechanik*, 17, 1, doi:10.1002/ZAMM.19370170103, 1937. [488](#)
- Cushman-Roisin, B., Subduction, in *Dynamics of the oceanic surface mixed-layer*, pp. 181–196, Hawaii Institute of Geophysical Special Publications, 1987. [1018](#)

- Cushman-Roisin, B., and J.-M. Beckers, *Introduction to Geophysical Fluid Dynamics*, Academic Press, Amsterdam, 828, 2011. iv, 620, 621, 684, 762
- Dahler, J., and L. Scriven, The angular momentum of continua, *Nature*, 4797, 36–37, 1961. 376
- Davidson, P., *Turbulence: an introduction for scientists and engineers*, Oxford University Press, Oxford, UK, 2015. 598, 830, 871
- Davies-Jones, R., Comments on "A Generalization of Bernoulli's Theorem", *Journal of the Atmospheric Sciences*, 60, 2039–2041, doi:10.1175/1520-0469(2003)060<2039:COAGOB>2.0.CO;2, 2003a. 901
- Davies-Jones, R., An expression for effective buoyancy in surroundings with horizontal density gradients, *Journal of the Atmospheric Sciences*, 60, 2922–2925, doi:10.1175/1520-0469(2003)060, 2922:AEFEBI.2.0.CO;2, 2003b. 575
- Davis, R. E., Diapycnal mixing in the ocean: equations for large-scale budgets, *Journal of Physical Oceanography*, 24, 777–800, 1994. 435, 470
- DeGroot, S. R., and P. Mazur, *Non-Equilibrium Thermodynamics*, Dover Publications, New York, 510 pp, 1984. 311, 317, 457, 470, 472
- DeSzeoke, R. A., and A. F. Bennett, Microstructure fluxes across density surfaces, *Journal of Physical Oceanography*, 23, 2254–2264, 1993. 1138
- Doering, C., and J. Gibbon, *Applied analysis of the Navier-Stokes equations*, Cambridge University Press, Cambridge, UK, 217 pp, 2004. 389
- Doswell, C., and P. Markowski, Is buoyancy a relative quantity?, *Monthly Weather Review*, 132, 853–863, doi:10.1175/1520-0493(2004)132<0853:IBARQ>2.0.CO;2, 2004. 575
- Drazin, P., and R. Johnson, *Solitons: an introduction*, Cambridge University Press, Cambridge, UK, 226 pp, 1989. 662
- Drijfhout, S. S., D. P. Marshall, and H. A. Dijkstra, Conceptual models of the wind-driven and thermohaline circulation, in *Ocean Circulation and Climate, 2nd Edition: A 21st Century Perspective, International Geophysics Series*, vol. 103, edited by G. Siedler, S. M. Griffies, J. Gould, and J. Church, pp. 257–282, Academic Press, 2013. 957
- Duchateau, P., and D. Zachmann, *Partial differential equations*, Schaum's Outline Series in Mathematics, McGraw-Hill, New York, 1986. 45
- Dukowicz, J. K., and R. D. Smith, Stochastic theory of compressible turbulent fluid transport, *Physics of Fluids*, 9, 3523–3529, 1997. 1158
- Dyke, M. V., *Perturbation methods in fluid mechanics*, 1st ed., The Parabolic Press, Stanford, California, USA, 271, 1975. 397
- Eckart, C., An analysis of the stirring and mixing processes in incompressible fluids, *Journal of Marine Research*, 7, 265–275, 1948. 1067, 1168
- Einstein, A., Über die von der molekularkinetischen theorie der wärme geforderte bewegung von in ruhenden flüssigkeiten suspendierten teilchen, *Annalen der Physik (in German)*, 322, 549–560, 1905. 1069, 1168

- Einstein, A., *Autobiographical Notes*, Open Court, 1949. 423
- Eliassen, A., and E. Palm, On the transfer of energy in stationary mountain waves, *Geofysiske Publikasjoner*, 22(1–23), 1960. 1051
- Ertel, H., Ein neuer hydrodynamicscher Wirbelsatz, *Meteorol. Z. (Braunschweig)*, 59, 271–281, 1942. 879, 882
- Falkovich, G., *Fluid Mechanics: A short course for physicists*, Cambridge University Press, 167pp, 2011. 403
- Farrell, W., and J. Clark, On postglacial sea level, *Geophysical Journal of the Royal Astronomical Society*, 46, 646–667, 1976. 643
- Feistel, R., Equilibrium thermodynamics of seawater revisited, *Progress in Oceanography*, 31, 101–179, 1993. 439
- Ferrari, R., J. C. McWilliams, V. M. Canuto, and M. Dubovikov, Parameterization of eddy fluxes near oceanic boundaries, *Journal of Climate*, 21, 2770–2789, 2008. 1159, 1168
- Ferrari, R., S. M. Griffies, A. J. G. Nurser, and G. K. Vallis, A boundary-value problem for the parameterized mesoscale eddy transport, *Ocean Modelling*, 32, 143–156, 2010. 1159
- Ferreira, D., and J. Marshall, Formulation and implementation of a residual-mean ocean circulation model, *Ocean Modelling*, 13, 86–107, 2006. 1153
- Fetter, A. L., and J. D. Walecka, *Theoretical Mechanics of Particles and Continua*, McGraw-Hill Book Company, New York, 570 pp, 1980. 402, 403, 662
- Feynman, R. P., R. B. Leighton, and M. L. Sands, *The Feynman lectures on physics*, Addison-Wesley Publishing Company, Reading, Mass, 1963. 472, 486
- Flanders, H., *Differential Forms with Applications to the Physical Sciences*, Dover Publications, New York, 224 pp, 1989. 165, 1224
- Fox-Kemper, B., R. Ferrari, and J. Pedlosky, On the indeterminacy of rotational and divergent eddy fluxes, *Journal of Physical Oceanography*, 33, 478–483, doi:10.1175/1520-0485(2003)033<0478:OTIORA>2.0.CO;2, 2003. 40, 1127
- Frankel, T., *The Geometry of Physics: An Introduction*, Cambridge University Press, 694, 2004. 165
- Frisch, U., *Turbulence, the Legacy of A.N. Kolmogorov*, Cambridge University Press, 1995. 259
- Frisch, U., and B. Villone, Cauchy's almost forgotten Lagrangian formulation of the Euler equation for 3D incompressible flow, *European Physical Journal*, 39, 325–351, doi:10.1140/epjh/e2014-50016-6, 2014. 273
- Garrett, C., Turbulent dispersion in the ocean, *Progress in Oceanography*, 70, 113–125, 2006. 1126, 1130
- Garrett, C., P. MacCready, and P. Rhines, Boundary mixing and arrested ekman layers: Rotating stratified flow near a sloping boundary, *Annual Review of Fluid Mechanics*, 25, 291–323, doi:10.1146/annurev.fl.25.010193.001451, 1993. 640

- Gent, P. R., and J. C. McWilliams, Isopycnal mixing in ocean circulation models, *Journal of Physical Oceanography*, 20, 150–155, 1990. 602, 1157, 1158
- Gent, P. R., J. Willebrand, T. J. McDougall, and J. C. McWilliams, Parameterizing eddy-induced tracer transports in ocean circulation models, *Journal of Physical Oceanography*, 25, 463–474, doi:10.1175/1520-0485(1995)025<463:PEITTI>2.0.CO;2, 1995. 551, 1085, 1138, 1145, 1149, 1150, 1151, 1152, 1153, 1154, 1155, 1160
- Gill, A., *Atmosphere-Ocean Dynamics, International Geophysics Series*, vol. 30, Academic Press, London, 662 + xv pp, 1982. iv, 201, 203, 204, 235, 393, 483, 567, 591, 611, 615, 621, 644, 1168
- Gnanadesikan, A., A global model of silicon cycling: Sensitivity to eddy parameterization and dissolution, *Global Biogeochemical Cycles*, 13, 199–220, 1999. 1157
- Gnanadesikan, A., and R. Weller, Structure and instability of the ekman spiral in the presence of surface gravity waves, *Journal of Physical Oceanography*, 25, 3148–3171, doi:10.1175/1520-0485(1995)025<3148:SAIOTE>2.0.CO;2, 1995. 640
- Göckeler, M., and T. Schücker, *Differential Geometry, Gauge Theories, and Gravity*, 230 pp., Cambridge University Press, Cambridge, UK, 1989. 165
- Godfrey, J., A Sverdrup model of the depth-integrated flow for the World Ocean allowing for island circulations, *Geophysical and Astrophysical Fluid Dynamics*, 45, 89–112, doi:10.1080/03091928908208894, 1989. 819, 820
- Goldstein, H., *Classical Mechanics*, Addison-Wesley Publishing, Reading, MA, 1980. 180
- Graham, F., and T. McDougall, Quantifying the nonconservative production of Conservative Temperature, potential temperature, and entropy, *Journal of Physical Oceanography*, 43, 838–862, 2013. 1144
- Greatbatch, R. J., A note on the representation of steric sea level in models that conserve volume rather than mass, *Journal of Geophysical Research*, 99, 12,767–12,771, 1994. 334, 1180
- Greatbatch, R. J., and K. G. Lamb, On parameterizing vertical mixing of momentum in non-eddy resolving ocean models, *Journal of Physical Oceanography*, 20, 1634–1637, 1990. 1153
- Gregg, M. C., Entropy generation in the ocean by small-scale mixing, *Journal of Physical Oceanography*, 14, 688–711, 1984. 470
- Gregory, J., Vertical heat transports in the ocean and their effect on time-dependent climate change, *Climate Dynamics*, 15, 501–515, 2000. 530
- Gregory, J., S. M. Griffes, C. Hughes, J. Lowe, J. Church, I. Fukimori, N. Gomez, R. Kopp, F. Landerer, R. Ponte, D. Stammer, M. Tamisiea, and R. van den Wal, Concepts and terminology for sea level-mean, variability and change, both local and global, *Surveys in Geophysics*, 40, 1251–1289, doi:10.1007/s10712-019-09555-7, 2019. 203, 644, 1180
- Griffies, S., and R. Hallberg, Biharmonic friction with a Smagorinsky-like viscosity for use in large-scale eddy-permitting ocean models, *Monthly Weather Review*, 128, 2935–2946, doi:10.1175/1520-0493(2000)128<2935:BFWASL>2.0.CO;2, 2000. 466

- Griffies, S., A. Adcroft, and R. Hallberg, A primer on the vertical lagrangian-remap method in ocean models based on finite volume generalized vertical coordinates, *Journal of Advances in Modeling Earth Systems*, 12, doi:10.1029/2019MS001954, 2020. 243, 322, 323, 684, 982, 1020, 1030, 1159
- Griffies, S. M., The Gent-McWilliams skew-flux, *Journal of Physical Oceanography*, 28, 831–841, 1998. 1082, 1150, 1158
- Griffies, S. M., *Fundamentals of Ocean Climate Models*, Princeton University Press, Princeton, USA, 518+xxxiv pages, 2004. 30, 129, 164, 239, 339, 359, 387, 391, 466, 775, 1030, 1075, 1082, 1115, 1138, 1147
- Griffies, S. M., and A. J. Adcroft, Formulating the equations for ocean models, in *Ocean Modeling in an Eddying Regime, Geophysical Monograph*, vol. 177, edited by M. Hecht and H. Hasumi, pp. 281–317, American Geophysical Union, 2008. 520
- Griffies, S. M., and R. J. Greatbatch, Physical processes that impact the evolution of global mean sea level in ocean climate models, *Ocean Modelling*, 51, 37–72, doi:10.1016/j.ocemod.2012.04.003, 2012. 306, 334, 675, 1115, 1146, 1149, 1163, 1180
- Griffies, S. M., A. Gnanadesikan, R. C. Pacanowski, V. Larichev, J. K. Dukowicz, and R. D. Smith, Isoneutral diffusion in a z -coordinate ocean model, *Journal of Physical Oceanography*, 28, 805–830, doi:10.1175/1520-0485(1998)028<0805:IDIAZC>2.0.CO;2, 1998. 1075, 1157, 1158
- Griffies, S. M., C. W. Böning, F. O. Bryan, E. P. Chassignet, R. Gerdes, H. Hasumi, A. Hirst, A.-M. Treguier, and D. Webb, Developments in ocean climate modelling, *Ocean Modelling*, 2, 123–192, 2000. 1032, 1155
- Griffies, S. M., J. Yin, P. J. Durack, P. Goddard, S. Bates, E. Behrens, M. Bentsen, D. Bi, A. Biastoch, C. W. Böning, A. Bozec, C. Cassou, E. Chassignet, G. Danabasoglu, S. Danilov, C. Domingues, H. Drange, R. Farneti, E. Fernandez, R. J. Greatbatch, D. M. Holland, M. Illicak, J. Lu, S. J. Marsland, A. Mishra, W. G. Large, K. Lorbacher, A. G. Nurser, D. Salas y Mélia, J. B. Palter, B. L. Samuels, J. Schröter, F. U. Schwarzkopf, D. Sidorenko, A.-M. Treguier, Y. Tseng, H. Tsujino, P. Uotila, S. Valcke, A. Volodire, Q. Wang, M. Winton, and Z. Zhang, An assessment of global and regional sea level for years 1993–2007 in a suite of interannual CORE-II simulations, *Ocean Modelling*, 78, 35–89, doi:10.1016/j.ocemod.2014.03.004, 2014. 505, 686, 1180
- Groeskamp, S., J. Zika, T. McDougall, B. Sloyan, and F. Laliberté, The representation of ocean circulation and variability in thermodynamic coordinates, *Journal of Physical Oceanography*, 44, 1735–1750, doi:10.1175/JPO-D-13-0213.1, 2014. 1185
- Groeskamp, S., R. P. Abernathey, and A. Klocker, Water mass transformation by cabbeling and thermobaricity, *Geophysical Research Letters*, doi:10.1002/2016GL070860, 2016. 1163
- Groeskamp, S., S. M. Griffies, D. Iudicone, R. Marsh, A. G. Nurser, and J. D. Zika, The water mass transformation framework for ocean physics and biogeochemistry, *Annual Review of Marine Science*, 11, 1–35, doi:10.1146/annurev-marine-010318-095421, 2019. 1011, 1163, 1176, 1185, 1208, 1213, 1214
- Gula, J., M. Molemaker, and J. McWilliams, Gulf Stream dynamics along the southeastern U.S. seaboard, *Journal of Physical Oceanography*, 45, 690–715, doi:10.1175/JPO-D-14-0154.1, 2015. 409, 862, 869

- Haine, T. W. N., E. Delhez, S. Griffies, D. Waugh, M. Holzer, F. Primeau, S. Khatiwala, E. Deleersnijder, and G. Gebbie, Tracers in ocean modeling, *Journal of Advances in Modeling Earth Systems (JAMES)*, *in prep*, 2021. 1067
- Hallberg, R., A thermobaric instability in Lagrangian vertical coordinate ocean models, *Ocean Modelling*, 8, 227–300, 2005. 1039
- Hallberg, R., and P. Rhines, Buoyancy-driven circulation in a ocean basin with isopycnals intersecting the sloping boundary, *Journal of Physical Oceanography*, 26, 913–940, 1996. 857, 912
- Haller, G., Lagrangian coherent structures, *Annual Reviews of Fluid Mechanics*, 47, 137–162, doi: 10.1146/annurev-fluid-010313-141322, 2015. 269
- Haltiner, G. T., and R. T. Williams, *Numerical Prediction and Dynamic Meteorology*, John Wiley and Sons, New York, USA, 1980. 769
- Haney, R. L., On the pressure gradient force over steep topography in sigma-coordinate ocean models, *Journal of Physical Oceanography*, 21, 610–619, 1991. 1032
- Hasselmann, K., Stochastic climate models. Part I: Theory, *Tellus*, 28, 473–485, 1976. 102
- Hasselmann, K., R. Sausen, E. Maier-Reimer, and R. Voss, On the cold start problem in transient simulations with coupled atmosphere-ocean models, *Climate Dynamics*, 9, 53–61, doi:10.1007/BF00210008, 1993. 107
- Haynes, P. H., and M. E. McIntyre, On the evolution of vorticity and potential vorticity in the presence of diabatic heating and frictional or other forces, *Journal of Atmospheric Sciences*, 44, 828–841, 1987. 898, 901
- Haynes, P. H., and M. E. McIntyre, On the conservation and impermeability theorems for potential vorticity, *Journal of Atmospheric Sciences*, 47, 2021–2031, 1990. 901
- Hildebrand, F., *Advanced Calculus for Applications*, Prentice-Hall Publishers, Englewood Cliffs, New Jersey, 1976. 45, 48, 57
- Holloway, G., Eddy transport of thickness and momentum in layer and level models, *Journal of Physical Oceanography*, 27, 1153–1157, 1997. 1155
- Holloway, G., and P. Rhines, Angular momenta of modeled ocean gyres, *Journal of Geophysical Research*, 27, 843–846, doi:10.1029/90JC02256, 1991. 783
- Holmes, R. M., J. Zika, and M. H. England, Diathermal heat transport in a global ocean model, *Journal of Physical Oceanography*, 49, 141–161, doi:10.1175/JPO-D-18-0098.1, 2019. 1195, 1218
- Holton, J. R., *An Introduction to Dynamic Meteorology*, Academic Press, San Diego, USA, 507 pp, 1992. 363, 364, 415, 421, 454, 605, 611, 615, 617, 619, 620, 776, 803, 848, 865
- Hoskins, B., Towards and $\text{pv}-\theta$ view of the general circulation, *Tellus*, 43AB, 27–35, 1991. 884
- Hoskins, B., I. Draghici, and H. Davies, A new look at the ω equation, *Quarterly Journal of the Royal Meteorological Society*, 104, 31–38, 1978. 977, 978
- Hoskins, B., I. James, and G. White, The shape, propagation, and mean flow interaction of large-scale weather systems, *Journal of the Atmospheric Sciences*, 40, 1595–1612, doi: 10.1175/1520-0469(1983)040<1595:TSPAMF>2.0.CO;2, 1983. 765

- Hoskins, B. J., and I. N. James, *Fluid Dynamics of the Midlatitude Atmosphere*, Wiley Blackwell, Chichester, UK, 2014. 284, 286
- Huang, K., *Statistical Mechanics*, John Wiley and Sons, New York, 493 pp, 1987. 231, 240, 1072
- Huang, R. X., and R. W. Schmitt, Goldsbrough-Stommel circulation of the World Oceans, *Journal of Physical Oceanography*, 23, 1277–1284, 1993. 951
- Hughes, C. W., and B. de Cueves, Why western boundary currents in realistic oceans are inviscid: A link between form stress and bottom pressure torques, *Journal of Physical Oceanography*, 31, 2871–2885, 2001. 421, 724, 816, 857
- IOC, SCOR, and IAPSO, *The international thermodynamic equation of seawater-2010: calculation and use of thermodynamic properties*, Intergovernmental Oceanographic Commission, Manuals and Guides No. 56, UNESCO, 196pp, 2010. 439, 472, 480, 557, 560, 1144, 1166, 1169
- Jackett, D. R., and T. J. McDougall, A neutral density variable for the world's oceans, *Journal of Physical Oceanography*, 27, 237–263, 1997. 1186
- Jackson, J. D., *Classical Electrodynamics*, John Wiley and Sons, New York, USA, 848 pp, 1975. 68, 581, 1083
- Jackson, L., C. Hughes, and R. Williams, Topographic control of basin and channel flows: the role of bottom pressure torques and friction, *Journal of Physical Oceanography*, 36, 1786–1805, doi:10.1175/JPO2936.1, 2006. 816
- Jeevanjee, N., and D. Romps, Effective buoyancy, inertial pressure, and the mechanical generation of boundary layer mass flux by cold pools, *Journal of Atmospheric Sciences*, 72, 3199–3213, doi:10.1175/JAS-D-14-0349.1, 2015a. 575, 775
- Jeevanjee, N., and D. Romps, Effective buoyancy at the surface and aloft, *Quarterly Journal of the Royal Meteorological Society*, 142, 811–820, doi:10.1002/qj.2683, 2015b. 575
- Killworth, P. D., On the parameterization of eddy transfer Part I: Theory, *Journal of Marine Research*, 55, 1171–1197, 1997. 1155
- Klinger, B., and T. Haine, *Ocean Circulation in Three Dimensions*, Cambridge University Press, 466, 2019. 820
- Klocker, A., and R. Abernathey, Global patterns of mesoscale eddy properties and diffusivities, *Journal of Physical Oceanography*, 44, 1030–1046, 2014. 1168
- Klocker, A., and T. J. McDougall, Influence of the nonlinear equation of state on global estimates of dianeutral advection and diffusion, *Journal of Physical Oceanography*, 40, 1690–1709, 2010. 1163
- Kopp, R. E., J. X. Mitrovica, S. M. Griffies, J. Yin, C. C. Hay, and R. J. Stouffer, The impact of Greenland melt on regional sea level: a preliminary comparison of dynamic and static equilibrium effects, *Climatic Change Letters*, 103, 619–625, doi:10.1007/s10584-010-9935-1, 2010. 643
- Kraichnan, R. H., and D. Montgomery, Two-dimensional turbulence, *Reports on Progress in Physics*, 43, 547–619, 1980. 769

Kundu, P., I. Cohen, and D. Dowling, *Fluid Mechanics*, Academic Press, 921 + xxiv pp, 2016. iv, 240, 256, 267, 385, 386, 391, 403, 1072

Kushner, P. J., and I. M. Held, Potential vorticity thickness fluxes and wave-mean flow interaction, *Journal of Atmospheric Sciences*, 56, 948–958, 1999. 1138

Landau, L. D., and E. M. Lifshitz, *Mechanics*, Pergamon Press, Oxford, UK, 170 pp, 1976. 209

Landau, L. D., and E. M. Lifshitz, *Fluid Mechanics*, Pergamon Press, Oxford, UK, 539 pp, 1987. iv, 403, 457, 459, 470, 472, 473

Large, W., J. McWilliams, and S. Doney, Oceanic vertical mixing: a review and a model with a nonlocal boundary layer parameterization, *Reviews of Geophysics*, 32, 363–403, 1994. 1141, 1168, 1176

LeCorre, M., J. Gula, and A.-M. Treguier, Barotropic vorticity balance of the North Atlantic subpolar gyre in an eddy-resolving model, *Ocean Science*, 16, 451–468, doi:10.5194/os-16-451-2020, 2020. 862

Lee, M.-M., D. Marshall, and R. Williams, On the eddy transfer of tracers: advective or diffusive?, *Journal of Marine Research*, 55, 483–505, 1997. 1051

Lemarié, F., L. Debreu, A. F. Shchepetkin, and J. C. McWilliams, On the stability and accuracy of the harmonic and biharmonic isoneutral mixing operators in ocean models, *Ocean Modelling*, 52-53, 9–35, 2012. 1157

Lilly, J., Kinematics of a fluid ellipse in a linear flow, *Fluids*, 3(16), doi:10.3390/fluids3010016, 2018. 286

Lin, S. J., A finite volume integration method for computing pressure gradient force in general vertical coordinates, *Quarterly Journal of the Royal Meteorological Society*, 123, 1749–1762, 1997. 1031, 1032, 1034

Lorenz, E. N., Available potential energy and the maintenance of the general circulation, *Tellus*, 7, 157–167, 1955. 538, 539

MacCready, P., and P. Rhines, Buoyant inhibition of ekman transport on a slope and its effect on stratified spin-up, *Journal of Fluid Mechanics*, 223, 631–661, doi:10.1017/S0022112091001581, 1991. 640

MacCready, P., and P. Rhines, Slippery bottom boundary layers on a slope, *Journal of Physical Oceanography*, 23, 5–22, doi:10.1175/1520-0485(1993)023<0005:SBBLOA>2.0.CO;2, 1993. 640

MacKinnon, J., Louis St. Laurent, and A. C. Naveira Garabato, Diapycnal mixing processes in the ocean interior, in *Ocean Circulation and Climate, 2nd Edition: A 21st century perspective, International Geophysics Series*, vol. 103, edited by G. Siedler, S. M. Griffies, J. Gould, and J. Church, pp. 159–183, Academic Press, 2013. 466, 1145, 1149, 1168

Maddison, J., D. Marshall, and J. Shipton, On the dynamical influence of ocean eddy potential vorticity fluxes, *Ocean Modelling*, 92, 169–182, doi:10.1016/j.ocemod.2015.06.003, 2015. 40

Maddison, J., and D. Marshall, The Eliassen-Palm flux tensor, *Journal of Fluid Mechanics*, 729, 69–102, doi:10.1017/jfm.2013.259, 2013. 705, 714, 972, 1049, 1061

- Marion, J. B., and S. T. Thornton, *Classical Dynamics of Particles and Systems*, Harcourt Brace Jovanovich, San Diego, USA, 602 pp, 1988. 180, 792
- Markowski, P., and Y. Richardson, *Mesoscale Meteorology in Midlatitudes*, Wiley-Blackwell Publishers, Oxford, UK, 2010. 211, 527, 575, 865
- Marshall, D., Vertical fluxes of potential vorticity and the structure of the thermocline, *Journal of Physical Oceanography*, 30, 3102–3112, 2000. 891, 904
- Marshall, J., and R. A. Plumb, *Atmosphere, Ocean, and Climate Dynamics*, 1st ed., Academic Press, Amsterdam, 319pp, 2008. iv, 363, 508, 591, 621, 626, 726, 736, 737
- Marshall, J., D. Jamous, and J. Nilsson, Entry, flux, and exit of potential vorticity in ocean circulation, *Journal of Physical Oceanography*, 31, 777–789, doi:10.1175/1520-0485(2001)031<0777:EFAEOP>2.0.CO;2, 2001. 891, 901, 903, 904, 912
- Marshall, J., K. Armour, J. Scott, Y. Kostov, U. Hausmann, D. Ferreira, T. Shepherd, and C. Bitz, The ocean's role in polar climate change: asymmetric Arctic and Antarctic responses to greenhouse gas and ozone forcing, *Philosophical Transactions of the Royal Society A*, A372, doi: 10.1098/rsta.2013.0040, 2014. 107
- Marshall, J. C., E. Shuckburgh, H. Jones, and C. Hill, Estimates and implications of surface eddy diffusivity in the Southern Ocean derived from tracer transport, *Journal of Physical Oceanography*, 36, 1806–1821, 2006. 1110
- McCabe, R., P. MacCready, and G. Pawlak, Form drag due to flow separation at a headland, *Journal of Physical Oceanography*, 36, 2136–2152, doi:10.1175/JPO2966.1, 2006. 411
- McDougall, T., R. Feistel, and R. Pawlowicz, Thermodynamics of seawater, in *Ocean Circulation and Climate, 2nd Edition: A 21st Century Perspective, International Geophysics Series*, vol. 103, edited by G. Siedler, S. M. Griffies, J. Gould, and J. Church, pp. 141–158, Academic Press, 2013. 560
- McDougall, T. J., Neutral surfaces, *Journal of Physical Oceanography*, 17, 1950–1967, 1987a. 567, 1168
- McDougall, T. J., Thermobaricity, cabbeling, and water-mass conversion, *Journal of Geophysical Research*, 92, 5448–5464, 1987b. 567, 1147, 1148, 1161, 1162, 1168
- McDougall, T. J., Potential enthalpy: a conservative oceanic variable for evaluating heat content and heat fluxes, *Journal of Physical Oceanography*, 33, 945–963, doi:10.1175/1520-0485(2003)033<0945:PEACOV>2.0.CO;2, 2003. 482, 483, 1166
- McDougall, T. J., and R. Feistel, What causes the adiabatic lapse rate, *Deep-Sea Research*, 50, 1523–1535, 2003. 444
- McDougall, T. J., and D. R. Jackett, On the helical nature of neutral trajectories in the ocean, *Progress in Oceanography*, 20, 153–183, 1988. 569
- McDougall, T. J., and D. R. Jackett, The thickness of the ocean in $s - \theta - p$ space and the implications for mean diapycnal advection, *Journal of Physical Oceanography*, 37, 1714–1732, 2007. 569

McDougall, T. J., and P. C. McIntosh, The temporal-residual-mean velocity. Part II: isopycnal interpretation and the tracer and momentum equations, *Journal of Physical Oceanography*, 31, 1222–1246, 2001. 1092, 1115, 1132, 1138, 1141, 1150

McDougall, T. J., and Y. You, Implications of the nonlinear equation of state for upwelling in the ocean interior, *Journal of Geophysical Research*, 95, 13,263—13,276, 1990. 1163

McDougall, T. J., S. Groeskamp, and S. M. Griffies, On geometric aspects of interior ocean mixing, *Journal of Physical Oceanography*, 44, 2164–2175, 2014. 567, 1145, 1147, 1168

McIntosh, P. C., and T. J. McDougall, Isopycnal averaging and the residual mean circulation., *Journal of Physical Oceanography*, 26, 1655–1660, 1996. 1138, 1150

McWilliams, J., *Fundamentals of Geophysical Fluid Dynamics*, Cambridge University Press, Cambridge, Cambridge, UK, 2006. iv, 761, 781, 1050

Mellor, G. L., L.-Y. Oey, and T. Ezer, Sigma coordinate pressure gradient errors and the seamount problem, *Journal of Atmospheric and Oceanic Technology*, 15, 1122–1131, 1998. 1032

Mertz, G., and D. Wright, Interpretations of the JE BAR term, *Journal of Physical Oceanography*, 22, 301–305, doi:10.1175/1520-0485(1992)022<301:JOTJT>2.0.CO;2, 1992. 957

Meyer, R., *Introduction to mathematical fluid dynamics*, Dover Publications, New York, 1971. iv

Middleton, J. F., and J. W. Loder, Skew fluxes in polarized wave fields, *Journal of Physical Oceanography*, 19, 68–76, 1989. 1082, 1126, 1127, 1128, 1130, 1160

Misner, C., K. Thorne, and J. Wheeler, *Gravitation*, W.H. Freeman and Co., 1279 pages, 1973. 165, 1218

Mitrovica, J. X., M. E. Tamisiea, J. L. Davis, and G. A. Milne, Recent mass balance of polar ice sheets inferred from patterns of global sea-level change, *Nature*, 409, 1026–1029, 2001. 643, 644

Moffatt, H., Transport effects associated with turbulence with particular attention to the influence of helicity, *Reports on Progress in Physics*, 46, 621–664, 1983. 1082, 1130

Molemaker, M., J. McWilliams, and W. Dewar, Submesoscale instability and generation of mesoscale anticyclones near a separation of the California Undercurrent, *Journal of Physical Oceanography*, 45, 613–629, doi:10.1175/JPO-D-13-0225.1, 2015. 409, 869

Morse, P. M., and H. Feshbach, *Methods of Theoretical Physics Part I and II*, McGraw-Hill Book Company, New York, 1953. 39, 67, 87, 204

Mory, M., Inertial oscillations, in *Rotating Fluids in Geophysical and Industrial Applications*, International Centre for Mechanical Sciences (Courses and Lectures), Springer, Vienna, doi: 10.1007/978-3-7091-2602-8_8, 1992. 598

Müller, P., Ertel's potential vorticity theorem in physical oceanography, *Reviews of Geophysics*, 33, 67–97, 1995. 883

Müller, P., *The Equations of Oceanic Motions*, 1st ed., Cambridge University Press, Cambridge, 302pp, 2006. 472, 836

- Müller, P., and C. Garrett, From stirring to mixing in a stratified ocean, *Oceanography*, 15, 12–19, 2002. [1168](#)
- Munk, W., and E. Palmén, Note on the dynamics of the Antarctic Circumpolar Current, *Tellus*, 3, 53–55, 1951. [421](#), [720](#), [723](#)
- Munk, W., and C. Wunsch, Abyssal recipes II: Energetics of tidal and wind mixing, *Deep Sea Research*, 45, 1977–2010, doi:10.1016/S0967-0637(98)00070-3, 1998. [1168](#)
- Munk, W. H., On the wind-driven ocean circulation, *Journal of Meteorology*, 7, 79–93, 1950. [421](#)
- Munk, W. H., Abyssal recipes, *Deep-Sea Research*, 13, 707–730, 1966. [1168](#)
- Nakahara, M., *Geometry, Topology and Physics*, 2nd ed., Institute of Physics, Bristol and Philadelphia, 2003. [165](#)
- Nakamura, N., A new look at eddy diffusivity as a mixing diagnostic, *Journal of the Atmospheric Sciences*, 58(24), 3685–3701, doi:10.1175/1520-0469(2001)058<3685:ANLAED>2.0.CO;2, 2001. [1168](#)
- Naveira Garabato, A., E. Frajka-Williams, C. Spingys, A. Legg, K. Polzin, A. Forryan, E. Abrahamsen, C. Buckingham, S. Griffies, S. McPhail, K. Nicholls, L. Thomas, and M. Meredith, Rapid mixing and exchange of deep-ocean waters in an abyssal boundary current, *Proceedings of the National Academy of Sciences*, doi:10.1073/pnas.1904087116, 2019. [912](#)
- Noether, E., Invariante variationsprobleme, *Nachrichten von der Gesellschaft der Wissenschaften zu Göttingen, Math-phys. Klasse*, pp. 235–257, 1918. [209](#)
- Noether, E., and M. Tavel, Invariant variational problems, *arXiv:physics/0503066 [physics.hist-ph]*, 2018. [209](#)
- Nurser, A. G., and S. M. Griffies, Relating diffusive surface salinity fluxes to boundary freshwater and salt fluxes, *Journal of Physical Oceanography*, 49, 2365–2376, doi:10.1175/JPO-D-19-0037.1, 2019. [1173](#)
- Nurser, A. G., and M.-M. Lee, Isopycnal averaging at constant height. Part I: The formulation and a case study, *Journal of Physical Oceanography*, 34, 2721–2739, 2004a. [1138](#)
- Nurser, A. G., and M.-M. Lee, Isopycnal averaging at constant height. Part II: Relating to the residual streamfunction in Eulerian space, *Journal of Physical Oceanography*, 34, 2740–2755, 2004b. [1138](#)
- Nurser, A. G., R. Marsh, and R. Williams, Diagnosing water mass formation from air-sea fluxes and surface mixing, *Journal of Physical Oceanography*, 29, 1468–1487, doi:10.1175/1520-0485(1999)029<1468:DWMFFA>2.0.CO;2, 1999. [539](#)
- Olbers, D. J., J. Willebrand, and C. Eden, *Ocean Dynamics*, 1st ed., Springer, Berlin, Germany, 704 pages, 2012. [iv](#), [240](#), [241](#), [245](#), [278](#), [435](#), [472](#), [480](#), [483](#), [567](#), [1144](#)
- Ottino, J. M., *The kinematics of mixing: stretching, chaos, and transport*, Cambridge University Press, Cambridge, UK, 396 pp, 1989. [269](#)

Otto, A., F. Otto, O. Boucher, J. Church, G. Hegerl, P. Forster, N. Gillett, J. Gregory, G. Johnson, R. Knutti, N. Lewis, U. Lohmann, J. Marotzke, G. Myhre, D. Shindell, B. Stevens, and M. Allen, Energy budget constraints on climate response, *Nature Geosciences*, 6, 415–416, doi:10.1038/ngeo1836, 2013. 1179

Patmore, R., P. Holland, D. Munday, A. Naveira Garabato, D. Stevens, and M. Meredith, Topographic control of Southern Ocean gyres and the Antarctic Circumpolar Current: a barotropic perspective, *Journal of Physical Oceanography*, 49, 3221–3244, doi:10.1175/JPO-D-19-0083.1, 2019. 816

Pedlosky, J., *Geophysical Fluid Dynamics*, 2nd ed., Springer-Verlag, Berlin Heidelberg New York, 710 + xv pp, 1987. iv

Pedlosky, J., L. Pratt, M. Spall, and K. Helfrich, Circulation around islands and ridges, *Journal of Marine Research*, 55, 1199–1251, 1997. 820

Plumb, R. A., Eddy fluxes of conserved quantities by small-amplitude waves, *Journal of Atmospheric Sciences*, 36, 1699–1704, 1979. 1130

Polton, J., and D. Marshall, Overturning cells in the Southern Ocean and subtropical gyres, *Ocean Science*, 3, 17–30, 2007. 891, 904

Polzin, K. L., J. M. Toole, J. R. Ledwell, and R. W. Schmitt, Spatial variability of turbulent mixing in the abyssal ocean, *Science*, 276, 93–96, 1997. 1168

Pratt, L. J., and J. A. Whitehead, *Rotating Hydraulics: Nonlinear Topographic Effects in the Ocean and Atmosphere*, 589 pp., Springer Science + Business Media, New York, 2008. 255, 701, 809

Pugh, D. T., *Tides, surges, and mean sea-level*, 472 pp., John Wiley and Sons, 1987. 643

Redi, M. H., Oceanic isopycnal mixing by coordinate rotation, *Journal of Physical Oceanography*, 12, 1154–1158, 1982. 1158

Reif, F., *Fundamentals of Statistical and Thermal Physics*, McGraw-Hill, New York, 1965. 231, 240, 423, 426, 427, 1072

Rhines, P. B., Basic dynamics of the large-scale geostrophic circulation, in *WHOI 1982 Summer Study Program*, Woods Hole Oceanographic Institute, 1982. 1136

Rhines, P. B., and W. R. Holland, A theoretical discussion of eddy-driven mean flows, *Dynamics Of Atmospheres And Oceans*, 3, 289–325, doi:10.1016/0377-0265(79)90015-0, 1979. 1113

Rhines, P. B., and W. R. Young, Homogenization of potential vorticity in planetary gyres, *Journal of Fluid Mechanics*, 122, 347–367, 1982. 1218

Ringler, T., J. Saenz, P. Wolfram, and L. Van Roekel, A thickness-weighted average perspective of force balance in an idealized Circumpolar Current, *Journal of Physical Oceanography*, 47, 285–302, doi:10.1175/JPO-D-16-0096.1, 2017. 539

Rintoul, S. R., The global influence of localized dynamics in the Southern Ocean, *Nature*, 558, 209–218, doi:10.1038/s41586-018-0182-3, 2018. 422

- Rintoul, S. R., and A. C. Naveira Garabato, Dynamics of the Southern Ocean circulation, in *Ocean Circulation and Climate, 2nd Edition: A 21st Century Perspective, International Geophysics Series*, vol. 103, edited by G. Siedler, S. M. Griffies, J. Gould, and J. Church, pp. 471–492, Academic Press, 2013. [422](#)
- Rintoul, S. R., C. W. Hughes, and D. Olbers, The Antarctic Circumpolar Current system, in *Ocean Circulation and Climate, 1st Edition, International Geophysics Series*, vol. 103, edited by G. Siedler, J. Gould, and J. Church, pp. 271–301, Academic Press, 2001. [422](#)
- Roberts, M. J., and D. Marshall, Do we require adiabatic dissipation schemes in eddy-resolving ocean models?, *Journal of Physical Oceanography*, 28, 2050–2063, 1998. [1143](#)
- Rossby, C.-G., Planetary flow patterns in the atmosphere, *Quarterly Journal of the Royal Meteorological Society (suppl)*, 66, 68–87, 1940. [744](#), [799](#)
- Sadourny, R., and C. Basdevant, Parameterization of subgrid scale barotropic and baroclinic eddies in quasi-geostrophic models: anticipated potential vorticity method, *Journal of Atmospheric Sciences*, 42, 1353–1363, 1985. [784](#)
- Salmon, R., *Lectures on Geophysical Fluid Dynamics*, Oxford University Press, Oxford, England, 378 + xiii pp., 1998. [iv](#), [231](#), [240](#), [241](#), [385](#), [673](#), [794](#)
- Samelson, R., *The Theory of Large-Scale Ocean Circulation*, Cambridge University Press, Cambridge, UK, 193 pp., 2011. [591](#), [950](#), [957](#)
- Schär, C., A generalization of Bernoulli's Theorem, *Journal of the Atmospheric Sciences*, 50, 1437–1443, doi:10.1029/2003JC001823, 1993. [891](#), [903](#), [904](#)
- Schey, H., *Div, grad, curl and all that: an informal text on vector calculus*, W.W. Norton and Company, Inc., 176 pp, 2004. [21](#)
- Schmitt, R. W., Double diffusion in oceanography, *Annual Review of Fluid Mechanics*, 26, 255–285, 1994. [1148](#), [1168](#)
- Schneider, T., I. M. Held, and S. Garner, Boundary effects in potential vorticity dynamics, *Journal of Atmospheric Sciences*, 60, 1024–1040, doi:10.1175/1520-0469(2003)60<1024:BEIPVD>2.0.CO;2, 2003. [539](#), [972](#)
- Schutz, B. F., *Geometrical Methods of Mathematical Physics*, Cambridge University Press, Cambridge, UK, 250 pp, 1980. [126](#), [165](#)
- Schutz, B. F., *A First Course in General Relativity*, Cambridge University Press, Cambridge, UK, 392 pp, 1985. [121](#)
- Shchepetkin, A., and J. McWilliams, A method for computing horizontal pressure-gradient force in an ocean model with a non-aligned vertical coordinate, *Journal of Geophysical Research*, 108, 35.1–35.34, 2002. [1032](#)
- Smagorinsky, J., General circulation experiments with the primitive equations: I. The basic experiment, *Monthly Weather Review*, 91, 99–164, 1963. [494](#)
- Smagorinsky, J., Some historical remarks on the use of nonlinear viscosities, in *Large Eddy Simulation of Complex Engineering and Geophysical Flows*, edited by B. Galperin and S. A. Orszag, pp. 3–36, Cambridge University Press, 1993. [391](#)

Smith, K. S., and J. Marshall, Evidence for enhanced eddy mixing at middepth in the southern ocean, *Journal of Physical Oceanography*, 39, 50–69, 2009. 1155, 1159

Smith, K. S., and G. K. Vallis, The scales and equilibration of midocean eddies: freely evolving flow, *Journal of Physical Oceanography*, 31, 554–570, 2001. 591, 769

Smith, K. S., and G. K. Vallis, The scales and equilibration of midocean eddies: forced-dissipative flow, *Journal of Physical Oceanography*, 32, 1699–1721, 2002. 1155

Smith, R. D., The primitive equations in the stochastic theory of adiabatic stratified turbulence, *Journal of Physical Oceanography*, 29, 1865–1880, 1999. 1141

Spence, P., R. M. Holmes, A. McC. Hogg, S. M. Griffies, K. D. Stewart, and M. H. England, Localized rapid warming of West Antarctic subsurface waters by remote winds, *Nature Climate Change*, doi:10.1038/NCLIMATE3335, 2017. 641

Spiegel, M., *Theory and Problems of Vector Analysis*, Schaum's Outline Series, McGraw-Hill International Book Company, New York, 1974. 21

Stakgold, I., *Boundary value problems of mathematical physics, volume I*, SIAM, Philadelphia, 340 pp, 2000a. 45, 61, 67, 72, 73, 84, 107

Stakgold, I., *Boundary value problems of mathematical physics, volume II*, SIAM, Philadelphia, 408 pp, 2000b. 45, 61, 67, 70, 84

Stanley, G. J., Neutral surface topology, *Ocean Modelling*, doi:10.1016/j.ocemod.2019.01.008, 2019. 569

Starr, V. P., A quasi-Lagrangian system of hydrodynamical equations, *Journal of Meteorology*, 2, 227–237, doi:10.1175/1520-0469(1945)002<0227:AQLSOH>2.0.CO;2, 1945. 982

Stewart, R., *An Introduction to Physical Oceanography*, 345 pp., Texas A& M, College Station, Texas, USA, 2008. 643

Stommel, H., The westward intensification of wind-driven ocean currents, *Transactions of the American Geophysical Union*, 29, 202–206, 1948. 421

Straub, D. N., On the transport and angular momentum balance of channel models of the Antarctic Circumpolar Current, *Journal of Physical Oceanography*, 23, 776–782, doi:10.1175/1520-0485(1993)023<0776:OTTAAM>2.0.CO;2, 1993. 422

Straub, D. N., On thermobaric production of potential vorticity in the ocean, *Tellus*, 51A, 314–325, 1999. 886

Sun, S., R. Bleck, C. Rooth, J. Dukowicz, E. Chassignet, and P. D. Killworth, Inclusion of thermobaricity in isopycnic-coordinate ocean models, *Journal of Physical Oceanography*, 29, 2719–2729, 1999. 1039

Symon, K., *Mechanics*, Addison-Wesley Publishing Co., Reading, MA, USA, 639 pp, 1971. v

Talley, L. D., G. L. Pickard, W. J. Emery, and J. H. Swift, *Descriptive Physical Oceanography*, 6th ed., Elsevier, 555pp, 2011. 685

- Tarshish, N., N. Jeevangee, and D. Lecoanet, Buoyant motion of a turbulent thermal, *Journal of the Atmospheric Sciences*, 75, 3233–3244, doi:10.1175/JAS-D-17-0371.1, 2018. 554, 575
- Taylor, G., Diffusion by continuous movements, *Proceedings of the London Mathematical Society*, 20, 196–212, 1921. 1070, 1168
- Tennekes, H., and J. Lumley, *A First Course in Turbulence*, 300 pp., MIT Press, Cambridge, USA, 1972. 397, 748
- Thomas, L., A. Tandon, and A. Mahadevan, Submesoscale processes and dynamics, in *Eddy resolving ocean models*, edited by M. Hecht and H. Hasumi, Geophysical Monograph 177, pp. 17–38, American Geophysical Union, 2008. 901, 912
- Thomas, L., J. R. Taylor, R. Ferrari, and T. Joyce, Symmetric instability in the Gulf Stream, *Deep Sea Research II*, 91, 96–110, doi:10.1016/j.dsr2.2013.02.025, 2013. 912
- Thorne, K., and R. Blandford, *Modern Classical Physics*, Princeton University Press, Princeton, USA, 1511 + xl pp, 2017. 2, 129, 366, 472, 473, 839
- Tomczak, M., and J. S. Godfrey, *Regional Oceanography: An Introduction*, Pergamon Press, Oxford, England, 422 + vii pp, 1994. 505, 686, 820
- Treguier, A. M., I. M. Held, and V. D. Larichev, On the parameterization of quasi-geostrophic eddies in primitive equation ocean models, *Journal of Physical Oceanography*, 27, 567–580, 1997. 1155, 1159, 1168
- Truesdell, C., Notes on the history of the general equations of hydrodynamics, *American Mathematical Monthly*, 60, 445–458, doi:10.2307/2308407, 1953. 243
- Truesdell, C., *The kinematics of vorticity*, Dover Publications, Mineola, New York, 1954. 321
- Vallis, G., and B.-L. Hua, Eddy viscosity of the anticipated potential vorticity method, *Journal of the Atmospheric Sciences*, 45, 617–627, doi:10.1175/1520-0469(1988)045<0617:EVOTAP>2.0.CO;2, 1988. 785
- Vallis, G. K., *Atmospheric and Oceanic Fluid Dynamics: Fundamentals and Large-scale Circulation*, 1st ed., Cambridge University Press, Cambridge, 745 + xxv pp, 2006. 692, 736
- Vallis, G. K., *Atmospheric and Oceanic Fluid Dynamics: Fundamentals and Large-scale Circulation*, 2nd ed., Cambridge University Press, Cambridge, 946 + xxv pp, 2017. iv, xvi, 158, 204, 206, 240, 357, 366, 393, 422, 444, 472, 483, 489, 505, 511, 519, 520, 538, 546, 549, 557, 560, 567, 577, 591, 602, 605, 611, 621, 673, 684, 686, 733, 740, 777, 778, 781, 785, 808, 816, 845, 886, 915, 929, 930, 940, 976, 1072, 1082, 1106, 1144, 1145
- Vallis, G. K., *Essentials of Atmospheric and Oceanic Dynamics*, 1st ed., Cambridge University Press, Cambridge, 2019. iv, xvi, 824
- van Heijst, G., Dynamics of vortices in rotating and stratified flows, in *Fronts, Waves, and Vortices in Geophysical Flows*, Lecture notes in Physics 805, p. 192, Springer, 2010. 610, 620
- Van Roekel, L., A. Adcroft, G. Danabasoglu, S. M. Griffies, B. Kauffman, W. Large, M. Levy, B. Reichl, T. Ringler, and M. Schmidt, The KPP boundary layer scheme for the ocean: revisiting its formulation and benchmarking one-dimensional simulations relative to LES, *Journal of Advances in Modeling Earth Systems*, doi:10.1029/2018ms001336, 2018. 1176

Venaille, A., G. K. Vallis, and S. M. Griffies, The catalytic role of beta effect in barotropization processes, *Journal of Fluid Dynamics*, doi:10.1017/jfm.2012.344, 2012. 769

Veronis, G., Large scale ocean circulation, *Advances in Applied Mechanics*, 13, 2–92, 1973. 204

Veronis, G., The role of models in tracer studies, in *Numerical Models of Ocean Circulation*, pp. 133–146, National Academy of Sciences, 1975. 1143

von Arx, W. S., *An Introduction to Physical Oceanography*, Addison-Wesley Publishing Company, Reading, Massachusetts, USA, 422 pp, 1962. 616, 631, 635

Wåhlin, A., R. Muench, L. Arneborg, G. Björk, H. Ha, S. Lee, and A. Alsén, Some implications of Ekman layer dynamics for cross-shelf exchange in the amundsen sea, *Journal of Physical Oceanography*, 42, 1461–1474, doi:10.1175/JPO-D-11-041.1, 2012. 641, 642

Walin, G., A theoretical framework for the description of estuaries, *Tellus*, 29(2), 128–136, doi: 10.1111/j.2153-3490.1977.tb00716.x, 1977. 1195

Walin, G., On the relation between sea-surface heat flow and thermal circulation in the ocean, *Tellus*, 34, 187–195, doi:10.1111/j.2153-3490.1982.tb01806.x, 1982. 1195

Wallace, J., and P. Hobbs, *Atmospheric Science: An Introductory Survey*, Academic Press, 2006. 736

Ward, M., and A. Hogg, Establishment of momentum balance by form stress in a wind-driven channel, *Ocean Modelling*, 40(133–146), doi:10.1016/j.ocemod.2011.08.004, 2011. 714

Waterhouse, A. F., J. A. MacKinnon, J. D. Nash, M. H. Alford, E. Kunze, H. L. Simmons, K. L. Polzin, L. C. St. Laurent, O. M. Sun, R. Pinkel, L. D. Talley, C. B. Whalen, T. N. Huussen, G. S. Carter, I. Fer, S. Waterman, A. C. Naveira Garabato, T. B. Sanford, and C. M. Lee, Global Patterns of Diapycnal Mixing from Measurements of the Turbulent Dissipation Rate., *Journal of Physical Oceanography*, 44(7), 1854–1872, doi:10.1175/JPO-D-13-0104.1, 2014. 1168

Waterman, S., and B. J. Hoskins, Eddy shape, orientation, propagation, and mean flow feedback in western boundary current jets, *Journal of Physical Oceanography*, 43, 1666–1690, 2013. 765

Waterman, S., and J. Lilly, Geometric decomposition of eddy feedbacks in barotropic systems, *Journal of Physical Oceanography*, 45, 1009–1024, 2015. 765

Webb, D., R. Holmes, P. Spence, and M. England, Barotropic Kelvin wave-induced bottom boundary layer warming along the West Antarctic Peninsula, *Journal of Geophysical Research: Oceans*, 124, 1595–1615, doi:10.1029/2018JC014227, 2019. 641

Weiss, J., The dynamics of enstrophy transfer in two-dimensional hydrodynamics, *Physica D*, 273–294, 1991. 286

Whalen, C. B., L. D. Talley, and J. A. MacKinnon, Spatial and temporal variability of global ocean mixing inferred from argo profiles, *Geophysical Research Letters*, 39(18), n/a–n/a, doi: 10.1029/2012GL053196, 2012. 1168

Wolfe, C., Approximations to the ocean's residual circulation in arbitrary tracer coordinates, *Ocean Modelling*, 75, 20–35, doi:10.1016/j.ocemod.2013.12.004, 2014. 1138

Young, W. R., Dynamic enthalpy, Conservative Temperature, and the seawater Boussinesq approximation, *Journal of Physical Oceanography*, 40, 394–400, doi:10.1175/2009JPO4294.1, 2010. 534, 536, 537, 538

Young, W. R., An exact thickness-weighted average formulation of the Boussinesq equations, *Journal of Physical Oceanography*, 42, 692–707, doi:10.1175/JPO-D-11-0102.1, 2012. 539, 1049, 1056, 1057, 1063, 1141, 1153

Zanna, L., S. Khatiwala, J. Gregory, J. Ison, and P. Heimbach, Global reconstruction of historical ocean heat storage and transport, *Proceedings of the National Academy of Science*, doi:10.1073/pnas.1808838115, 2019. 107

Zhao, R., and G. K. Vallis, Parameterizing mesoscale eddies with residual and Eulerian schemes, and a comparison with eddy-permitting models, *Ocean Modelling*, 23, 1–12, 2008. 1153

Index

- AABW, 1181
absolute circulation, 847
absolute momentum, 211
absolute reference frame, 178
Absolute Salinity, 1168
absolute salinity, 434, 557
absolute temperature, 236
absolute vorticity, 790
 impermeability, 900
absolute vorticity invariance, 775
acceleration, 177
 centrifugal, 191, 195
 centripetal, 191
 Coriolis, 191, 195
 inertial, 187, 193
 planetary Cartesian, 191
 spherical, 194
 spherical metric, 196
acoustic wave, 50
acoustic waves, 365, 486
action/reaction law, 369
active tracer, 46
adiabatic, 425
adiabatic lapse rate, 443
adjoint diffusion equation, 93
adjoint operator, 94
advection
 reversible, 468
advection equation, 46
advection of velocity vector identity, 30
advection operator, 253
advection-diffusion equation, 326, 1096
advection boundary transport, 1174
advection flux, 1170
advection time scale, 919
advection tracer flux, 315
ageostrophic components, 962
ageostrophic flow, 622
ageostrophic velocity, 622
air-water interface, 399
anelastic approximation, 332, 519
angular momentum, 178, 208, 216, 376, 792,
 798, 837
 axial, 223
 barotropic model, 783
 conservation, 216
 free vortex, 751
 shallow water, 724
 tracer, 1226
 water mass, 1226
angular velocity, 180
anisotropic tensor, 19
anti-cyclonic, 212
anti-cyclonic motion, 583
anticipated potential vorticity, 784
arc length, 31, 111, 117
arc-length, 335
Archimedean buoyancy, 514, 553, 554
Archimedes' Principle, 554
area
 surface, 115
area evolution
 incompressible, 341
arrested Ekman layer, 640

- aspect ratio, 498, 521, 522, 918, 920
association versus causation, 525
associative relations, v
asymptotic methods, 925
atmospheric form stress, 406, 409
atmospheric pressure torque, 815, 861
available potential energy, 538
 approximate, 544
 exact, 543
 QG fluid, 974
Avogadro's number, 236
axial angular momentum, 191, 361
 atmosphere, 363
 depth integrated, 416
 ocean, 414
 ring of air, 362
 steady state, 418
 zonal acceleration, 219
axial vector, 10

back-reaction, 553
backward diffusion equation, 93
balanced model, 959
baroclinic, 503, 588, 838
baroclinic velocity, 713, 859
baroclinicity, 583, 834, 838, 882
 Boussinesq, 852
 generalized vertical coordinates, 1035
 ideal gas, 865
 seawater, 884
baroclinicity
 solenoid, 838
barotropic, 503, 838
barotropic flow, 591, 835
barotropic model, 761
barotropic velocity, 713, 859
barotropization, 587
barycenter, 314
barycentric velocity, 243, 245, 313, 316, 321,
 371, 1169
basic one-form, 185
basis one-forms, 134
basis vector, 185
basis vectors, 125, 132
 changes, 184
Beltrami flow, 42
Bernoulli function, 902
Bernoulli potential, 472, 1042
 Boussinesq and hydrostatic, 892
 Bernoulli theorem
 traditional derivation, 474
 Bernoulli's Principle, 474
 Bernoulli's theorem, 473
 pipe flow, 475
 beta effect, 765, 775, 794, 811, 814, 848, 860,
 947
 two-dimensional example, 849
 beta plane, 765
 beta-effect
 topographic, 801, 928
 beta-plane approximation, 494, 495
 bi-orthogonality relation, 134
 Bjerknes circulation theorem, 848
 body force
 homogeneous layer, 810
 body forces, x, 354, 369
 boldface notation, 126
 Boltzmann constant, 449
 bolus velocity, 1051, 1057
 bottom drag, 531
 bottom geostrophic velocity, 861
 bottom kinematic boundary condition, 679
 bottom pressure torque, 815, 816, 840, 861,
 947
 bottom transformation, 1197
 bottom vertical velocity, 948, 949
 boundary Green's function, 85, 87
 boundary layer, 628
 boundary layers, 397, 622
 boundary propagator, 87, 100
 normalization, 102
 boundary stress, 531
 boundary velocity, 321
 boundary water mass transformation, 1202
 Boussinesq
 energetics, 532
 mass continuity, 515
 momentum equation, 514
 Boussinesq approximation, 331, 332, 512
 generalized, 547
 ocean equations, 516
 reference density, 516
 traditional, 519
 Boussinesq continuity
 density evolution, 518
 Boussinesq dynamic enthalpy, 534, 535

- Boussinesq equations
non-dimensional, 936, 942
- Boussinesq fluid, 578
energetics, 548
- Boussinesq reference density, 332
- Buckingham-II theorem, 916
- budget analysis, 318
- bulk viscosity, 386
- buoyancy, 500, 553, 1165
effective, 569, 571
globally referenced, 561
homogeneous fluid, 519
- buoyancy and height
relation, 542
- buoyancy budget
surface ocean, 1175
- buoyancy coordinates, 123
- buoyancy force, 556
- buoyancy frequency, 564
ideal gas, 565
- buoyancy sorting, 545
- buoyancy stratification, 561
- buoyancy work, 533
- Burger number, 921, 975
- cabbeling, 517
- capillary pressure, 400, 402
- capillary tube, 398
- capillary waves, 402
- capillary-gravity waves, 394
- Cartesian coordinates, 123, 151
summary, 153
- Cauchy problem
hyperbolic, 57
parabolic, 54
- Cauchy's equation of motion, 356
- Cauchy's fundamental lemma, 371
- Cauchy's solution, 272, 273
- Cauchy's stress principle, 371
- Cauchy's theorem, 373
- Cauchy-Green strain tensor, 272
- Cauchy-Stokes decomposition, 278
- causal free space Green's function, 92
- Causal Green's function, 92
- causal Green's function, 103
- causal relations, iv
- causality condition, 92
- causation versus association, 525
- center of mass motion, 307
- center of mass velocity, 314, 1170
- centrifugal acceleration, 202, 505, 606–608, 772
- particle motion, 212
- planetary orbital, 649
- planetary rotation, 649
- spherical, 196
- centripetal acceleration, 202, 606–608
- channel, 717
- channel flow, 598
- characteristic curves, 48
advection equation, 49
wave equation, 57
- chemical energy flux, 469
- chemical potential, 432, 434, 439
- chemical reactions, 311
- chemical work, 433
- Christoffel symbols, 145
metric connection, 146
related to metric, 147
- circuit, 32, 110
- circulation, 32, 39, 749
absolute, 847
around a streamline, 810
free vortex, 751
friction effects, 809
Rayleigh drag, 810
rotating fluids, 846
shallow water, 809
solid-body rotation, 753
tornado, 832
wind stress, 810
- circulation induction, 848
- circulation theorem, 809
- co-tangent space, 134
- coarse graining, 1116
- coherent structure, 269
- cold core eddy, 700
- column vorticity, 817
- compatibility
total mass + tracer mass, 315
- compression, 285
- conservation equation
Eulerian flux-form, 319
material form, 319
- conservation laws, 177, 209, 214, 230, 485
- conservation of indices, 125

- conservative force, 179
conservative property, 486
Conservative Temperature, 445, 480, 482, 557, 1165
conservative vector field, 41
constituent mass flux, 314
constitutive relation, 384, 385
contact force
 pressure, 372
contact forces, x, 354, 369
contact pressure force, 705
contact pressure forces, 703
continuity, 332
continuity equation, 292
continuum approximation, 231, 232
continuum hypothesis, 231
contour, 110
contraction, 9
contravariant tensor index, 122, 185
convective time derivative, 253
converging flow
 PV constraints, 803
coordinate invariance, 120
coordinate representation
 Cartesian, 6
Coriolis acceleration, 191, 217
 large-scale motion, 197
 shallow atmosphere, 197
 spherical, 197
Coriolis parameter, 197
Couette flow, 385, 395
couplet, 381
covariant curl, 149
covariant derivative
 metric tensor, 147
 one-form, 146
 scalar, 145
 vector, 145
covariant differentiation, 121
covariant divergence
 vector, 147
covariant Laplacian
 scalar, 148
covariant tensor index, 122, 185
Crocco's theorem, 487
curl integrated over a closed surface, 36
curl of a curl, 29
curl of cross product, 29
curl of vector, 26
curl-free vector, 27
curvature, 400
 curve, 113
 Gaussian, 116
 surface, 115
curvature circle, 113
curves on a surface, 117
cyclonic, 212
cyclonic eddy, 700
cyclonic motion, 583
cyclostrophic balance, 616
cylindrical coordinates, 123
cylindrical-polar coordinates, 158
D'Alembert wave solution, 58
deformation gradient tensor, 272
deformation radius, 808, 940, 961, 975
 shallow water, 731, 807, 921, 931
deformation rate, 275
deformation rates, 285
deformation tensor, 248
deformational flow, 285
delta sheets of potential vorticity, 539
density, 558
depth of no motion, 589
determinant
 exponential, 148
developing flow, 252
deviator, 387
deviatoric friction tensor, 387
deviatoric rate of strain tensor, 387, 464
deviatoric stress, 355, 386
dia-surface transfer
 shallow water, 690
dia-surface transport, 327, 1003, 1007
 incompressible, 1011
dia-surface velocity
 compared to vertical, 1014
dia-surface velocity component, 1010
diabatic, 425
diagnostic equation, v, 257
dianeutral direction, 563
diapycnal transport velocity, 305
diapycnal velocity component, 1040
differential forms, 165, 1218
differential mass form, 1220
diffusion, 1169

- diffusion dissipation functional, 1076
diffusion equation, 54, 92
diffusion tensor, 1129, 1170
diffusive fluxes, 316, 321
diffusive tracer flux, 315
diffusivity, 1170
 eddy, 530
dilation, 278
dimensionless numbers, 916
Dirac delta, 69
 Cartesian, 71
 sifting property, 70
 spherical, 71
Dirac delta distribution, 1197, 1208
Dirac delta sheet, 90, 100
direct stress, 386
direction cosines, 17
Dirichlet boundary condition, 52, 54, 80, 1097
Dirichlet boundary conditions, 77, 1098
dispersion relation
 inertia-gravity waves, 731
 inertial waves, 594
 shallow water gravity wave, 729
displacement tensor, 272
dissipation, 355, 465
distance, 8
distance to polar axis, 190
divergence, 389
divergence of cross product, 29
divergence of curl, 28
divergence operator
 Cartesian, 25
divergence theorem, 34, 291
 scalar fields, 35
diverging flow, 284
domain of influence, 59
Doppler shift, 779
dot product, 8
double integral formulate, 1205
double integral identity, 63
drag coefficient, 531
dual form stress, 705, 710
dual pressure form stress, 1060
duality
 Eulerian and Lagrangian, 271
duality between one-form and vector, 133
duality vectors & one-forms, 134
Dufour effect, 478
Duhamel's integral
 heat equation, 56
 wave equation, 60
dynamic enthalpy, 534
dynamic pressure, 610, 688
dynamic pressure gradient, 359
dynamic topography, 503
dynamic viscosity, 386
dynamical constraints, 208, 209, 214, 230
dynamical pressure, 500, 789
- Eötvös correction, 197
earth radius, 644
earth's angular rotation, 180
earth's mass, 199
earth's radius, 180, 199
earth's solid body speed, 180
eddy diffusivity, 530
eddy tracer fluxes, 1124
eddy viscosity, 531, 622
 Ekman layer, 629
eddy-induced velocity, 530, 551
effective beta, 802, 929
effective buoyancy, 553, 560, 569, 571
effective free surface height, 675
effective gravity, 495
Einstein index convention, 6
Einstein summation convention, 126
Ekman
 balance, 622
Ekman balance
 mass transport, 629
 natural coordinates, 624
 Rayleigh drag, 625
 spiral motion, 626
Ekman boundary layers, 397, 621, 622
Ekman downwelling, 633
Ekman layer
 arrested, 640
 bottom, 636
 horizontal transport, 629
 ocean surface, 629
 velocity profile, 636
Ekman layer thickness, 628
Ekman mechanics, 621
Ekman number, 627, 628
Ekman pumping, 633
Ekman suction, 633

- Ekman transport, 632
Ekman upwelling, 633
Ekman velocity, 948
Eliasson-Palm fluxes, 713
Eliassen-Palm flux, 711, 1051, 1061
Eliassen-Palm flux tensor, 704, 1061
elliptic PDE
 classification, 50
 properties, 50
energy
 total, 470
ensemble average, 1055
ensemble mean, 1117
enthalpy, 437, 480
 budget, 469
enthalpy budget, 471, 479
entropy, 429, 430, 436
 perfect fluid, 468
entropy budget, 476
entropy flux, 477
entropy source, 477, 479
epsilon product identity, 29
epsilon tensor product identity, 15
equation of motion, 356
 Cartesian, 201
 particle, 201
 rotating, 357
 spherical, 202, 358
equation of state, 434, 436–438, 448, 557
 linear, 517
 oceanic Boussinesq, 518
equatorial radius, 203
equilibrium tide, 645
equivalent barotropic depth, 768, 769
equivalent barotropic flow, 768
Ertel potential vorticity, 879, 881
Euclidean metric, 8
Euclidean space, 119
Euler equation, 357, 364, 389
 reversible, 366
Eulerian and Lagrangian duality, 271
Eulerian coordinates
 Cartesian, 123
 cylindrical, 123
 spherical, 123
Eulerian flux-form, 485
Eulerian mean, 1118
Eulerian reference frame, 230, 245
Eulerian region, 320
Eulerian time derivative, 122, 252
evolving flow, 252
exact differential, 37, 179, 429, 433
 hiker analogy, 39
exact differential form, 167, 169
exact differentials, 336
exact geostrophic balance, 611
exact hydrostatic balance, 359
Exner function, 455
extensive property, 318, 426
exterior algebra, 167
exterior calculus, 167
exterior derivative, 167, 1220
exterior moment of λ , 1193
external forces, 369
external pressure, 500
external tracer mass, 1216
external velocity, 713, 859
f-plane approximation, 211, 494, 495
fetch, 722
Fick's law of diffusion, 478
field point, 69, 75
first kinematic viscosity, 386
first law of thermodynamics, 428, 431, 471
 moving fluid, 468
 potential temperature, 470
flat space, 119
flow lines, 262
fluid dynamics, x
fluid element, 245, 312, 316, 553
fluid kinematics, x
fluid material region, 244
fluid parcel, 244
 constituents, 313
fluid particle, 243
fluid particle trajectories, 246
fluid particle trajectory, 262
fluid region, 245
flux, 43
flux-form conservation law, 290, 887
flux-form conservation laws, 485
force couplet, 381
force potential, 179
forces
 body, 354
 contact, 354

- form stress, 372, 703, 705, 708, 709
dual, 705, 710
geostrophic eddies, 598
isopycnal layer, 601
isopycnal surface, 599
mathematical expression, 407
formation, 1202
Fourier's heat law, 327
Fourier's law, 469
Fourier's law of conduction, 478
free particle, 214
free space Green's function, 75
free vortex, 749, 758
freely falling particle, 226
freshwater boundary condition, 1170, 1171
freshwater budget, 1168
freshwater velocity, 1170
friction
 vertical shear, 390
friction driven velocity, 622
friction tensor, 385
frictional dissipation, 464, 471
frictional force, 388
frictional torque, 952
Froude number, 255, 522, 692, 917, 919, 921
functional degrees of freedom, 332
functional variations, 1076
functions versus scalars, 135
fundamental solution
 Laplace operator, 75
fundamental theorem of calculus, 32
fundamental thermodynamic relation, 430

Galilean boost, 178
Galilean invariance, 177, 182, 257, 386, 784
 kinetic energy, 222
Galilean transformation, 178, 257, 319, 734
 material time operator, 258
gauge freedom, 889, 911
gauge function, 472, 801
gauge symmetry, 336, 338, 472
 pressure, 355
gauge transformation, 39
Gauss's divergence theorem, 34, 149, 1031
Gaussian curvature, 116
general covariance, 120
general orthogonal coordinates, 163
general vertical coordinates, 322
basis one-forms, 989
basis vectors, 988
circulation, 1001
common confusion, 987
contravariant velocity, 993
covariant velocity, 994
diffusion operator, 999
divergence, 998
divergence theorem, 998
Jacobian, 995
layer integrated diffusion, 1000
Levi-Civita tensor, 996
material time derivative, 998
metric tensor, 994
partial derivatives, 996
position vector, 991
related to Cartesian, 986
specific thickness, 995
triple product identity, 990
vector, 992
vector cross product, 996
velocity, 993
volume element, 996
vorticity, 1000
generalized Lagrangian mean, 1115
generalized vertical coordinates, 247, 982
 examples, 1004
geopotential, 200, 518
 height, 452
 reference, 460
 thickness, 452
geopotential coordinates, 203
geopotential height
 ideal gas, 504
geostrophic adjustment, 804
 potential vorticity, 807
geostrophic advection, 969
geostrophic balance, v, 577, 581
 exact, 611
geostrophic contours, 802, 1113
geostrophic eddies, 598
geostrophic eddy, 699
geostrophic number, 938
geostrophic streamfunction, 926, 928
geostrophic transport, 698
geostrophy
 isopycnal models, 1042
 pressure coordinates, 584

- shallow water, 696
geothermal heating, 327
Gibbs potential, 438, 467
Gibbs relation, 430
Gibbs-Duhem, 433
GLM, 1115, 1120
 isopycnal, 1131
GLM and tracers, 1123
GLM kinematics, 1117
global conservation law, 887
global conservation laws, 486
Godfrey's island rule, 819
gradient operator, 22
 Cartesian and GVC, 997
 general vertical coordinate, 997
gradient wind
 regular high, 619
gradient wind balance, 617
 barotropic flow, 771
gravest vertical mode, 587
gravitational acceleration, 198, 646
 approximate, 199
 effective, 200
gravitational potential, 644
 earth-moon, 649
 general case, 644
gravitational potential energy, 199, 459
 mixing, 462
 regional, 460
 stratification, 461
gravitational stability, 564
gravitational work, 475
gravity wave speed, 729
gravity waves, 748
 non-rotating, 806
Green's function
 advection-diffusion equation, 1098
 diffusion, 92
 diffusion adjoint, 93
 diffusion equation, 92
 diffusion free space, 92
 free space for Laplace, 75
 non-closed reciprocity, 1101
 one-dimensional Poisson, 77
 reciprocity, 83, 94, 1099
Green's function method, 68
Green's functions
 passive tracers, 1096
 Green's identities, 36
 Green's theorem, 170
 group velocity
 Rossby waves, 780
 GVCs, 982
 gyre circulation, 813
 gyres and channels, 724
Hadley circulation, 363
haline, 558
haline contraction, 441
haline contraction coefficient, 558, 559
Hamilton's operator, 22
harmonic function, 26, 28, 336, 748
 mean-value property, 51
harmonic functions, 51
Haynes-McIntyre PV flux, 898
heat capacity, 440
heat equation, 54, 92
heating, 428
Heaviside step function, 72, 92, 805
height and buoyancy
 relation, 542
Helmholtz decomposition, 39, 336
Helmholtz decomposition of Coriolis, 43
Helmholtz equation, 60
Helmholtz free energy, 438
Helmholtz's first theorem, 832
Helmholtz's second theorem, 832
Helmholtz's third theorem, 833
heuristics, 68
homogeneous function, 432
homogeneous solution, 50
homogeneous tensor, 18
Hooke's Law, 386
Hughes gyre model, 816
hydrostatic approximation, 359, 518, 570, 683
 background state, 499
 ocean models, 517
 scaling, 498, 520
 vertical motion, 492, 504
hydrostatic balance, 359, 378
 approximate, 492, 496
 exact versus approximate, 361
 forces, 379
 torques, 381
hydrostatic fluid, 221
hydrostatic number, 938

- hydrostatic pressure, 555, 674
evolution, 496, 524
expressions, 496
- hydrostatic primitive equations, 492
- hyperbolic PDE
classification, 50
domain of influence, 59
- hyperbolic PDEs, 57
- hypsometric equation, 452
- ice skater, 792
- ideal gas, 448
adiabatic lapse rate, 451
buoyancy frequency, 565
compressibility, 451
enthalpy, 450
equation of state, 448
geopotential thickness, 452
heat capacity, 449
internal energy, 449
lapse rate, 565
potential temperature, 453
sound speed, 451
thermal expansion coefficient, 450
- ideal gas law, 236
- impermeability vs. material invariance, 901
- Impermeability, 972
- impermeability
absolute vorticity components, 900
kinematics, 899
- impermeability theorem, 897, 908
confusions, 901
planetary geostrophy, 945
seawater, 901
- impulse, 74
- impulse response function, 74, 78, 103, 1105
- in situ density, 559
- in situ temperature, 445, 480
evolution, 480
- incompressible flow, 332
- incrop, 539
- incropping buoyancy, 539
- index convention, 6
- index gymnastics, 14, 125
- index raising and lowering, 134
- induction vector, 972
- inertia-gravity waves, 598, 730, 731
topographic forcing, 734
- inertial acceleration, 187
- inertial frequency, 580
- inertial motion, 613
- inertial oscillation, 614
- inertial oscillations, 211, 592, 940
ocean, 213
period, 213
- inertial reference frame, 177, 178, 230
- inertial velocity, 187
spherical, 194
- inertial waves, 577, 592
dispersion relation, 594
group velocity, 595
near, 596
- inexact differential, 38, 428, 429, 433
integrating factor, 38
- inexact differential form, 169
- influence function, 78
- information entropy, 1108
- initial value problem, 805
- initial value problems, 102
- injection work, 472
- inner product, 8
- integral curve, 243, 262
- integral surface, 48
- integrating factor, 429, 433
- integration in λ -space, 1189
- intensive property, 318, 426
- interfacial form stress, 406, 412, 705, 708
- interior moment of λ , 1193
- interior transformation, 1197
- intermolecular forces, 399
- internal energy, 427, 436, 470, 537
- internal energy budget, 471
- internal forces, 354, 369
- internal pressure, 500
- internal tracer mass, 1216
- internal velocity, 713, 859
- inverse barometer response, 675
- inverse energy cascade, 769
- inversion, 804
- irreversible process, 430
- irreversible thermodynamics, 457
- irrotational, 747
- irrotational flow, 337, 748, 755
- irrotational vector, 27
- island rule, 819
- isopycnal coordinates, 123

- isopycnal ensemble, 1131
isopycnal GLM, 1131
isopycnal layer transport, 1130
isopycnal mean, 1131
isopycnal primitive equations, 1038
isopycnals, 305
isotropic tensor, 18
- Jacobian
ratio of volumes, 250
Jacobian evolution, 283
Jacobian of transformation, 132, 139
Jacobian on a curve, 118, 1113
Jacobian operator, 118, 766, 969, 1113
JEBAR, 956
jets, 778
Joule heating, 471
Joule unit of energy, 237
- Kelvin waves, 598, 740
Kelvin's circulation theorem, 833, 839, 880, 882
Kelvin-Helmholtz instability, 939
kinematic b.c.
buoyancy surface, 305
geometric derivation, 301
incompressible flow, 332
material interface, 300
moving material, 299
ocean free surface, 304
permeable surface, 302
static material, 298
kinematic boundary condition, 298
kinematic boundary conditions, 679
tracers, 1170
kinematic free surface equation, 305
kinematic viscosity, 386
kinematic water mass transformation, 1196
kinematics of two-dimensional flow, 283
kinetic energy, 179, 222, 463
axial angular momentum, 223
Boussinesq, 532
cartesian coordinates, 222
geopotential coordinates, 223
hydrostatic Boussinesq, 548
spherical coordinates, 223
kinetic energy dissipation, 464, 465
kinetic stress, 382, 701, 858, 1059
- kinetic stress tensor, 764
kinetic theory, 236
Knudsen number, 233, 456
Kronecker tensor, 8, 131, 185
- laboratory reference frame, 178, 245
Lagrangian coordinates, 124
Lagrangian mean, 1118
Lagrangian reference frame, 230, 245
Lagrangian region, 316
Lagrangian time derivative, 122, 253
laminar subregion, 397
Laplace's equation, 50, 51, 748
properties, 51
Laplacian friction, 627
Laplacian operator, 26
law of cosines, 649
layer mass continuity, 1020
compressible, 1021
pressure coordinates, 1022
Legendre transformation, 435, 437, 438
Leibniz's rule, 320, 854
Leibniz-Reynolds
brute force, 324
Leibniz-Reynolds transport theorem, 320, 908
length scale
gradient, 235
macroscopic, 232
mean free path, 232, 237
measurement, 235
molecular, 232
simulation, 235
level of no motion, 503, 683
Levi-Civita tensor
general coordinates, 141
volume element, 142
Levi-Civita tensor, 9, 140, 770
lid pressure, 769
linear equation of state, 560
linear momentum, 177
conservation, 177
finite volume, 383
Lagrangian volume, 384
local conservation laws, 486
local Rossby number, 610
local thermodynamic equilibrium, 457, 466
locally referenced potential density, 565
longitudinal waves, 365, 729

- longwave radiation, 1174
Lorenz convention, 539
- Mach number, 366, 511
macroscopically small, 231
Magnus acceleration, 359, 610, 789
manifold, 165
Margules' relation, 697, 700
Marshall potential vorticity flux, 902
mass
 λ -layer, 1189
mass budget
 λ -layer, 1202
 fluid column, 306, 308
 weak formulation, 313
mass budgets, 321
mass conservation, 230, 294
 arbitrary Eulerian region, 291
 Eulerian flux-form, 290
 finite Eulerian cube, 290
 Jacobian derivation, 293
 Lagrangian, 292
 Lagrangian derivation, 295
 multi-components, 313
mass continuity
 generalized vertical coordinates, 1018
 layer integrated, 1020
mass density, 557
mass density function, 1190
mass distribution function, 1190
mass equation
 general vertical coordinates, 1028
mass flux, 290
mass flux through a surface, 302
mass-Sverdrup, 1186
material area evolution, 278
material constant, 319
material coordinates, 124, 246, 248
 choices, 247
material curve, 271
 evolution, 273
material fluid parcel, 244
material invariance, 255, 257
material invariant, 319
material line elements, 841
material surface, 255
material time derivative, 122, 253
 general vertical coordinates, 1013
invariance, 259, 260
isopycnal models, 1040
space-time, 262
material volume evolution, 282
materially closed system, 427
materially invariant volume, 332
mathematically rigorous, 68, 69
matrices, 18
matter concentration, 431
matter conservation, 485
Maxwell relations, 442
mean tracer equation, 1139
mechanical energy, 463, 465
 Boussinesq, 532–534, 536, 548
 conservation, 224
 particle, 222
mechanical energy conservation, 179
mechanical energy dissipation, 393, 464
mechanical flux, 465
mechanical pressure, 384, 388
mechanical stress, 382
mechanical work, 179, 222, 429
meridional overturning circulation, 342
meridional transport, 947
method of characteristics, 48
metric acceleration, 196
metric acceleration terms, 359
metric tensor, 130, 272
 Cartesian coordinates, 130
 coordinate representation, 131
 coordinate transformation, 132
 determinant, 139
 general, 131
 inverse, 134, 135
 relating vectors & one-forms, 134
metricity condition, 147
microscopically large, 231
mixed boundary condition, 1097
mixing
 surface boundary, 1171
modified mean, 1131
moist static energy, 483
mole, 236
molecular composition
 air, 236
 water, 236
molecular diffusivity, 548
molecular viscosity, 393, 548

- moment arm, 191, 216
moment of inertia, 792
moment-arm, 361
moments of λ , 1193
momentum
 angular, 178
 approximate axial angular, 221
 axial angular, 191, 215, 361, 508
 axial angular and Coriolis, 217
 linear, 214
 potential, 210, 214
momentum equation, 701
 flux form, 382, 1029
 general vertical coordinates, 1028
 horizontal, 1029
 natural coordinates, 609
 spherical, 358
 vector-invariant, 359, 1029
 vertical, 1029
momentum flux, 290
Monge gauge, 109
Montgomery potential, 601, 1039
motional forces, 569
motional vertical acceleration, 572
Munk gyre model, 816

NADW, 1181
natural boundary condition, 1076
natural coordinates, 605, 606
 isobars, 623
 vorticity, 753
Navier-Stokes equation, 357, 389, 391
 non-dimensional, 392
near inertial waves, 596
nearly horizontal motion, 221
net stress tensor, 390
Neumann boundary condition, 52, 54, 328, 1076, 1097
Neumann boundary conditions, 77, 88, 1098
neutral density, 1186
neutral directions, 560, 562
neutral helicity, 568
neutrality condition, 563
neutrally buoyant, 553, 556
Newton's equation of motion, 356
Newton's gravitational constant, 198
Newton's gravitational law, 198
Newton's second law, 177
Newton's third law, 369, 371, 394, 709
 pressure, 372
Newtonian fluid, 385, 386
Newtonian gravity, 69
Newtonian gravity field, 649
Newtonian time, 119, 122
no normal flow b.c., 298
no-flux boundary condition, 327, 640
no-slip boundary condition, 395
Noether's Theorem, 209, 230
non-advection flux, 1170
non-advection transport, 1174
non-conservative forces, 179
non-conservative process, 486
non-dimensionalization, 922
 Ekman balance, 627
 hydrostatic approximation, 521
non-dispersive waves, 729
non-divergent barotropic, 762
non-divergent barotropic model, 761
non-divergent velocity field, 332
non-hydrostatic pressure, 676
non-hydrostatic primitive equations, 494
non-inertial reference frame, 178
non-Newtonian constitutive relations, 391
non-orthogonal coordinates, 124
normal derivative, 24
normal direction, 24, 110
normal stress, 386
normal vector, 335

obduction, 1017
oblate spheroid, 203
oblique coordinates, 136
observation point, 75
ocean buoyancy, 1165
ocean free surface, 1005
ocean gyres, 811
 topographic form stress, 814
ocean mesoscale eddy, 699
ocean mixed layer base, 1005
ocean-atmosphere equations, 483
oceanic Boussinesq, 511, 512
 mass, 519
 weight, 519
oceanic form stress, 406
one-forms, 133
 basis, 134

- coordinate representation, 133
one-way wave equation, 57
orbital motion, 649
orientable, 167
orthogonal curvilinear coordinates, 163
oscillator equation, 212
outcrop, 539
outcropping buoyancy, 539
outer product, 18, 121, 382, 701
outward normal
solid-earth bottom, 298
overturning streamfunction, 342
- parabolic PDE
classification, 50
parabolic PDEs, 54
smoothing property, 55
parameterizations, 235
partial derivatives
transformation, 144
particle displacement, 1124
particle mechanics, 177
particle trajectory
fall to equatorial plane, 226
free fall to center, 226
spherical motion, 226
particular solution, 50
passive tracer, 46
passive tracer source, 1098
passive tracers, 317, 1096
path, 110
orientable, 110
simple, 110
path dependent integrals, 428
path integral
arc length, 31
scalar function, 30
vector function, 32
pathlines, 262
analog to car flow, 263
perfect fluid, iii, 244, 425
periodic channel, 717
permutation symbol, 9, 140, 770
product identity, 15
phase averaging, 1116
phase speed
Rossby waves, 779
physically formal, 68, 69
- physics as geometry, 120
planar curves, 111
planetary beta effect, 587
planetary Cartesian coordinates, 188, 189
planetary centrifugal acceleration, 191, 202
planetary centripetal acceleration, 191
planetary geostrophic potential vorticity, 802
planetary geostrophy, 584, 923
continuously stratified, 935
equations, 943
potential vorticity, 944
vorticity equation, 585
planetary induction, 848
planetary spherical coordinates, 188
planetary vorticity, xi, 587
Poincaré waves, 730
point mass, 69
point transformation, 248
points, 6
Poisson equation, 198, 768, 770
Green's function, 80, 88
pressure, 572
Poisson's equation, 50, 53, 75
Dirichlet boundaries, 85
max-min principle, 53
Neumann boundaries, 89
polar coordinates, 158
polar radius, 203
polarized gravity waves, 733
position coordinates, 248
position vector
Cartesian representation, 188
spherical representation, 188
postulate of total energy conservation, 479
potential density, 559
locally referenced, 565
potential energy, 179
Boussinesq, 533
shallow water layer, 1060
potential enthalpy, 481
potential flow, 748
potential momentum, 202, 214
potential properties, 557
potential scalar properties, 487
potential temperature, 444, 445, 480, 557
evolution, 481
specific entropy, 446
potential vorcity

- Eulerian flux-form budget, 887
potential vorticity, 1001
 substance, 888
 baroclinic fluid, 882
 barotropic fluid, 880
Bernoulli potential, 892
Boussinesq, 1036
cylinder, 792
delta sheet, 539
derivation, 889
dynamical tracer, 883
Ertel, 879, 881, 967
flux, 892, 902
flux vector, 888, 890
flux-form equation, 800
fluxes, 898
gauge choice, 892
generalized vertical coordinates, 1034
hydrostatic Boussinesq, 889
impermeability, 897, 945
induction vector, 972
inversion, 804
isopycnal coordinates, 1043
Kelvin circulation theorem, 796
kinematic derivation, 891
Marshall flux, 902
name, 796, 883
non-conservative processes, 799, 886
ocean layer, 907
perfect fluid, 880
planetary geostrophy, 802, 931, 944
potential density, 884
preferred form of flux?, 911
quasi-geostrophic, 967
quasi-geostrophy, 927, 928, 931
Rossby, 795
seawater, 901
seawater equation of state, 884
shallow water, 794
substance, 898
symmetric instability, 911
TWA, 1063
 two-dimensional non-divergent, 766
potential vorticity “stuff”, 801
potential vorticity budget
 layer, 907
potential vorticity concentration, 904
potential vorticity flux, 801
air-sea boundary, 910
land-sea boundary, 909
 steady state constraints, 903
potential vorticity impermeability
 kinematics, 899
potential vorticity substance
 integrated, 904
power, 222
practical salinity, 1168
Prandtl ratio, 940
pressure
 contact force, 1032
 contact stress, 1031
 dynamic, 688
 dynamical pressure, 513
 gradient force, 1031
 interfacial form stress, 1033
 mechanical, 384, 388
 Pascal unit, 237
 poisson equation, 770
 Poisson equation, 524
 sources, 525
 stagnation, 474
 standard atmosphere, 236
 stress, 371
 thermodynamic, 384
 total, 474
pressure driven velocity, 622
pressure force, 705
 two-dimensional flow, 772
pressure forces, 703
pressure form stress, 372, 703, 705
 dual, 1060
 shallow water, 708
pressure gradient
 bottom, 501
 external, 500, 501
 hydrostatic, 500, 501, 523
 ideal gas, 504
 internal, 500, 501, 523
pressure gradients
 hydrostatic, 583
pressure torque, 947
 atmosphere, 951
 bottom, 952
pressure torques, 712, 860
pressure transport, 486
pressure work, 429, 475

- primitive equations, 218
primitive equations, 492, 578
 nomenclature, 494
process water mass transformation, 1196
prognostic equation, v, 61, 257
pseudo angular momentum, 211
pseudo vector, 10
pseudo-vector, 746
pycnocline, 699
Pythagoras' Theorem, 8
- quasi-equilibrium thermodynamics, 457
quasi-geostrophic potential vorticity, 927, 928
quasi-geostrophy, 924
quasi-Lagrangian coordinates, 124
quasi-static process, 231, 238, 426, 457
quasi-Stokes transport, 1133, 1137
quotient rule, 375
- radius of curvature, 113, 115, 607, 614, 753
rarefied gas, 234
rate of strain
 vorticity source, 842
rate of strain tensor, 275, 276, 284, 285, 464,
 526, 746, 770
Rayleigh drag, 179, 386, 624, 810, 855
re-entrant channel, 717, 778
real fluid, iii
reciprocity
 diffusion equation, 94
 Poisson equation, 79, 83
rectification, 1052
rectified, 1115
reduce gravity model, 699
reduced gravity, 685, 697
reduced gravity model, 503, 683
reference density, 332
reference frames
 absolute, 182
 inertial, 182
 non-inertial, 182
 rotating, 182
reference geopotential, 460
reference pressure, 559
region following the flow, 293
relative chemical potential, 434
rescaled geopotential coordinate, 1023
response function, 74, 103, 440
- reversibility, 209
reversible process, 430
Reynolds averaging, 1055
Reynolds decomposition, 1116
Reynolds number, 234, 392, 627, 641
 atmosphere, 393
 Gulf Stream rings, 393
 turbulence, 392
Reynolds stress, 382, 1058, 1059
Reynolds transport theorem, 293, 296
 linear momentum, 384
 multi-component fluid, 322
Richardson number, 522, 939
Riemannian differential geometry, 120, 129,
 165
right hand rule, 9
rigid body rotation, 278
rigid lid approximation, 769, 956
Robin boundary condition, 1097
Rossby deformation radius, 940
Rossby number, 579, 641, 919–921, 938
 Gulf Stream ring, 580
 kitchen sink, 580
 local, 610
Rossby potential vorticity, 799
Rossby waves, 495, 598, 778
 dispersion relation, 779
 group velocity, 780
 phase speed, 779
 stationary, 779
rotating tank, 505, 724, 772
 parabolic free surface, 507
 solid body motion, 506
rotation matrix, 17
rotation tensor, 275, 277, 285, 526, 746, 770,
 842
rotational flow, 285
rotational fluxes, 1126
- salinity, 557, 1165
salt boundary condition, 1170, 1171
salt budget, 1168
salt concentration, 1168
salt velocity, 1170
scalar field, 6
scalar potential, 27
scalar product, 8
 general tensors, 136

- invariance, 17
scalar streamfunctions, 340
scale analysis, 520, 916
scale height, 452
scaling
 homogeneous function, 432
sea breeze, 865
seawater chemical potential, 434, 439
seawater equation of state, 884
seawater mass distribution, 1190
second kinematic viscosity, 386, 388
second law of thermodynamics, 429–431
self-adjoint, 84, 1075
self-adjoint operator, 94
self-advection, 770
self-attraction and loading, 653
sgn function, 72
shallow fluid approximation, 218, 221, 492
shallow water
 gravity wave dispersion, 731
 gravity wave speed, 729
 hydrostatic approximation, 683
 inertia-gravity waves, 730
 Kelvin waves, 740
 linear potential vorticity, 730
 linear relative vorticity, 727
 linearized equations, 726
 momentum equation, 701
 non-dimensional, 922
 non-rotating gravity waves, 728
 planetary geostrophy, 923
 quasi-geostrophy, 924
shallow water columns, 683
shallow water equations
 N -layers, 706
shallow water model
 formulation, 674
 stacked, 686
shallow water momentum, 677
shallow water potential vorticity, 794
shallow water pressure, 674
shallow water PV invariance, 795
shallow water Sverdrup balance, 812
shallow water thickness, 677
shallow water vertical velocity, 681
shear flow, 547
shear stresses, 386
shearing strain, 285
shortwave radiation, 1174
sign-function, 805
simple ideal gas, 448, 449
simply connected surface, 568
singular limit, 628
skew symmetric stirring tensor, 1129
skew symmetric tracer fluxes, 1124
slippery Ekman layer, 640
slope of a surface, 303
smooth velocity field, 262
solenoid, 838
solid body rotation, 182
solid body velocity
 spherical, 191
solid-body rotation, 41, 752, 757, 758
Sorret effect, 478
sorting buoyancy, 545
sound speed, 365, 441, 558
 ideal gas, 451
source point, 69, 75
Southern Ocean channel, 717
space increments
 Eulerian, 270
 Lagrangian, 270
specific entropy, 446
specific gas constant
 air, 448
 atmosphere, 452
specific thickness, 1006
specific volume, 467
spherical coordinates, 123, 153, 186
 basis one-forms, 155
 basis vectors, 155
 differential operators, 157
 Levi-Civita tensor, 156
 metric, 156
 position, 156
 summary, 158
 transformation from Cartesian, 154
vector components, 157
vector cross product, 157
velocity, 156
 volume element, 156
spin, 526
splat, 526
stacked shallow water model, 686
standard atmosphere, 236
standard pressure, 232

- standard temperature, 232, 236
state function, 428
static equilibrium sea level, 644, 645
static forces, 553, 569
stationary Rossby waves, 779
steady flow, 252, 253
steady state, 252, 257, 319, 425, 467, 723
steady state solution, 50
steepest descent, 23
step response function, 104
steric sea level, 503
stirring
 reversible, 468
Stoke's theorem, 149
Stokes drift, 1051, 1122, 1128
Stokes mean, 1117, 1120
Stokes' theorem, 33, 170, 749
 generalized, 170
Stokes-Cartan theorem, 170
Stokesian fluid, 385
Stommel gyre model, 816
storage, 1202
strain, 285
stratification, 561
streaklines, 264
streamfunction, 336, 762
 meridional-depth overturning, 342
quasi-geostrophy, 928
streamlines, 334
 transport, 335, 338–340
two-dimensional flow, 334
vector, 338
 vertical gauge, 338
streamlines, 263, 298, 334
streamlines versus pathlines, 264
streamtubes, 263
stress
 boundary, 531
 contact, 354
 deviatoric, 355, 386
 direct, 386
 friction, 355
 normal, 386
 pressure, 355
 shallow water, 692
 shear, 386
stress boundary condition
 Lagrangian interface, 396
permeable interface, 396
solid boundary, 394
stress on an interface, 394
stress tensor, 354, 369, 373
 net, 390
 symmetry, 376
stress-strain relation, 384
stresses
 local equilibrium, 371, 394
stretching, 843
 material lines, 276
strong formulation, 242, 324, 379, 381
sub-inertial frequency, 598
subduction, 1017
subgrid-scale parameterizations, 585
subharmonic function, 53
substantial time derivative, 253
super-inertial frequency, 598
super-inertial waves, 731
superposition property, 69
surface
 orientable, 110
surface area, 115
 horizontal projection, 303, 304, 326
surface derivative operator, 279
surface gravity waves, 676
surface tension, 369, 371, 394, 398, 429
 gas bubbles, 402
surface transformation, 1197
 buoyancy, 1208
surface velocity, 299
surface water mass transformation, 1206
 circulation, 1207
Sverdrup, 1186
Sverdrup balance, 812
symmetric mixing tensor, 1129
symmetric tracer fluxes, 1124
symmetries, 207, 209, 230
 spatial, 214
symmetry, 177
tangent direction, 23, 110
tangent plane approximation, 494
tangent plane Cartesian coordinates, 189
tangent space, 134
tangent vector, 31, 335
Taylor columns, 587, 683
Taylor curtains, 591

Taylor series, 22
Taylor-Bretherton identity, 767, 1051, 1063
Taylor-Proudman effect, 577, 586, 591, 597
temperature, 445, 557
tensile force, 399
tension strain, 285
tensor analysis
 algebra summarized, 184
 contraction of indices, 121
 covariant differentiation, 121
 permissible operations, 121
 quotient rule, 121
 tensor fields, 6
 tensors, 6
 tensors and matrices, 18
 uses for GFM, 2
tensor operations, 121
tensor product, 121
tensor quotient rule, 121
test charges, 243
test fluid element, 245, 553, 560
test paddle wheel, 746
test particles, 243
tetrahedron fluid region, 373
theorem of stress means, 366
thermal equation of state, 448, 557
thermal expansion, 333, 441
thermal expansion coefficient, 558, 559
thermal wind, 700
 atmosphere, 590
 shallow water, 697
thermal wind balance, 577, 584, 963
 Antarctic Circumpolar Current, 588
 diagnostics, 589
 ocean, 588
thermo-hydrodynamical equations, 483
thermobaricity, 517
thermodynamic configuration space, 535
thermodynamic equilibrium, 425
thermodynamic potentials, 435
thermodynamic pressure, 384
 incompressible flow, 385
thermodynamic relation
 specific, 434
thermodynamic relations
 steady state, 467
thermodynamic systems
 exchanges, 425

thermodynamics
 moving fluid, 466
thermohaline circulation, 512, 518
thickness
 specific, 1006, 1028
 specific thickness, 1040
thickness diffusion
 Gent-McWilliams, 601
thickness equation, 1023
 isopycnal models, 1040
thickness weighted averaging, 1056
thickness weighted mean, 1139
tidal acceleration, 646
tides
 earth-moon system, 647
 realistic effects, 652
 semi-diurnal, 647
tilting, 843
 material lines, 276
time
 coordinate, 122
 parameter, 122
time averages, 61
time derivative notation, 297
time derivatives
 summary, 324
time mean, 1116
time reversal symmetry, 209
time scale
 molecular collisions, 238
time tendency, 61, 252
topographic beta, 801, 928, 929
topographic form stress, 406, 409
 components, 411
 gyres, 814
 zonal ridge, 409
topography
 two-dimensional non-divergent, 766
topological manifold, 427, 1183
tornado, 832
tornadoes, 617
torque, 178
torque density, 377
torques, 837, 946
total energy, 470
total time derivative, 252
tracer boundary condition
 no-normal derivative, 327

- tracer boundary conditions, 326
air-sea boundary, 328
bottom, 327
- tracer budgets, 321, 326
- tracer concentration, 314, 431
- tracer equation, 312, 314
derived, 314
Eulerian and Lagrangian forms, 315
general vertical coordinates, 1028
layer integrated, 1024
mean, 1139
shallow water, 682
- tracer fluxes
rotational, 1126
skew symmetric, 1124
symmetric, 1124
- tracer mass
external, 1216
internal, 1216
- tracer mass analysis, 1181, 1210
external, 1216
internal, 1216
special cases, 1214
- tracer mechanics, xi, 230
- tracer transport tensor, 1140
- tracers
passive, 317
- traditional approximation, 493
- trajectories, 246
- trajectory, 122, 177, 262
generalized vertical coordinates, 1016
- transformation
Cartesian, 17
Cartesian coordinates, 15
geometric, 17
inverse, 16
jacobian, 248
material and position, 248
orthogonal, 16
rotation, 16
tensor, 18
- transformation matrix, 132, 272
coordinates, 132
determinant, 140, 143
discrete algorithm, 249
general vertical coordinates, 986
inverse, 144
Jacobian, 132, 139
- space-time components, 143
- transformation of vectors, 186
- transformed residual mean, 1132
- transforming Cartesian to spherical, 191, 195
- transition to turbulence, 392
- translation, 278
- translation motion, 752
- transport, 43
- transport streamfunction, 335
- transport theorem
scalar field, 321
- transverse waves, 365, 594
- TRM, 1132
- TRM velocity, 1134
- turbulence, 389, 392
- turbulence closure, 235
- turbulent cascade and vortex stretching, 845
- turbulent cascade and vorticity, 748
- turbulent energy cascade, 393
- turbulent transport, 1174
- two-dimensional flow, 761
kinematics, 283
- two-dimensional incompressible, 849
- two-dimensional turbulence, 769
- two-layer shallow water, 686
- ultraviolet catastrophe, 393
- uniqueness
diffusion equation, 99
Dirichlet problem, 80, 81
Neumann problem, 89
- unit vector, 8
- unit vectors
rotation, 24
- universal gas constant, 236, 448
- vanishing layers, 1043
- vector field, 6
- vector invariant, 610
- vector invariant velocity equation, 359, 788
- vector product, 9
general coordinate, 142
geometric, 11
- vector rotation, 182
- vector streamfunction, 338
- vector-invariant velocity equation, 836
- vectors, 6
- velocity, 177

- basis vectors, 187
coordinate, 187
particle motion, 187
spherical, 194
surface, 299
velocity circulation, 749
velocity equation
 vector-invariant, 359, 1029, 1042
velocity gradient tensor, 275
velocity potential, 336, 341
velocity scale
 molecular rms speed, 238
velocity self-advection, 772, 836
velocity vector, 124
 generalized vertical coordinates, 1016
 planetary Cartesian representation, 189
 spherical representation, 190
vertical Ekman transport, 632
vertical gauge, 338
vertical shear, 276
vertical stiffness, 591, 592
vertical stiffening, 586
vertical stratification, 564
 energetics, 461
vertical velocity, 1014
 decomposed, 1014
 shallow water, 681
 two-dimensional non-divergent, 764
viscosity, 385
 air, 235
 dynamic, 235
 eddy, 531
 kinematic, 235
 water, 235
viscous dissipation, 465
viscous flux, 465
viscous tensor, 385
volume
 n -space, 14
 between isosurfaces, 880
 defined by vectors, 13
 element for integration, 14
volume budget
 column of water, 305
 material region, 293
volume element
 general coordinates, 142
 invariant, 139
volume equation, 293
volume evolution, 321
 incompressible, 341
volume-Sverdrup, 1186
vortex force, 359
vortex line, 830
vortex lines, 276, 843
vortex stretching, 795, 947
vortex tilting, 795
vortex tube, 830
vortex tubes, 843
vortical flow, 337
vortical motion, 914
vorticity, 39, 277, 283, 285, 389, 474, 746, 765
 absolute, 752, 790
 baroclinicity, 834
 barotropic flow, 835
 beta effect, 947
 bottom boundary, 855
 Boussinesq, 850
 curvature, 755, 775
 depth averaged velocity, 954
 depth integrated, 854
 depth integrated velocity, 953
 dynamics, 836
 free vortex, 749
 frozen-in, 842
 general vertical coordinate, 1029
 generalized vertical coordinates, 1034
 hydrostatic, 850
 isopycnal coordinates, 1044
 isopycnal models, 1042
 line element rotation, 746
 meridional transport, 951
 natural coordinates, 753
 non-divergent, 747
 of a column, 817
 orbital, 755, 775
 planetary, 752
 planetary geostrophic, 585
 quasi-geostrophy, 927
 relative, 752, 1029
 rotating fluids, 846
 rotating reference frame, 747
 shallow water, 789
 shear, 755, 775
 stretching, 843, 947
 surface boundary, 856

- tilting, 843
torques, 837
TWA, 1063
two-dimensional non-divergent, 765
vertical component, 853
vorticity and turbulent cascade, 748
vorticity budget
depth integrated, 946
vorticity dynamics, 789
vorticity equation, 836
hydrostatic, 851
planetary geostrophy, 853
vorticity filaments, 841
vorticity flux vector, 854
warm core eddy, 700
water mass, 445, 1181
water mass analysis, 1181
mathematics, 1188
tracers, 1210
water mass configuration space, 1183, 1218
mathematics, 1218
water mass coordinates, 1218
water mass formation, 1185, 1202
buoyancy, 1186
water mass transformation, 1185, 1195
boundary, 1202
buoyancy, 1186, 1208
dia-surface flux, 1195
interior, 1195
kinematic, 1196
kinematic method, 1198
process, 1196
process method, 1197
processes, 1187, 1205
surface, 1206
wave equation, 735
domain of influence, 59
waves
inertia-gravity, 598
inertial, 592
Kelvin, 598
Rossby, 598
weak formulation, 242, 313, 324, 379, 381
wedge product, 167, 1220
western boundary intensification, 811
western boundary layer, 777
western intensification, 813
wind driven circulation, 633
wind stress
homogeneous layer, 810
shallow water, 692
wing flow, 474
work, 222
work-energy theorem, 179, 475
working, 428
Young-Laplace formula, 400
derivation, 400
zero buoyancy layer thickness, 539
zonal channel, 778
zonal flow
absolute vorticity constraints, 775
zonal mean, 1116
zonal ridge
topographic form stress, 409
zonally re-entrant, 717