Elements of geophysical fluid mechanics

with ocean applications

Stephen M. Griffies
Program in Atmospheric and Oceanic Sciences
Princeton University
SMG@princeton.edu

Draft from August 29, 2020

Copyright ©2020 by Stephen M. Griffies
All Rights Reserved

Griffies, Stephen M., 1962-

Elements of geophysical fluid mechanics with ocean applications

This book was typeset using LaTeX.

Contents

Preface vii

List of Symbols xiii

Princeton University AOS 571 xvii

Part I. Geophysical fluid mathematics 1

1 Cartesian tensor algebra 5

2 Cartesian tensor calculus 19

3 Partial differential equations 41

4 Geometry of curves and surfaces 65

5 General tensors in brief 75

6 General tensor analysis 83

7 Orthogonal coordinates 101

8 General vertical coordinates 115

Part II. Geophysical particle mechanics 133

9 Particle kinematics and dynamics 135

10 Symmetries and conservation laws 161

Part III. Fluid kinematics 179

11 The continuum hypothesis 181

12 Fundamentals of fluid kinematics 191

13 Material fluid objects 217

14 Mass conservation 231

15 Tracer conservation 251

16 Incompressible flow 269

17 General vertical coordinate kinematics 283

Part IV. Geophysical fluid dynamics 305

18 Momentum dynamics 307

19 Stress in fluids 317

20 Pressure form stress 335

21 Buoyancy 353

22 Equilibrium thermodynamics 369

23 Energy dynamics 395

24 Filtered equations 415

25 Oceanic Boussinesq fluid 431

26 Geostrophic mechanics 459

27 Balanced inviscid horizontal flows 479

28 Ekman mechanics 493

Part V. Shallow water mechanics 505

29 Shallow water models 507

30 Shallow water dynamics 525

31 Shallow water gravity waves 553

Part VI. Vorticity and potential vorticity 571

32 Vorticity and circulation 573

33 Shallow water vorticity 583

34 Vorticity mechanics 611

35 Angular momentum, vorticity, and strain 639

36 Potential vorticity mechanics 647

37 Potential vorticity budgets and impermeability 665

38 General vertical coordinate dynamics 681

Part VII. Balanced models 691

39 Two-dimensional barotropic flow 693

40 Shallow water planetary and quasi-geostrophy 703

41 Continuously stratified planetary geostrophy 719

42 Continuously stratified quasi-geostrophy 743

Part VIII. Topics in ocean fluid mechanics 761

43 Space-time dependent gravity 763

44 Surface tension 775

45 Surface gravity waves and Stokes drift 783

46 Isopycnal ocean equations 799

47 Advection and diffusion 811

48 Tracer kinematics 847

49 Subgrid scale tracer transport 877

50 Evolution of ocean buoyancy and global sea level 899

51 Water mass and tracer mass analysis 917

Contents

Preface

Synopsis

Geophysical fluid mechanics (GFM) is an area of physics concerned with fluid motion on a rotating and gravitating spherical planet. In this book, the primary application of GFM concerns terrestrial fluid motion with particular emphasis on the needs of ocean analysis. Even so, we consider concepts and methods generally relevant to the atmosphere and extra-terrestrial planetary fluid motions. A key feature of geophysical fluid motion is that the fluid is in near solid-body motion with the planet, thus prompting a description of the motion from the rotating planetary reference frame.

We study the physics of perfect geophysical fluids, where fundamental processes are limited to the reversible and mechanical. We also study realistic geophysical fluids including those with multiple matter constituents exposed to irreversible processes such as mixing and heating. This book is in part a systematic formulation of theoretical geophysical fluid mechanics, and in part an eclectic sample of topics to support understanding and to provide avenues for research. As part of this development we draw on a variety of topics from mathematics, mechanics, thermodynamics, and tracer mechanics.

Though guided by observed phenomena, significant weight is here given to appreciating the mechanics at the core of geophysical fluid mechanics. We thus study certain topics to a degree unnecessary when primarily focused on phenomenological goals. Presentations are deductive to the extent available yet without being burdensome to the physicist. Derivations are offered in detail to expose the physics and to develop a mastery of the maths. In so doing, we cultivate an appreciation for the physical principles and mathematical structure forming the governing equations. This appreciation builds understanding and supports the development of methods for analysis and numerical simulation.

Some treatments of rotating and stratified fluids, and applications to circulating atmospheres and oceans, eschew certain of the foundational physical concepts and associated mathematical methods. Conversely, some treatments venture into the foundations and methods while assuming a specialized knowledge outside the purview of many practitioners. This perception motivated the style for this book. Namely, we study geophysical fluid mechanics from an elementary mathematical physics viewpoint at a level accessible to the entering graduate student. Doing so is partly an intellectual exercise in weaving together a variety of physical and mathematical topics. It is also a practical exercise that offers a platform for research and teaching. This perspective motivates our focus on physics forming the core of the equations rather than on physics emerging from solutions

to the equations. This emphasis has relevance to research that makes use of numerical simulations as well as process field measurements, each of which are deepened by physical frameworks and mathematical tools that support analysis. There is a synergy between the physical content of geophysical fluid mechanics and its mathematical expression, whereby physical concepts inform the maths and mathematical clarity reveals the physics. This synergy is the heart of mathematical physics and this book is a humble attempt to exemplify it.

This book is targeted to entering graduate students with an undergraduate training in physics, engineering, and/or applied mathematics. The reader is assumed to have some prior exposure to Newtonian mechanics, equilibrium thermodynamics, fluid mechanics (or another continuum field theory such as classical electromagnetism), vector calculus, tensor analysis, and partial differential equations. Even so, most of these pre-requisites are reviewed within this book where needed. Hence, students finding the entry fee rather expensive will see their patience and persistence rewarded. Those aiming to master the subject should work through the in-text derivations and pursue exercises at the end of chapters. There is no substitute for hands-on practice.

Two pillars of geophysical fluid mechanics

There are two pillars to theoretical geophysical fluid mechanics. The *formulation pillar* comprises the physical and mathematical formulation of conceptual models used to garner insight into the rich variety of phenomena realized by rotating and stratified fluids. This pillar is not generally concerned with finding specific solutions to the resulting mathematical equations. Rather, it is concerned with understanding how physical concepts are mathematically expressed in a mechanical description of geophysical flows. In the very act of formulating the equations, we expose physical principles and processes encapsulated by the equations. The particular study of geophysical fluid phenomena and corresponding solutions to equations, such as waves, instabilities, turbulence and circulation, forms the *emergent phenomena pillar*. The bulk of this book focuses on the conceptual formulation pillar, though we introduce elements of emergent phenomena where it serves our needs. This choice to focus on formulation is a product of the course from which this book grew, which is largely a course that lays the foundations for the subject by formulating the equations and exploring their physical content, rather than studying specific solutions to the equations.

A book focused on formulating the equations of geophysical fluid mechanics may seem dry and uninteresting. Nevertheless, many, such as this author, find physical principles and their mathematical realization to offer a deeply compelling aspect of theoretical science. Therefore, rather than avoiding principles and eschewing mathematics, we embrace them. Besides having an aesthetic value, the formulation pillar has an objective payoff by refining understanding of emergent phenomena. Indeed, the formulation and emergent phenomena pillars are synergistic and they offer lessons and guidance for the other. Accepting the proposition that the formulation pillar is under-emphasized by existing books on geophysical fluid mechanics, we here aim to partly remedy that situation while remaining accessible and relevant to students of oceanography, atmospheric sciences, and associated fields in physics, engineering, and applied mathematics.

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 in physical oceanography. Although this book is quite 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. The following surveys the parts to this book.

- GEOPHYSICAL FLUID MATHEMATICS: Part I provides a suite of mathematics chapters of use in studying geophysical fluids. 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 firm understanding.

- GEOPHYSICAL PARTICLE MECHANICS: Part II provides a resume 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.

- FLUID KINEMATICS: Mechanics is comprised of kinematics (the study of intrinsic properties of motion) and dynamics (the study of causes for motion). In Part III we initiate a study of fluid mechanics by focusing on fluid kinematics. Our treatment is rather thorough, exposing the reader to both the Eulerian and Lagrangian viewpoints and emphasizing the variety of kinematic notions and tools key to describing fluid motion. It is in these chapters that we learn how to think about fluid motion. As such motion has many features fundamentally distinct from point particle and rigid body motion, we require distinct tools and concepts that take practice to intellectually digest.

  Quite often a course in fluid mechanics skims over fluid kinematics, moving quickly to "more interesting topics" related to dynamics. Indeed, some kinematic topics can seem a bit esoteric, particularly the study of Lagrangian kinematics. However, an incomplete understanding of fluid kinematics can lead to many difficulties with later subjects in dynamics. The reader is thus encouraged to study kinematics in detail, and to revisit the material as the need arises during later chapters.

- GEOPHYSICAL FLUID DYNAMICS: Part IV presents the core of geophysical fluid dynamics. Within these chapters we study how Newton's Laws and thermodynamics on a rotating and gravitating sphere, along with kinematics, leads to the basic features of fluid dynamics of the atmosphere and ocean. We give much attention to how forces acting on fluid elements lead to accelerations and thus to motion. These forces act throughout the volume of a fluid element (gravity, Coriolis, centrifugal) as well as on the boundary of a fluid element (pressure, 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 occurs 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 study of shallow water models a very popular activity for 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. 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 kinematics (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 from motion on a rotating and gravitating sphere. These constraints provide conceptual insights and predictive power.

- balanced models: 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 motions associated with geostrophy. 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, with the subject worthy of a book of this size.

- topics in ocean fluid mechanics: Part VIII is where the author's prejudices are exposed, in that we here encounter a variety of topics in ocean fluid mechanics from the small scale (e.g., surface tension) to the large scale (water mass analysis). Some chapters offer the reader a summary of topics not normally found in a GFM book, whereas others aim to bring the reader to the cutting edge of ongoing research.

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 prompt attention to be spread rather than focused. 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, generous use of cross-referencing identifies allied material treated elsewhere in the book.

No book is an island, with this book generously making use of other books, review articles, research articles, and online tutorials. Furthermore, 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 a broad study of the field, throughout this book we provide pointers to written and/or video presentations to support further studies. Many more resources are available through a quick internet search.

From course notes to book

This book grew from class notes developed for a first semester graduate course in geophysical fluid mechanics at Princeton University. The course offers an overview of fluid mechanics and the formulation of a variety of conceptual geophysical fluid models. The course drew on some tremendous resources, such as *Vallis* (2017) and its more succinct offspring *Vallis* (2019), as well as the more ocean focused texts from *Cushman-Roisin and Beckers* (2011) and *Olbers et al.* (2012). This course also drew from the pioneering texts by *Gill* (1982) and *Pedlosky* (1987), both of which

served a generation of students studying *geophysical fluid dynamics* (GFD). Their treatments of GFD offer an elegant application of Newton's laws of motion to rotating and stratified fluids, with an emphasis on describing large-scale circulation features through the use of applied mathematical methods.

The study of rotating and stratified fluids has evolved much over its history. During recent decades we have 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. What has emerged is a recognition that a fruitful study of rotating and stratified fluids makes use of ideas that go beyond the traditional notions of GFD. A contemporary practitioner thus develops insights by weaving together concepts and tools from a variety mathematical, kinematical, dynamical, thermodynamical, tracer mechanical, numerical, experimental, and observational perspectives. Acknowledging this broadening of the practice motivates the term "mechanics" in this book's title, rather than the more focused "dynamics". It is a trivial change in verbiage but a nontrivial change in perspective.

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*.

Elements of geophysical fluid mechanics are essential as numerical circulation models, now the core tool for Climate Science 2.0, admit enhanced details of the complex multi-scaled fluid flow. Indeed, GFM is essential for the development of robust numerical models along with mechanistic interpretations of model output, which in turn further supports confidence in model projections for future climate. Additionally, GFM is key as observational field measurements become more refined and process-oriented, with advanced technologies facilitating increasingly ambitious investigations into geophysical fluid mechanical questions within the natural environment. As technologies are enhanced for both simulations and measurements, the concepts and tools from geophysical fluid mechanics become more relevant as a means to conceptually digest the underlying physics and to communicate the scientific findings to the broader community. Geophysical fluid mechanics, as part of the scientific investigation of the earth's climate, thus serves an essential role in society's quest for sustainability, equity, and justice.

Gratitudes

This book benefitted from interactions with the many bright and inquisitive students taking part in Princeton University's AOS 571. Prompted by their comments, queries and sincere search for understanding, I was led to introduce and to refine a variety of topics. Further inspiration was offered by students, postdocs, and fellow research scientists who I encountered as part of my nearly 30 years of research. I wish to particularly highlight a journal club held as part of my research group, where a great deal of head scratching led to insights sprinkled through these pages.

I remain amazed and profoundly grateful for being part of the unique research environment cultivated at the 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 many thinkers whose style, questions, and insights have taken root into my work. This environment has also afforded me the opportunity to travel the world to interact with various gurus of the field. Their 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, have been among the most satisfying experiences of my career. Safe spaces for honest, diverse, and inclusive learning are precious, and I am grateful to those in my life who have nurtured that space.

A book of this nature is a not a simple endeavor. It starts modestly, grows a bit over time, and eventually becomes an obsession. There is little objective guidance to know when the book is ready to live outside the author's hands. That decision ultimately must come from intuition, exhaustion, or some combination. 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. At each step of this seemingly neverending project, I have been supported by my wife, Adi, and our son, Francisco. I am deeply grateful for their patience and trust as I satisfied the needs of writing this book through countless nights, weekends, and holidays secluded in our home attic office. I treasure being part of this family and I dedicate this work to you two amazing human beings.

caveats and limitations

Although reaching some level of maturity, this book remains a work in progress that is not 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 also needed to unify notation, complete the index, develop more exercises, enhance figures, improve discussions, and correct errors.
- Many sections and chapters have yet to be scrutinized by others. I hope to garner readers interested in providing candid feedback to help 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.

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.

English-based symbols

SYMBOL	MEANING
$A_{\rm v}$	Avogadro's number: $A_{\rm v} = 6.022 \times 10^{23}$ mole^{-1}
b	buoyancy of a fluid element, with $b > 0$ for fluid relatively light: $b = -g\,(\rho - \rho_0)/\rho_0$
C	tracer concentration = mass of tracer per mass of fluid = tracer mass fraction
$C_{\rm D}$	dimensionless bottom drag coefficient: $C_{\rm D} > 0$
$\mathcal{C}$	circulation of velocity around the boundary of a surface $\mathcal{C} \equiv \oint_{\partial \mathcal{S}} \boldsymbol{v} \cdot \mathrm{d}\boldsymbol{r}$
f, f_0	Coriolis parameter and a reference value, also the planetary vorticity: $f = 2\,\Omega\,\sin\phi$
$\boldsymbol{F}$	frictional acceleration
F_i^m	components to the transformation matrix between material & position coordinates; also the deformation tensor
G	Newton's gravitational constant: $G = 6.674 \times 10^{-11}$ N m^2 kg^{-2} = 6.674×10^{-11} m^3 kg^{-1} s^{-2}
g	effective gravitational acceleration at earth's surface from central gravity + planetary centrifugal: $g \approx 9.8$ m s^{-2}
h	layer thickness for a shallow water fluid
h	layer thickness for a continuously stratified fluid: $h = \overline{\hbar}\,\delta\sigma$
$\hbar$	specific thickness for a generalized vertical coordinate: $\hbar = \partial z/\partial\sigma$
H	vertical depth scale
H	sometimes used as depth of the ocean bottom: $z = -H(x,y) = \eta_{\rm b}(x,y)$
i	i $= \sqrt{-1}$ used for imaginary numbers
$\boldsymbol{J}$	tracer flux; for material tracers the dimensions are mass per time per area
K	kinetic energy for a point particle of mass m: $K = m\,\boldsymbol{V}\cdot\boldsymbol{V}/2$
$\mathcal{K}$	kinetic energy per mass of a fluid element: $\mathcal{K} = \boldsymbol{v}\cdot\boldsymbol{v}/2$
L	horizontal length scale
p	pressure at a point in the fluid
Q	potential vorticity either for a continuous fluid (Ertel PV) or shallow water fluid (Rossby PV)
$Q_{\rm m}$	mass flux per horizontal area crossing the ocean upper surface, with $Q_{\rm m} > 0$ for mass entering ocean
r	radial position relative to an origin typically taken as center of a spherical earth
R	radius of sphere
$R_{\rm e}$	mean earth radius: $R_{\rm e} = 6.367 \times 10^6$ m
$R_{\rm g}$	universal gas constant: $R_{\rm g} = 8.314$ J mole^{-1} K^{-1} = 8.314 kg m^2 s^{-2} mole^{-1} K^{-1}
Re	Reynolds number: $Re = U\,L/\nu$
Ro	Rossby number: $Ro = U/(f\,L)$
S	salinity = mass of salt in a fluid element per mass of seawater
s	expression for a generic surface:$s = s(x,y,z,t)$.
T	thermodynamic or *in situ* temperature
$\mathbb{T}$	stress tensor
t	time (universal Newtonian time)
$\boldsymbol{u}$	horizontal velocity of a fluid element: $\boldsymbol{u} = \hat{\boldsymbol{x}}\,u + \hat{\boldsymbol{y}}\,v$
U	horizontal velocity scale
$\boldsymbol{U}$	depth integrated horizontal velocity: $\boldsymbol{U} = \int_{\eta_{\rm b}}^{\eta} \boldsymbol{u}\,\mathrm{d}z$
$\boldsymbol{v}$	velocity of a fluid element: $\boldsymbol{v} = \mathrm{D}\boldsymbol{x}/\mathrm{D}t = \hat{\boldsymbol{x}}\,u + \hat{\boldsymbol{y}}\,v + \hat{\boldsymbol{z}}\,w$
W	vertical velocity scale
w	vertical component to the velocity: $w = \mathrm{D}z/\mathrm{D}t$
$w^{(\dot\sigma)}$	dia-surface velocity = fluid volume per *horizontal area* per time crossing a σ surface: $w^{(\dot\sigma)} = (\partial z/\partial\sigma)\,\mathrm{D}\sigma/\mathrm{D}t$
$w^{\rm dia}$	dia-surface flux = fluid volume per *surface area* per time crossing a σ-surface: $w^{\rm dia} = (1/\lvert\nabla\sigma\rvert)\,\mathrm{D}\sigma/\mathrm{D}t$
(x,y,z)	triplet of Cartesian coordinates
$\boldsymbol{x}$	spatial position with Cartesian representation: $\boldsymbol{x} = \hat{\boldsymbol{x}}\,x + \hat{\boldsymbol{y}}\,y + \hat{\boldsymbol{z}}\,z$
$(\hat{\boldsymbol{x}}, \hat{\boldsymbol{y}}, \hat{\boldsymbol{z}})$	triplet of Cartesian unit vectors oriented in a righthand sense
z_σ	specific thickness for a generalized vertical coordinate: $z_\sigma = \partial z/\partial\sigma = \hbar$

Greek symbols

SYMBOL	MEANING
α	thermal expansion coefficient: $\alpha = -\rho^{-1}\,\partial\rho/\partial\theta$
β	saline contraction coefficient: $\beta = \rho^{-1}\,\partial\rho/\partial S$
β	meridional derivative of planetary vorticity: $\beta = \partial_y f$
$\eta = \eta_0$	vertical position of the free upper surface of a fluid domain: $z = \eta(x, y, t)$
η_k	interface depth for layer-k in a stacked shallow water fluid
$\eta_{\mathrm{b}} = -H$	vertical position of rigid and static lower solid-earth boundary of a fluid domain: $z = \eta_{\mathrm{b}}(x, y) = -H(x, y)$
(λ, ϕ)	longitude and latitude position on the sphere: $0 \le \lambda \le 2\pi$ and $-\pi/2 \le \phi \le \phi/2$
κ	molecular kinematic diffusivity
κ^{eddy}	eddy kinematic diffusivity: $\kappa^{\mathrm{eddy}} \gg \kappa$
ν	molecular kinematic viscosity
ν_{air}	molecular kinematic viscosity of air: $\nu_{\mathrm{air}} \approx 1.3 \times 10^{-5}\ \mathrm{m}^2\ \mathrm{s}^{-1}$
ν_{water}	molecular kinematic viscosity of fresh water: $\nu_{\mathrm{water}} \approx 10^{-6}\ \mathrm{m}^2\ \mathrm{s}^{-1}$
ν^{eddy}	eddy viscosity: $\nu^{\mathrm{eddy}} \gg \nu$
$\boldsymbol{\omega}$	relative vorticity vector: $\boldsymbol{\omega} = \nabla \wedge \boldsymbol{v}$
$\boldsymbol{\Omega}$	earth's angular velocity oriented through north pole with magnitude $\Omega = 7.2921 \times 10^{-5}\ \mathrm{s}^{-1}$
Φ	geopotential, which in its simplest form is $\Phi = g\,z$
ψ	streamfunction for two-dimensional incompressible flow: $\nabla \cdot \boldsymbol{u} = 0 \Longrightarrow \boldsymbol{u} = \hat{\boldsymbol{z}} \wedge \nabla\psi$
$\boldsymbol{\Psi}$	vector streamfunction for three-dimensional incompressible flow: $\nabla \cdot \boldsymbol{v} = 0 \Longrightarrow \boldsymbol{v} = \nabla \wedge \boldsymbol{\Psi}$
ρ	*in situ* 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 $\sigma(x, y, z, t) =$ constant, commonly used for generalized vertical coordinates
$\boldsymbol{\tau}$	stress vector
θ	potential temperature or 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 $\zeta_{\mathrm{a}} = f + \partial_x v - \partial_y u$

Mathematical operations

SYMBOL	MEANING
∇	gradient operator that acts on a scalar
∇_z	horizontal gradient operator: $\nabla_z = \hat{\boldsymbol{x}}\,(\partial/\partial x) + \hat{\boldsymbol{y}}\,(\partial/\partial y) = \hat{\boldsymbol{x}}\,\partial_x + \hat{\boldsymbol{y}}\,\partial_y$
$\nabla\cdot$	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: $\nabla_\sigma = \hat{\boldsymbol{x}}\,(\partial/\partial x)_\sigma + \hat{\boldsymbol{y}}\,(\partial/\partial y)_\sigma$
∂_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, $\boldsymbol{x}$
$[\partial/\partial t]_\sigma$	time derivative computed on constant σ-surface
$\mathrm{D}/\mathrm{D}t$	material, Lagrangian, or substantial time derivative following a fluid element
δ	differential increment often written for a field following the fluid motion
Δ	finite difference increment in space: $\Delta_x, \Delta_y, \Delta_z, \Delta_\sigma$
$\mathrm{d}A$	infinitesimal horizontal area element: $\mathrm{d}A = \mathrm{d}x\,\mathrm{d}y$
$\mathrm{d}\mathcal{S}$	infinitesimal area element on a surface
$\mathrm{d}V$	infinitesimal volume element
δV	infinitesimal volume for a region moving with the fluid (Lagrangian region)
$\int_{\mathcal{R}} \mathrm{d}V$	integral over an arbitrary volume region, $\mathcal{R}$
$\int_{\mathcal{R}(\boldsymbol{v})} \mathrm{d}V$	integral over a region that follows the fluid motion (Lagrangian integral)
$\int_{\mathcal{S}} \mathrm{d}\mathcal{S}$	integral over a surface $\mathcal{S}$
$\oint_{\partial\mathcal{R}} \mathrm{d}\mathcal{S}$	integral over a closed surface $\partial\mathcal{R}$ that bounds the volume $\mathcal{R}$
$\oint \mathrm{d}l$	integral over a periodic domain
$\oint_{\partial\mathcal{S}} \mathrm{d}l$	counter-clockwise closed line integral over the boundary of a surface, $\partial\mathcal{S}$
$\sim$	similar to or scales as
$\approx$	approximately equal to
$\dot{A}$	short-hand notation for material time derivative $\dot{A} = \mathrm{D}A/\mathrm{D}t$.

Princeton University AOS 571

0.1 Welcome . . . **xvii**
0.2 Class structure and expectations . . . **xviii**
0.2.1 Cultivating a safe space for healthy learning . . . xviii
0.2.2 Class notes . . . xix
0.2.3 Grade = homeworks (75%) + final exam (25%) . . . xix
0.2.4 During the class . . . xix
0.3 Written and spoken communication . . . **xx**
0.3.1 Clear thinking leads to clear communication . . . xx
0.3.2 Empathy is key . . . xx
0.3.3 Clarity helps, but some material is just tough . . . xx
0.4 Pointers on problem solving . . . **xxi**
0.4.1 Dimensional analysis . . . xxi
0.4.2 Tensorial consistency . . . xxi
0.4.3 Words and pictures . . . xxi
0.4.4 Mathematical sophistication . . . xxii
0.4.5 More than one path to a solution . . . xxii
0.4.6 Balance between thorough and brief . . . xxii
0.4.7 Questions for clarification . . . xxii
0.4.8 Stay positive . . . xxiii
0.5 Specifics for marking assignments . . . **xxiii**
0.6 Course syllabus . . . **xxiv**

0.1 Welcome

Welcome to Princeton University's AOS 571 for the autumn 2020 semester. Due to the ongoing COVID-19 pandemic, the class will be run this year fully online via Zoom. Please reach out to me if an online class causes you difficulties due to technology issues or the like. Although this situation places us at a bit of a disadvantage due to the lack of in-person time, I hope that there will be advantages that outweigh the limitations. For example, classes will be recorded so that you can access them later to fill in gaps.

The goal of AOS 571 is to develop certain of the foundations for geophysical fluid mechanics, which is a physics discipline concerned with the motion of rotating and stratified fluids such as those in the earth's atmosphere (mostly a gas) and the ocean (mostly a liquid). This subject is rich in physics, maths, and phenomenology. Our explorations will reveal basic features about the observed patterns of planetary fluid motion, and provide a mathematical physics foundation for

later courses and research. Here are details of your teacher and the course times.

Dr. Stephen M. Griffies
SMG@princeton.edu
https://stephengriffies.github.io/
Office hours by email appointment
Class Monday and Wednesday 2:30pm-4:00pm + 30 minutes on most days for "precept" discussions
Virtual class on Zoom and organized by Blackboard
Classes are recorded with video availability to be determined.

0.2 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 will work though pieces of these notes in class. Garnering an understanding of geophysical fluid mechanics requires great effort and much practice. 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.

0.2.1 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, which is a particularly important task during the pandemic. Indeed, even in non-pandemic times, the first semester of graduate school can be very tough. Even so, it can be a time for building deep friendships as we share in the process of learning how to thrive as budding researchers and teachers.

0.2.2 Class notes

The class lectures closely follow selected material from the Griffies (2020) course notes available online. These notes started in 2014 as annotations to the textbooks from *Vallis* (2017) and *Vallis* (2019). Both of these books remain highly recommended references for the course. However, the class notes have grown over recent years into an autonomous document from which the course material directly derives.

Material for each class comes from the class notes as well as online videos. For autumn 2020 we will meet virtually via Zoom. However, the structure of the course will remain largely as in-person, with interactions and questions encouraged throughout the lectures. You are asked to read through the class notes and to view the 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. In this way, the class is partially "flipped".

0.2.3 Grade = homeworks (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 (if within time zone) and participate by asking questions. Note that the class will be recorded and available for later viewing.

Worked homework exercises = 75%

Worked homework exercises are due at the start of class 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.

Take-home final exam = 25%

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 references provided to the student.

0.2.4 During the 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. This request is particularly important given that everyone is within their own bubble during the virtual classes. To get the most from the lectures and class time requires focused attention and participation.

0.3 Written and spoken communication

Reading class 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 pill. Others rush through more material than even they can digest, whereas others are happy to pedagogically engage in a lecture even without any prior knowledge of the material. To succeed in research and teaching, you will need to master elements of both written and spoken communication. It is therefore critical to nurture these skills throughout your career whether you seek employment in academics or elsewhere. Furthermore, I am grateful for your feedback on the lectures in case you find the style unsupportive of your pedagogy.

0.3.1 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 and organize thoughts and, if time allows, to edit and edit yet again. 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?

0.3.2 Empathy is key

A basic tenet of effective communication is empathy. 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. You will not need this document for homeworks or exams in AOS 571. However, it will come in handy when you begin the process of writing scientific documents and preparing scientific presentations.

0.3.3 Clarity helps, but some material is just tough

Although poor communication certainly does hinder our ability to digest new ideas and material, it is conversely 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 before some material can be sufficiently digested to allow for the core conceptual nugget to be revealed. So do

ask for clear communication, but do not expect clarity to be sufficient to remove the struggles we all experience when learning.

0.4 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 one given for a class or one forming part of a broader research question.

0.4.1 Dimensional analysis

The symbols we use in mathematical physics correspond to geometrical objects (e.g., points, vectors, tensors) describing a physical concept (e.g., location, 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

$$A = B, \tag{0.1}$$

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 important means to "debug" the maths.

0.4.2 Tensorial consistency

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

$$A = B, \tag{0.2}$$

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

$$\boldsymbol{A} = \boldsymbol{B} \tag{0.3}$$

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

$$\boldsymbol{A} = \nabla \cdot \boldsymbol{B} \tag{0.4}$$

does not make sense because the left hand side is a vector and the right hand side is a scalar. Maintaining basic tensorial rules can be considered the next level of sophistication above dimensional analysis.

0.4.3 Words and pictures

It is important to explain the problem and your solution using words and pictures. Hence, it is good practice to liberally include words/sentences in between the key equations, with the purpose to explain what the maths means using 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 and errors in the solution. 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.

0.4.4 Mathematical sophistication

Students enter a graduate level course in geophysical fluid mechanics with differing levels of maths training. For the more math-centric problems, it may prove useful to exercise your maths brain muscle. Do so if you feel it useful for the solution presentation. However, solutions will generally not be marked down if you fail to present the full depth of the maths available for a problem. Having said that, as this course progresses you are expected to evolve in your math skills, just as your understanding of the physics matures.

0.4.5 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.

0.4.6 Balance between 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 or in *Vallis* (2017). Proper referencing of the equation is all that you need; i.e., tell me something like "starting from equation (X.YY) from the 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.

0.4.7 Questions for 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 BlackBoard 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.

0.4.8 Stay positive

Everyone makes mistakes, some more than others. The toughest part about making mistakes is the self-imposed shame or embarrassment. Please try to keep a positive mind about your mistakes. As you will learn, mistakes offer significant opportunities for learning. I am a poster-child for this process!

So do not fret if you find many marks on your homeworks and exams. But do be sure 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.

0.5 Specifics for marking assignments

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 down.

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 nonzero credit.

Here are some further considerations for your homework assignments.

- PRESENTATION OF THE SOLUTION: Please write clearly and legibly. There is generally no need to submit solutions in LaTeX. But if your handwriting is horrible, then consider learning LaTeX to typeset the solutions. 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 accidents the only excuses for late assignments. Fairness is the fundamental reason to insist on this rule, particularly since we will generally work through the solutions on the day the homeworks are handed in.

- You may hand in assignments via email, but within the same deadline as for class hand-in.
- You may hand-in homework late but only with prior arrangement. Otherwise, the homework set will receive zero credit.

0.6 Course syllabus

The following is a syllabus for the course. The course material consists of the class notes plus a selection of online videos that are required viewing. Ideally, the notes are to be read and the videos viewed *prior to the class*. Salient points related to the material will be discussed during class, along with further discussion and questions. That is, we will *not* cover all material in class that is expected to be read as part of the reading assignments.

1. Course introduction and mathematics refresher
   - 9-minute video from Prof. N. Hall
   - 16-minute video from 3Blue1Brown on the divergence and curl operations, with examples taken from fluid mechanics.
   - Course introduction
   - Chapter 11: the continuum hypothesis used for describing fluids as a continuous media
   - Survey of maths from Part I
   - Chapter 1: cartesian tensors
   - Chapter 2: vector calculus
2. Geophysical particle mechanics
   - Chapter 9: 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 9: Newton's equation of motion for particle moving around rotating sphere, rotating reference frames, gravitational geopotential
3. Symmetries, conservation laws, and constrained motion
   - Chapter 10: 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 12: Fluid kinematics, Eulerian and Lagrangian descriptions, Galilean invariance, material time derivative, flow lines
5. Mass and tracer conservation
   - 5-minute video on mass conservation from Prof. Hogg.

- Chapter 14: continuity equation, mass budget for fluid elements and finite regions, kinematic boundary conditions
- Chapter 15: barycentric velocity, tracer equation, budgets for infinitesimal fluid elements, budgets for finite fluid regions, Leibniz-Reynolds transport theorem, boundary conditions

6. Kinematics of incompressible flow

   - 4-minute video on streamlines from Prof. Hogg.
   - Chapter 16: scalar streamfunction, vector streamfunction, area and volume conservation, meridional-depth overturning circulation

7. 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

8. 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

9. Buoyancy

   - Chapter 21: Archimedes' principle, buoyancy, stratification, gravitational stability, mass density for perfect and realistic fluids

10. Thermodynamics

    - Section 2.8: refresher on exact and inexact differentials
    - Chapter 22: First law of thermodynamics, thermodynamic potentials, ideal gas atmosphere

11. Energy dynamics and filtered equations

    - Chapter 23: thermodynamics of a moving fluid, mechanical energy, internal energy, total energy
    - 8-minute video on hydrostatic pressure from Prof. Hogg.
    - Chapter 24: primitive equations, hydrostatic approximation, tangent plane approximation, basics of time evolving a fluid state
    - Chapter 25: oceanic Boussinesq approximation, equations for an ocean model

12. Geostrophy and thermal wind

    - 26-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

13. Balanced inviscid horizontal flows

    - 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. N. Hall.
    - Chapter 29: thickness equation, momentum equation, reduced gravity model, stacked shallow water layers, shallow water layer in a rotating tank

16. Shallow water dynamics

    - Chapter 30: geostrophy, form stress, mechanical energy including available potential energy, angular momentum in a tank

17. Shallow water gravity waves and geostrophic adjustment

    - 14-minute video on gravity waves from Prof. N. Hall.
    - Chapter 31: gravity waves in single shallow water layer; geostrophic adjustment

18. Vorticity, circulation, and potential vorticity

    - 23-minute video on vorticity from Prof. Shapiro
    - Chapter 32: vorticity and circulation
    - Chapter 33: shallow water vorticity and potential vorticity

19. Vorticity mechanics

    - 21-minute video on vorticity from Prof. Shapiro.
    - 5-minute video on vortex rings and Helmholz's theorems from the Physics Girl.
    - Chapter 34: vortex lines and tubes, Kelvin's circulation theorem, mechanics of baroclinicity, β-effect

20. Potential vorticity mechanics

    - Chapter 36: PV material invariance for perfect fluid; PV evolution with friction and heating; impermeability theorem; isopycnal layer integrated PV

21. Balanced models I (single shallow water layer)

- Chapter 39: barotropic vorticity equation
- Chapter 40: Buckingham's Π theorem, asymptotic expansion in terms of small Rossby number, shallow water planetary geostrophy and quasi-geostrophy

22. 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.

23. Tidy up loose ends

Part I

Geophysical fluid mathematics

Fluid mechanics is a classical field theory based on Newton's laws of mechanics and classical thermodynamics, both applied to a continuous fluid media. Geophysical fluid mechanics (GFM) is concerned with buoyancy stratified fluids of multiple constituents moving on a rotating sphere. Rotation, stratification, multiple constituents, and spherical geometry each influence the maths encountered in GFM.

Our goal for this part of the book is to review mathematical topics with a focus on how they are useful for GFM. Depending on the background of the reader, this material can be readily skipped or accessed later if needed to fill gaps. Indeed, one power of the tensor analysis presented in many of these chapters is that if handled properly, much of the maths takes on the same appearance as when using familiar Cartesian coordinates. Nonetheless, at important points there is use in having skills with the mathematical machinery. These chapters introduce that machinery.

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, they are geometric objects such as points, vectors, surfaces, volumes, etc. In the study of geophysical fluid mechanics, we use physical principles to develop equations relating geometric objects. The tools of tensor analysis are then used to compute numbers as required to compare with experiments and field measurements, and to formulate discrete equations for numerical simulations.

The above perspective of "physics as geometry" is foundational to theoretical physics (e.g., *Thorne and Blandford* (2017)) and it has 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. One aim for this book is to develop mathematical tools to help unpack the physics encapsulated by the equations. This aim extends to those cases where analytical solutions are unavailable, which is the norm for nonlinear field theories such as fluid mechanics. 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 our study. Tensor analysis is the proven means for systematically performing such transformations, hence motivating its use in fluid mechanics.

The following is 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 and burdensome manipulations that can obfuscate the underlying physical concepts. Indeed, imagine the tedium required to write 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, and its generalization to tensor analysis, 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 descriptions.

- SPHERICAL PLANET: Geophysical fluids move on a spherical planet, making spherical coordinates the preferred choice for studying and modeling 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 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 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. Transforming between a Cartesian and such *generalized vertical coordinates* requires the mathematical precision of general tensors.

SUMMARY OF THE MATH CHAPTERS

Some of the chapters in this part of the book are essential for nearly all subsequent chapters, whereas many others target the aficionados and 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 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.

- GEOMETRY OF CURVES AND SURFACES: Chapter 4 introduces rudimentary differential geometry used to describe properties of curves (such as fluid particle trajectories) and surfaces (such as isopycnals).

- GENERAL TENSOR INTRODUCTION: Chapter 5 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 6.

- GENERAL TENSOR ANALYASIS: Chapter 6 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 and isopycnal coordinates.

- ORTHOGONAL COORDINATES: Chapter 7 offers a reference for various locally orthogonal coordinate systems (Cartesian, spherical, cylindrical) used in this book and how various mathematical objects appear when written in these coordinates.

- GENERALIZED VERTICAL COORDINATES: Chapter 8 offers a reference for the mathematics of generalized vertical coordinates. These non-orthogonal coordinates, such as isopycnal coordinates, are commonly used for conceptual and numerical models of stratified flows. Their non-orthogonality presents some complexity in both concept and detail.

1

Cartesian tensor algebra

In this chapter we introduce the formalism of Cartesian tensor analysis, focusing on the basic algebraic relations. The use of Cartesian tensors restricts attention to Cartesian coordinates and their orthogonal transformations via 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 **7**
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 Orthogonality relations between cross products 10
1.4.3 Vector product of arbitrary vectors 10
1.4.4 Geometric interpretation of the vector product 11
1.4.5 Generalization to arbitrary vectors 12
1.5 Measuring volume **12**
1.5.1 Volume defined by three vectors 12
1.5.2 Cartesian volume element for integration 13
1.5.3 $n-$space volumes and the Levi-Civita tensor 13
1.6 Example vector identities using the Levi-Civita tensor **14**
1.6.1 Double vector product 14
1.6.2 Scalar product of two vector products 14
1.7 Transforming the coordinate representations **14**
1.7.1 Inverse transformation 15
1.7.2 Orthogonal transformation 15
1.7.3 Geometric interpretation of orthogonal transformations 16
1.7.4 Transforming the coordinate representation of a vector 16
1.7.5 Invariance of the scalar product 17
1.7.6 Transforming the coordinate representation of a second order tensor 17

1.7.7 Importance of distinguishing between tensors and matrices 17
1.8 Exercises **18**

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 at a point provides a single number so that a scalar field provides a number at each point in space-time. Example scalar fields include temperature, mass density, entropy, salinity, humidity, and mechanical energy. A vector connects two points and is specified by a direction and a magnitude, with a vector field providing a vector at each point in space-time. 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-time. 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-time 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 gives us nearly all of the formalism necessary to study general tensors in geophysical fluid mechanics.

1.2 Points and vectors

Consider a point, $\mathscr{P}$, 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 show in Figure 1.1. We write this coordinate representation as

$$\mathscr{P} \mapsto \vec{P} = \hat{\boldsymbol{x}}\, P_1 + \hat{\boldsymbol{y}}\, P_2 + \hat{\boldsymbol{z}}\, P_3. \tag{1.1}$$

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_2), that measure distance along their corresponding Cartesian unit vectors, $(\hat{\boldsymbol{x}}, \hat{\boldsymbol{y}}, \hat{\boldsymbol{z}})$. 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.

We find it useful to make use of alternative notations in which the position vector is written

$$\vec{P} = \boldsymbol{P} \tag{1.2a}$$
$$= \hat{\boldsymbol{x}}\, P_1 + \hat{\boldsymbol{y}}\, P_2 + \hat{\boldsymbol{z}}\, P_3 \tag{1.2b}$$
$$= \sum_{a=1}^{3} \vec{e}_a\, P_a \tag{1.2c}$$
$$= \vec{e}_a\, P_a. \tag{1.2d}$$

[1] The unit vectors are sometimes denoted $(\hat{\boldsymbol{\imath}}, \hat{\boldsymbol{\jmath}}, \hat{\boldsymbol{k}})$ in the literature. We generally avoid that notation in this book.

Figure 1.1: *An arbitrary point in space, $\mathscr{P}$, 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 $\mathscr{P} \mapsto \vec{P} = \boldsymbol{P} = \hat{\boldsymbol{x}}\, P_1 + \hat{\boldsymbol{y}}\, P_2 + \hat{\boldsymbol{z}}\, P_3$, with the Cartesian basis vectors the normalized triplet of unit vectors $(\hat{\boldsymbol{x}}, \hat{\boldsymbol{y}}, \hat{\boldsymbol{z}})$ and the triplet of Cartesian coordintates 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.*

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 5, we still maintain the boldface in this book given its common usage in the literature. Equation (1.2c) introduced a generic notation for the basis vectors

$$\vec{e}_1 = \hat{\boldsymbol{x}} \qquad \vec{e}_2 = \hat{\boldsymbol{y}} \qquad \vec{e}_3 = \hat{\boldsymbol{z}}. \tag{1.3}$$

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.

We emphasize that the tensor labels denote components of coordinates, P_a, and members from the set of basis vectors, $\vec{e}_a$. These labels are not to be confused with partial derivative operations.[2] We sometimes write the vector components and basis vectors in the form of a list or triplet

$$P_a = (P_1, P_2, P_3) \qquad \text{and} \qquad \vec{e}_a = (\vec{e}_1, \vec{e}_2, \vec{e}_3) = (\hat{\boldsymbol{x}}, \hat{\boldsymbol{y}}, \hat{\boldsymbol{z}}). \tag{1.4}$$

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 unit 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). Note that for the general tensors of Chapter 5, the most convenient basis vectors are not necessarily normalized.

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.

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

1.3.1 Distance between points

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

$$\mathscr{P} \mapsto \vec{P} = \vec{e}_a\, P_a \tag{1.5a}$$

$$\mathscr{P} + \mathrm{d}\mathscr{P} \mapsto \vec{P} + \mathrm{d}\vec{x} = \vec{e}_a\,(P_a + \mathrm{d}x_a). \tag{1.5b}$$

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

$$\begin{aligned}
[\mathrm{distance}(\mathscr{P}, \mathscr{P}+\mathrm{d}\mathscr{P})]^2 &= (\vec{P} + \mathrm{d}\vec{x} - \vec{P}) \cdot (\vec{P} + \mathrm{d}\vec{x} - \vec{P}) && (1.6a)\\
&= (\boldsymbol{P} + \mathrm{d}\boldsymbol{x} - \boldsymbol{P}) \cdot (\boldsymbol{P} + \mathrm{d}\boldsymbol{x} - \boldsymbol{P}) && (1.6b)\\
&= \mathrm{d}x_a\, \mathrm{d}x_b\, (\vec{e}_a \cdot \vec{e}_b) && (1.6c)\\
&= \mathrm{d}x_a\, \mathrm{d}x_b\, \delta_{ab} && (1.6d)\\
&= \mathrm{d}x_a\, \mathrm{d}x_a && (1.6e)\\
&= (\mathrm{d}x_1)^2 + (\mathrm{d}x_2)^2 + (\mathrm{d}x_3)^2. && (1.6f)
\end{aligned}$$

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

$$\vec{e}_a \cdot \vec{e}_b = \delta_{ab} = \begin{bmatrix} \delta_{11} & \delta_{12} & \delta_{13} \\ \delta_{21} & \delta_{22} & \delta_{23} \\ \delta_{31} & \delta_{32} & \delta_{33} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}. \tag{1.7}$$

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 6.1 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

Figure 1.2: *Illustrating the geometry associated with forming the scalar product between two vectors,* $\boldsymbol{P} \cdot \boldsymbol{Q} = |\boldsymbol{P}|\,|\boldsymbol{Q}|\,\cos\vartheta$.

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

$$|\boldsymbol{P}|^2 = \boldsymbol{P} \cdot \boldsymbol{P} = P_a\, P_b\, (\vec{e}_a \cdot \vec{e}_b) = P_a\, P_a = (P_1)^2 + (P_2)^2 + (P_3)^2. \tag{1.8}$$

Correspondingly, we have the scalar (or dot) product between two arbitrary vectors

$$\boldsymbol{P} \cdot \boldsymbol{Q} = P_a \, Q_b \, (\vec{e}_a \cdot \vec{e}_b) = P_a \, Q_a. \tag{1.9}$$

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

$$\cos\vartheta \equiv \frac{\boldsymbol{P} \cdot \boldsymbol{Q}}{|\boldsymbol{P}| \, |\boldsymbol{Q}|} = \frac{P_a \, Q_a}{\sqrt{P_a \, P_a} \, \sqrt{Q_b \, Q_b}}. \tag{1.10}$$

We illustrate this equation in Figure 1.2. It is useful to convince oneself that this definition is consistent with $-1 \leq \cos\vartheta \leq 1$.

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.[3] 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 relation[4]

$$\vec{e}_a \wedge \vec{e}_b = \epsilon_{abc} \, \vec{e}_c. \tag{1.11}$$

The left hand side introduces the vector (or cross) product of two basis vectors. The right hand side algebrically defines the vector product as the contraction of the Levi-Civita tensor with another basis vector. The Cartesian components of the Levi-Civita tensor are given by the totally anti-symmetric permutation symbol or ϵ-tensor

$$\epsilon_{123} = 1 \tag{1.12a}$$

$$\epsilon_{abc} = \begin{cases} 1, \text{ even permutation of } abc \; (123, 312, 231) \\ -1, \text{ odd permutation of } abc \; (321, 132, 213) \\ 0, \text{ all other } \; abc. \end{cases} \tag{1.12b}$$

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

$$\epsilon_{abc} = -\epsilon_{bac} = -\epsilon_{acb}, \tag{1.13}$$

whereas cycling indices (an even permutation) preserves the sign

$$\epsilon_{abc} = \epsilon_{cab} = \epsilon_{bca}. \tag{1.14}$$

[3] There are surfaces, such as the Möbius strip, that are not orientable. We only consider orientiable surfaces in this book.

[4] Many 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.

1.4.2 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 $\vec{e}_b$

$$\begin{aligned}
\vec{e}_a \cdot (\vec{e}_a \wedge \vec{e}_b) &= \vec{e}_a \cdot \epsilon_{abc}\, \vec{e}_c && \text{definition} && (1.15a)\\
&= \epsilon_{abc}\, \vec{e}_a \cdot \vec{e}_c && \text{rearrangement} && (1.15b)\\
&= \epsilon_{cba}\, \vec{e}_c \cdot \vec{e}_a && \text{relabel } a \text{ to } c \text{ and } c \text{ to } a && (1.15c)\\
&= -\epsilon_{abc}\, \vec{e}_c \cdot \vec{e}_a && cba \text{ is an odd permutation of } abc && (1.15d)\\
&\Rightarrow \vec{e}_a \cdot (\vec{e}_a \wedge \vec{e}_b) = 0. && && (1.15e)
\end{aligned}$$

To fully digest step (1.15c) is can be useful to reintroduce the summation symbol so that

$$\begin{aligned}
\epsilon_{abc}\, \vec{e}_a \cdot \vec{e}_c &= \sum_{a=1}^{3}\sum_{c=1}^{3} \epsilon_{abc} \vec{e}_a \cdot \vec{e}_c && \text{summation symbols reintroduced} && (1.16a)\\
&= \sum_{c=1}^{3}\sum_{a=1}^{3} \epsilon_{cba} \vec{e}_c \cdot \vec{e}_a && \text{swap } a \text{ and } c && (1.16b)\\
&= \epsilon_{cba}\, \vec{e}_c \cdot \vec{e}_a && \text{reintroduce summation convention} && (1.16c)
\end{aligned}$$

Additionally, to digest step (1.15d) we step through the permutations

$$\begin{aligned}
\epsilon_{cba} &= -\epsilon_{bca} && \text{swap } c \text{ with } b \text{ to pick up a minus sign} && (1.17a)\\
&= \epsilon_{bac} && \text{swap } c \text{ with } a \text{ to pick up a minus sign} && (1.17b)\\
&= -\epsilon_{abc} && \text{swap } b \text{ with } a \text{ to pick up a minus sign.} && (1.17c)
\end{aligned}$$

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 vanishes. We further illustrate this property in Exercise 1.2.

1.4.3 Vector product of arbitrary vectors

The expression (1.11) for the vector product of two basis vectors renders the vector product of arbitrary vectors.

$$\begin{aligned}
\boldsymbol{P} \wedge \boldsymbol{Q} &= P_a\, \vec{e}_a \wedge Q_b\, \vec{e}_b && (1.18a)\\
&= P_a\, Q_b\, \vec{e}_a \wedge \vec{e}_b && (1.18b)\\
&= P_a\, Q_b\, \epsilon_{abc}\, \vec{e}_c && (1.18c)\\
&= (P_2\, Q_3 - P_3\, Q_2)\, \vec{e}_1 + (P_3\, Q_1 - P_1\, Q_3)\, \vec{e}_2 + (P_1\, Q_2 - P_2\, Q_1)\, \vec{e}_3. && (1.18d)
\end{aligned}$$

We can write the vector product as a determinant

$$\boldsymbol{P} \wedge \boldsymbol{Q} = \det \begin{bmatrix} \vec{e}_1 & \vec{e}_2 & \vec{e}_3 \\ p_1 & p_2 & p_3 \\ q_1 & q_2 & q_3 \end{bmatrix}. \tag{1.19}$$

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

$$\boldsymbol{P} \cdot (\boldsymbol{P} \wedge \boldsymbol{Q}) = (P_d\, \vec{e}_d) \cdot (P_a\, Q_b\, \epsilon_{abc}\, \vec{e}_c) \tag{1.20a}$$
$$= P_c\, P_a\, Q_b\, \epsilon_{abc} \tag{1.20b}$$
$$= 0, \tag{1.20c}$$

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

1.4.4 Geometric interpretation of the vector product

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, $|\boldsymbol{P} \wedge \boldsymbol{Q}| = |\boldsymbol{P}|\,|\boldsymbol{Q}| \sin\vartheta$. 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.*

The expression (1.18d) leads to the identity

$$|\boldsymbol{P} \wedge \boldsymbol{Q}|^2 = |\boldsymbol{P}|^2\, |\boldsymbol{Q}|^2 - (\boldsymbol{P} \cdot \boldsymbol{Q})^2 \tag{1.21a}$$
$$= |\boldsymbol{P}|^2\, |\boldsymbol{Q}|^2\, (1 - \cos^2\vartheta), \tag{1.21b}$$
$$= |\boldsymbol{P}|^2\, |\boldsymbol{Q}|^2 \sin^2\vartheta, \tag{1.21c}$$

where we used the scalar product expression (1.10) to introduce the angle subtended by the two vectors. Trigonometry indicates that the area of the parallelogram defined by $\boldsymbol{P}$ and $\boldsymbol{Q}$ is given by $|\boldsymbol{P}|\,|\boldsymbol{Q}| \sin\vartheta$. Hence, the vector product has a magnitude given by this area

$$\text{area}(\boldsymbol{P}, \boldsymbol{Q}) = |\boldsymbol{P}|\,|\boldsymbol{Q}| \sin\vartheta = |\boldsymbol{P} \wedge \boldsymbol{Q}|. \tag{1.22}$$

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

$$\boldsymbol{P} \wedge \boldsymbol{Q} = \hat{\boldsymbol{n}}\, \text{area}(\boldsymbol{P}, \boldsymbol{Q}) = \hat{\boldsymbol{n}}\, |\boldsymbol{P}|\,|\boldsymbol{Q}| \sin\vartheta, \tag{1.23}$$

where $\hat{\boldsymbol{n}}$ 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.

1.4.5 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 area($\boldsymbol{P}, \boldsymbol{Q}$) 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, $(\boldsymbol{P} \wedge \boldsymbol{Q}) \cdot \boldsymbol{R}$. This scalar product projects that portion of the vector $\boldsymbol{R}$ onto the direction parallel to the normal to the plane defined by $\boldsymbol{P} \wedge \boldsymbol{Q}$. Given that $|(\boldsymbol{P} \wedge \boldsymbol{Q})|$ is the area of the parallelogram defined by $\boldsymbol{P}$ and $\boldsymbol{Q}$, we conclude that $(\boldsymbol{P} \wedge \boldsymbol{Q}) \cdot \boldsymbol{R}$ 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 $\boldsymbol{P} \wedge \boldsymbol{Q}$ and $\boldsymbol{R}$. 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* $|(\boldsymbol{P} \wedge \boldsymbol{Q}) \cdot \boldsymbol{R}| = |(\boldsymbol{R} \wedge \boldsymbol{P}) \cdot \boldsymbol{Q}| = |(\boldsymbol{Q} \wedge \boldsymbol{R}) \cdot \boldsymbol{P}|$

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

$$
\begin{aligned}
(\boldsymbol{P} \wedge \boldsymbol{Q}) \cdot \boldsymbol{R} &= (P_a\, \vec{e}_a \wedge Q_b\, \vec{e}_b) \cdot R_d\, \vec{e}_d && (1.24a)\\
&= P_a\, Q_b\, (\vec{e}_a \wedge \vec{e}_b) \cdot R_d\, \vec{e}_d && (1.24b)\\
&= P_a\, Q_b\, (\epsilon_{abc}\, \vec{e}_c) \cdot \vec{e}_d\, R_d && (1.24c)\\
&= P_a\, Q_b\, \epsilon_{abc}\, (\vec{e}_c \cdot \vec{e}_d)\, R_d && (1.24d)\\
&= P_a\, Q_b\, \epsilon_{abc}\, \delta_{cd}\, R_d && (1.24e)\\
&= P_a\, Q_b\, \epsilon_{abc}\, R_c && (1.24f)\\
&= R_c\, P_a\, Q_b\, \epsilon_{abc} && (1.24g)\\
&= R_a\, P_b\, Q_c\, \epsilon_{bca} && (1.24h)\\
&= R_a\, P_b\, Q_c\, \epsilon_{abc} && (1.24i)\\
&= (\boldsymbol{R} \wedge \boldsymbol{P}) \cdot \boldsymbol{Q}. && (1.24j)
\end{aligned}
$$

This identity yields the geometric result illustrated in Figure 1.4

$$\text{volume}(\boldsymbol{P}, \boldsymbol{Q}, \boldsymbol{R}) = |(\boldsymbol{P} \wedge \boldsymbol{Q}) \cdot \boldsymbol{R}| = |(\boldsymbol{R} \wedge \boldsymbol{P}) \cdot \boldsymbol{Q}| = |(\boldsymbol{Q} \wedge \boldsymbol{R}) \cdot \boldsymbol{P}|. \tag{1.25}$$

1.5.2 Cartesian volume element for integration

We are in 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

$$\boldsymbol{P} = \hat{\boldsymbol{x}}\, \mathrm{d}x \qquad \text{and} \qquad \boldsymbol{Q} = \hat{\boldsymbol{y}}\, \mathrm{d}y \qquad \text{and} \qquad \boldsymbol{R} = \hat{\boldsymbol{z}}\, \mathrm{d}z, \tag{1.26}$$

in which case the volume element is

$$\mathrm{d}V = (\boldsymbol{P} \wedge \boldsymbol{Q}) \cdot \boldsymbol{R} = \mathrm{d}x\, \mathrm{d}y\, \mathrm{d}z\, (\hat{\boldsymbol{x}} \wedge \hat{\boldsymbol{y}}) \cdot \hat{\boldsymbol{z}} = \mathrm{d}x\, \mathrm{d}y\, \mathrm{d}z. \tag{1.27}$$

This expression for the volume element could have been written down without the formalism of a vector triple product. However, in Chapter 5 we find the general relation $(\boldsymbol{P} \wedge \boldsymbol{Q}) \cdot \boldsymbol{R}$ 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, (1.23), with the algebraic specification (1.18d) by writing

$$\text{2-volume} = \epsilon(\boldsymbol{P}, \boldsymbol{Q}) = \epsilon_{ab}\, P_a\, Q_b = \det \begin{bmatrix} P_1 & Q_1 \\ P_2 & Q_2 \end{bmatrix}. \tag{1.28}$$

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

$$\begin{bmatrix} \epsilon_{11} & \epsilon_{12} \\ \epsilon_{21} & \epsilon_{22} \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}. \tag{1.29}$$

In words, the first equality in equation (1.28) 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

$$\text{3-volume} = \epsilon(\boldsymbol{P}, \boldsymbol{Q}, \boldsymbol{R}) = \epsilon_{abc}\, P_a\, Q_b\, R_c = \det \begin{bmatrix} P_1 & Q_1 & R_1 \\ P_2 & Q_2 & R_2 \\ P_3 & Q_3 & R_3 \end{bmatrix}. \tag{1.30}$$

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

$$\text{surface area} = \epsilon(\ \ , \boldsymbol{Q}, \boldsymbol{R}). \tag{1.31}$$

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

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* details involved with tensor manipulations. A bit of practice and confidence readily allows one to compress many of these steps.

1.6.1 Double vector product

Consider the double vector product

$$\begin{aligned}
\boldsymbol{P} \wedge (\boldsymbol{Q} \wedge \boldsymbol{R}) &= P_a\, Q_b\, R_c\, \vec{e}_a \wedge (\vec{e}_b \wedge \vec{e}_c) && (1.32a)\\
&= P_a\, Q_b\, R_c\, \vec{e}_a \wedge (\epsilon_{bcd}\, \vec{e}_d) && (1.32b)\\
&= P_a\, Q_b\, R_c\, \epsilon_{bcd}\, \epsilon_{ade}\, \vec{e}_e && (1.32c)\\
&= -P_a\, Q_b\, R_c\, \epsilon_{bcd}\, \epsilon_{aed}\, \vec{e}_e. && (1.32d)
\end{aligned}$$

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

$$\epsilon_{bcd}\, \epsilon_{aed} = \delta_{ba}\, \delta_{ce} - \delta_{be}\, \delta_{ca}. \tag{1.33}$$

This identity then leads to

$$\epsilon_{bcd}\, \epsilon_{aed}\, \vec{e}_e = \delta_{ba}\, \vec{e}_c - \delta_{ca}\, \vec{e}_b \tag{1.34}$$

so that

$$\begin{aligned}
\boldsymbol{P} \wedge (\boldsymbol{Q} \wedge \boldsymbol{R}) &= -P_a\, Q_b\, R_c\, (\delta_{ba}\, \vec{e}_c - \delta_{ca}\, \vec{e}_b) && (1.35a)\\
&= -(\boldsymbol{P} \cdot \boldsymbol{Q})\, \boldsymbol{R} + (\boldsymbol{P} \cdot \boldsymbol{R})\, \boldsymbol{Q}. && (1.35b)
\end{aligned}$$

1.6.2 Scalar product of two vector products

We make further use of the Levi-Civita identity (1.33) to write

$$\begin{aligned}
(\boldsymbol{P} \wedge \boldsymbol{Q}) \cdot (\boldsymbol{R} \wedge \boldsymbol{S}) &= (\epsilon_{abc}\, p_a\, q_b)\, (\epsilon_{dec}\, R_d\, s_e) && (1.36a)\\
&= p_a\, q_b\, R_d\, s_e\, \epsilon_{abc}\, \epsilon_{dec} && (1.36b)\\
&= p_a\, q_b\, R_d\, s_e\, (\delta_{ad}\, \delta_{be} - \delta_{ae}\, \delta_{bd}) && (1.36c)\\
&= (\boldsymbol{P} \cdot \boldsymbol{R})\, (\boldsymbol{Q} \cdot \boldsymbol{S}) - (\boldsymbol{P} \cdot \boldsymbol{S})\, (\boldsymbol{Q} \cdot \boldsymbol{R}). && (1.36d)
\end{aligned}$$

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

$$\vec{e}_{\overline{a}} = \mathcal{R}_{\overline{a}b}\, \vec{e}_b. \tag{1.37}$$

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

$$\mathcal{R}_{\bar{a}b} = \begin{bmatrix} \mathcal{R}_{\bar{1}1} & \mathcal{R}_{\bar{1}2} & \mathcal{R}_{\bar{1}3} \\ \mathcal{R}_{\bar{2}1} & \mathcal{R}_{\bar{2}2} & \mathcal{R}_{\bar{2}3} \\ \mathcal{R}_{\bar{3}1} & \mathcal{R}_{\bar{3}2} & \mathcal{R}_{\bar{3}3} \end{bmatrix}. \tag{1.38}$$

In Cartesian tensor analysis studied in this chapter, the transformation matrix is assumed to be independent of space.[5] 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.

1.7.1 Inverse transformation

Assuming the transformation is invertible leads to the inverse transformation

$$\vec{e}_a = (\mathcal{R}^{-1})_{a\bar{b}}\, \vec{e}_{\bar{b}}. \tag{1.39}$$

As a self-consistency check we combine this relation with equation (1.37) thus rendering

$$\vec{e}_a = (\mathcal{R}^{-1})_{a\bar{b}}\, \vec{e}_{\bar{b}} = (\mathcal{R})^{-1}_{a\bar{b}}\, (\mathcal{R}_{\bar{b}c}\, \vec{e}_c). \tag{1.40}$$

This relation holds since

$$(\mathcal{R}^{-1})_{a\bar{b}}\, \mathcal{R}_{\bar{b}c} = \delta_{ac}, \tag{1.41}$$

or as a matrix identity

$$\mathcal{R}^{-1}\, \mathcal{R} = I. \tag{1.42}$$

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

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

leads to the following constraint on the transformation matrix

$$\delta_{\bar{a}\bar{b}} = \vec{e}_{\bar{a}} \cdot \vec{e}_{\bar{b}} \tag{1.43a}$$
$$= \mathcal{R}_{\bar{a}a}\vec{e}_a \cdot \mathcal{R}_{\bar{b}b}\vec{e}_b \tag{1.43b}$$
$$= \mathcal{R}_{\bar{a}a}\,\mathcal{R}_{\bar{b}b}\,\vec{e}_a \cdot \vec{e}_b \tag{1.43c}$$
$$= \mathcal{R}_{\bar{a}a}\,\mathcal{R}_{\bar{b}b}\,\delta_{ab} \tag{1.43d}$$
$$= \mathcal{R}_{\bar{a}a}\,\mathcal{R}_{\bar{b}a} \tag{1.43e}$$
$$= \mathcal{R}_{\bar{a}a}\,(\mathcal{R}^T)_{a\bar{b}}, \tag{1.43f}$$

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

$$(\mathcal{R}^T)_{a\bar{b}} = \mathcal{R}_{\bar{b}a}. \tag{1.44}$$

Written as a matrix equation we see that

$$\mathcal{R}\,\mathcal{R}^T = I. \tag{1.45}$$

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

$$\mathcal{R}^{-1} = \mathcal{R}^T. \tag{1.46}$$

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*

$$\mathcal{R}_{\bar{a}a} = \begin{bmatrix} \cos\vartheta & \sin\vartheta \\ -\sin\vartheta & \cos\vartheta \end{bmatrix} = \begin{bmatrix} \cos\vartheta & \cos(\pi/2-\vartheta) \\ \cos(\pi/2+\vartheta) & \cos\vartheta \end{bmatrix} = \begin{bmatrix} \vec{e}_{\bar{1}}\cdot\vec{e}_1 & \vec{e}_{\bar{1}}\cdot\vec{e}_2 \\ \vec{e}_{\bar{2}}\cdot\vec{e}_1 & \vec{e}_{\bar{2}}\cdot\vec{e}_2 \end{bmatrix}. \tag{1.47}$$

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 dimensionals as well, thus leading to

$$\mathcal{R}_{\bar{a}a} = \begin{bmatrix} \vec{e}_{\bar{1}}\cdot\vec{e}_1 & \vec{e}_{\bar{1}}\cdot\vec{e}_2 & \vec{e}_{\bar{1}}\cdot\vec{e}_3 \\ \vec{e}_{\bar{2}}\cdot\vec{e}_1 & \vec{e}_{\bar{2}}\cdot\vec{e}_2 & \vec{e}_{\bar{2}}\cdot\vec{e}_3 \\ \vec{e}_{\bar{3}}\cdot\vec{e}_1 & \vec{e}_{\bar{3}}\cdot\vec{e}_2 & \vec{e}_{\bar{3}}\cdot\vec{e}_3 \end{bmatrix}. \tag{1.48}$$

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 (1.39) 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

$$\boldsymbol{P} = P_a\,\vec{e}_a \tag{1.49a}$$
$$= P_a\,(\mathcal{R}^T)_{a\bar{a}}\,\vec{e}_{\bar{a}} \tag{1.49b}$$
$$\equiv P_{\bar{a}}\,\vec{e}_{\bar{a}}, \tag{1.49c}$$

where we defined the transformation of the vector components

$$P_{\bar{a}} = P_a\,(\mathcal{R}^T)_{a\bar{a}} = \mathcal{R}_{\bar{a}a}\,P_a. \tag{1.50}$$

1.7.5 Invariance of the scalar product

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

$$\boldsymbol{P} \cdot \boldsymbol{Q} = P_a \, Q_a \tag{1.51}$$

is the same regardless the choice of Cartesian coordinates

$$\begin{aligned}
\boldsymbol{P} \cdot \boldsymbol{Q} &= P_a \, \vec{e}_a \cdot Q_b \, \vec{e}_b && (1.52a)\\
&= P_a \, (\mathcal{R}^T)_{a\bar{a}} \, Q_b \, (\mathcal{R}^T)_{b\bar{b}} \, (\vec{e}_{\bar{a}} \cdot \vec{e}_{\bar{b}}) && (1.52b)\\
&= P_a \, (\mathcal{R}^T)_{a\bar{a}} \, Q_b \, (\mathcal{R}^T)_{b\bar{b}} \, \delta_{\bar{a}\bar{b}} && (1.52c)\\
&= P_a \, (\mathcal{R}^T)_{a\bar{a}} \, Q_b \, (\mathcal{R}^T)_{b\bar{a}} && (1.52d)\\
&= P_{\bar{a}} \, Q_{\bar{a}}. && (1.52e)
\end{aligned}$$

We return to such *form invariance* in Section 5.2, where it is referred to as *general covariance* in the context of general tensor analysis.

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 35.3.3; and the diffusion tensor in Chapter 47 are examples of second order tensors. Second order tensors have a coordinate representation given by

$$\mathcal{T} = T_{ab} \, \vec{e}_a \, \vec{e}_b, \tag{1.53}$$

with T_{ab} the Cartesian coordinates that can be organized as a matrix. Notably, there is no scalar product between the basis vectors in equation (1.53).

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

$$\begin{aligned}
\mathcal{T} &= T_{ab} \, \vec{e}_a \, \vec{e}_b && (1.54a)\\
&= T_{ab} \, (\mathcal{R}^T)_{a\bar{a}} \, \vec{e}_{\bar{a}} \, (\mathcal{R}^T)_{b\bar{b}} \, \vec{e}_{\bar{b}} && (1.54b)\\
&= T_{ab} \, (\mathcal{R}^T)_{a\bar{a}} \, (\mathcal{R}^T)_{b\bar{b}} \, \vec{e}_{\bar{a}} \, \vec{e}_{\bar{b}} && (1.54c)\\
&\equiv T_{\bar{a}\bar{b}} \, \vec{e}_{\bar{a}} \, \vec{e}_{\bar{b}}. && (1.54d)
\end{aligned}$$

The final equality introduced the transformed tensor components

$$T_{\bar{a}\bar{b}} = T_{ab} \, (\mathcal{R}^T)_{a\bar{a}} \, (\mathcal{R}^T)_{b\bar{b}} = T_{ab} \, \mathcal{R}_{\bar{a}a} \, \mathcal{R}_{\bar{b}b}. \tag{1.55}$$

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 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: $A\,B = (A\,B)^T$?

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: $\mathrm{Tr}(A\,S) = 0$. Alternatively, in terms of tensors, show that the full contraction of an anti-symmetric tensor with a symmetric tensor vanishes: $A_{mn}\,S_{mn} = 0$.

2

Cartesian tensor calculus

READER'S GUIDE TO THIS CHAPTER

This chapter presents differential and integral calculus of use for our study of fluid mechanics. We build from the Cartesian tensor algebra of Chapter 1 to develop elements of Cartesian tensor calculus. The material is used throughout this book so the reader is encouraged to master the basics here. For further study consult any book on calculus with analytic geometry. Particularly effective treatments, with applications to physics, are given in the following.

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 **20**
2.1.1 Direction of steepest ascent 20
2.1.2 Tangent to an isosurface 21
2.1.3 Normal to an isosurface 21
2.1.4 Unit vectors change only by rotation 22
2.1.5 Showing that $\delta\hat{\boldsymbol{n}} \cdot \hat{\boldsymbol{n}} = 0$ 23
2.2 The divergence of a vector field **23**
2.2.1 Divergence of a scalar field times a vector field 24
2.2.2 Laplacian of a scalar field 24
2.3 The curl of a vector field **24**
2.3.1 Computing the curl 24
2.3.2 Curl-free vector fields 25
2.3.3 Curl-free and divergence-free fields 26
2.3.4 Identities involving the curl 26

2.4 Path integral of a scalar function **28**
2.5 Path integral of a vector function **29**
2.5.1 Circulation 30
2.5.2 Circulation example 30
2.5.3 Fundamental theorem of calculus 30
2.6 Stokes' theorem **31**
2.6.1 Statement of Stokes' theorem 31
2.6.2 Stokes' theorem for a rectangular region 31
2.7 Gauss's divergence theorem **32**
2.7.1 An example rectangular volume 32
2.7.2 Divergence theorem for scalar fields 33
2.7.3 First and second form of Green's identities 33
2.7.4 Integral of a curl over a closed surface 34
2.8 Exact and inexact differentials **35**
2.8.1 Exact differentials 35
2.8.2 Inexact differentials 35
2.8.3 Integrating factors 36
2.8.4 An example using the velocity field 36
2.8.5 Heuristic physics of exact and inexact differential operations 36
2.9 Helmholz decomposition **37**
2.10 Exercises **37**

2.1 The gradient of a scalar field

Consider a scalar field $\psi(\boldsymbol{x})$ defined on Euclidean space with position $\boldsymbol{x} = x_a\,\vec{e}_a$. For example, this field may be the temperature at a point, the mass density, or the specific entropy. We may estimate its value at an adjacent point $\boldsymbol{x} + \mathrm{d}\boldsymbol{x}$ through a Taylor series

$$\psi(\boldsymbol{x} + \mathrm{d}\boldsymbol{x}) = \psi(\boldsymbol{x}) + \frac{\partial\psi}{\partial x_1}\,\mathrm{d}x_1 + \frac{\partial\psi}{\partial x_2}\,\mathrm{d}x_2 + \frac{\partial\psi}{\partial x_3}\,\mathrm{d}x_3 + \mathcal{O}\left(\mathrm{d}\boldsymbol{x}\cdot\mathrm{d}\boldsymbol{x}\right) \tag{2.1a}$$

$$\approx \left[1 + \mathrm{d}x_a\,\partial_a\right]\psi(\boldsymbol{x}), \tag{2.1b}$$

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

$$\partial_a = \frac{\partial}{\partial x_a}. \tag{2.2}$$

We can introduce the gradient operator according to

$$\nabla = \vec{e}_a\,\partial_a = \hat{\boldsymbol{x}}\,\partial_x + \hat{\boldsymbol{y}}\,\partial_y + \hat{\boldsymbol{z}}\,\partial_z \tag{2.3}$$

in which case

$$\psi(\boldsymbol{x} + \mathrm{d}\boldsymbol{x}) \approx \left(1 + \mathrm{d}\boldsymbol{x}\cdot\nabla\right)\psi(\boldsymbol{x}). \tag{2.4}$$

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

2.1.1 Direction of steepest ascent

Using the approximate relation (2.4), and the geometric expression (1.10) for the scalar product, renders

$$\psi(\boldsymbol{x} + \mathrm{d}\boldsymbol{x}) - \psi(\boldsymbol{x}) \approx |\mathrm{d}\boldsymbol{x}|\,|\nabla\psi|\cos\vartheta, \tag{2.5}$$

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 $\hat{\boldsymbol{n}} = |\nabla\psi|^{-1}\,\nabla\psi$. The unit tangent vector, $\hat{\boldsymbol{s}}$, 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: $\hat{\boldsymbol{n}} \cdot \hat{\boldsymbol{s}} = 0$.*

where ϑ is the angle between the differential increment $\mathrm{d}\boldsymbol{x}$ and the gradient $\nabla\psi$. Orienting the increment $\mathrm{d}\boldsymbol{x}$ so that $\vartheta = 0$ ensures that $\psi(\boldsymbol{x}+\mathrm{d}\boldsymbol{x})-\psi(\boldsymbol{x})$ 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

$$\psi(\boldsymbol{x}) = \text{constant} \tag{2.6}$$

Figure 2.1 shows a two dimensional example where the isosurfaces are lines where ψ is a constant. As another example, consider $\psi(\boldsymbol{x}) = \psi(r)$, where $r^2 = \boldsymbol{x} \cdot \boldsymbol{x}$ 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 $\hat{\boldsymbol{s}}$ be a unit vector that points in the direction tangent to the isosurface at any point $\boldsymbol{x}$. By construction

$$\psi(\boldsymbol{x} + \hat{\boldsymbol{s}}\,\mathrm{d}s) - \psi(\boldsymbol{x}) = 0, \tag{2.7}$$

where $\mathrm{d}s$ 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

$$\hat{\boldsymbol{s}} \cdot \nabla\psi = \frac{\partial\psi}{\partial s} = 0. \tag{2.8}$$

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, $\psi(\boldsymbol{x}) = \psi(r)$, the tangent vector points in either of the two angular directions along the spherical surface.

2.1.3 Normal to an isosurface

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

$$\hat{\boldsymbol{n}} = |\nabla\psi|^{-1}\,\nabla\psi. \tag{2.9}$$

Figure 2.2: *The infinitesimal change to a unit vector is itself perpendicular to the unit vector: $\delta\hat{\boldsymbol{m}} \cdot \hat{\boldsymbol{m}} = 0$. 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.*

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

$$\hat{\boldsymbol{n}} \cdot \nabla\psi = \frac{\partial\psi}{\partial n} = |\nabla\psi|. \tag{2.10}$$

For the spherically symmetric example,

$$\hat{\boldsymbol{n}} = \frac{\boldsymbol{x}}{|\boldsymbol{x}|} = \hat{\boldsymbol{r}}, \tag{2.11}$$

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

$$\hat{\boldsymbol{n}} \cdot \nabla\psi = \frac{\partial\psi}{\partial r} \qquad \text{spherically symmetric } \psi. \tag{2.12}$$

2.1.4 Unit vectors change only by rotation

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

$$\hat{\boldsymbol{m}} \cdot \hat{\boldsymbol{m}} = 1. \tag{2.13}$$

Unit vectors can only be modified through changes in their direction since their magnitude is fixed at unity. Hence, they are only modifed 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

$$0 = \delta(1) = \delta\left(\hat{\boldsymbol{m}} \cdot \hat{\boldsymbol{m}}\right) = 2\,\hat{\boldsymbol{m}} \cdot \delta\hat{\boldsymbol{m}}. \tag{2.14}$$

In Section 9.4, we formally show that the constraint

$$\delta\hat{\boldsymbol{m}} \cdot \hat{\boldsymbol{m}} = 0 \tag{2.15}$$

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* $\boldsymbol{F} = x\,\hat{\boldsymbol{x}} + y\,\hat{\boldsymbol{y}}$ *the field diverges from the origin with a spatially constant divergence* $\nabla \cdot \boldsymbol{F} = 2$*. Note that this vector field has zero curl,* $\nabla \wedge \boldsymbol{F} = 0$*.*

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

As an illustration of the constraint (2.15), let us verify that it holds for the special case of a unit normal vector (2.9) defined according to surfaces of constant scalar field

$$\hat{\boldsymbol{n}} = |\nabla\psi|^{-1}\,\nabla\psi. \tag{2.16}$$

The proof follows first by writing

$$\delta\hat{\boldsymbol{n}} = |\nabla\psi|^{-1}\,\left[\delta(\nabla\psi) - \hat{\boldsymbol{n}}\,\delta|\nabla\psi|\right], \tag{2.17}$$

so that

$$|\nabla\psi|\,\hat{\boldsymbol{n}} \cdot \delta\hat{\boldsymbol{n}} = \hat{\boldsymbol{n}} \cdot \delta(\nabla\psi) - \delta|\nabla\psi| = \frac{\nabla\psi \cdot \delta(\nabla\psi)}{|\nabla\psi|} - \delta|\nabla\psi| = 0. \tag{2.18}$$

The last equality made use of the identity

$$\delta(|\nabla\psi|) = \delta(\sqrt{\nabla\psi \cdot \nabla\psi}) = \frac{1}{2\sqrt{\nabla\psi \cdot \nabla\psi}}\delta(\nabla\psi \cdot \nabla\psi) = \frac{\nabla\psi \cdot \delta(\nabla\psi)}{|\nabla\psi|}. \tag{2.19}$$

2.2 The divergence of a vector field

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

$$\mathrm{div}(\boldsymbol{F}) = \nabla \cdot \boldsymbol{F} = \partial_a F_a \begin{cases} > 0 \Rightarrow \text{diverging vector field,} \\ < 0 \Rightarrow \text{converging vector field} \\ = 0 \Rightarrow \text{divergence-free vector field.} \end{cases} \tag{2.20}$$

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 14).

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

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

$$\nabla \cdot (\phi\,\boldsymbol{F}) = \partial_a(\phi\,F_a) \tag{2.21a}$$
$$= \partial_a(\phi)\,F_a + \phi\,\partial_a F_a \tag{2.21b}$$
$$= \boldsymbol{F} \cdot \nabla\phi + \phi\,\nabla \cdot \boldsymbol{F}. \tag{2.21c}$$

2.2.2 Laplacian of a scalar field

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

$$\nabla^2\psi = \nabla \cdot \nabla\psi. \tag{2.22}$$

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

$$\nabla^2\psi = 0 \qquad \text{harmonic function.} \tag{2.23}$$

Familiar examples of harmonic functions are the sines and cosines used for Fourier analysis in flat 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 the mathematical discipline of 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 use of the vorticity field, which is defined as the curl of the velocity (Chapter 34).

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

$$\begin{aligned}
\operatorname{curl}(\boldsymbol{F}) &= \nabla \wedge \boldsymbol{F} && (2.24a)\\
&= \vec{e}_a\,\partial_a \wedge \vec{e}_b\,F_b && (2.24b)\\
&= \vec{e}_a \wedge \vec{e}_b\,\partial_a F_b + \vec{e}_a\,F_b \wedge \partial_b \vec{e}_a && (2.24c)\\
&= \epsilon_{abc}\,\vec{e}_c\,\partial_a F_b && (2.24d)\\
&= \left(\frac{\partial F_3}{\partial x_2} - \frac{\partial F_2}{\partial x_3}\right)\hat{\boldsymbol{x}} + \left(\frac{\partial F_1}{\partial x_3} - \frac{\partial F_3}{\partial x_1}\right)\hat{\boldsymbol{y}} + \left(\frac{\partial F_2}{\partial x_1} - \frac{\partial F_1}{\partial x_2}\right)\hat{\boldsymbol{z}}. && (2.24e)
\end{aligned}$$

To reach this result we set $\partial_b\,\vec{e}_a = 0$ since the Cartesian basis vectors are fixed in space.[1] We also made use of the relation (1.11) for the cross product of basis vectors. We can express the curl from equation (2.24e) as a determinant

$$\nabla \wedge \boldsymbol{F} = \det \begin{bmatrix} \vec{e}_1 & \vec{e}_2 & \vec{e}_3 \\ \partial_1 & \partial_2 & \partial_3 \\ F_1 & F_2 & F_3 \end{bmatrix}. \tag{2.25}$$

The horizontal vector field $\boldsymbol{F} = x\,\hat{\boldsymbol{x}} + y\,\hat{\boldsymbol{y}}$ shown in Figure 2.3 has zero curl yet non-zero divergence. Figure 2.4 shows another vector field, $\boldsymbol{F} = -y\,\hat{\boldsymbol{x}} + x\,\hat{\boldsymbol{y}}$, with zero divergence yet nonzero curl $\nabla \wedge \boldsymbol{F} = 2\,\hat{\boldsymbol{z}}$. As seen in Section 32.5, 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

$$\nabla \wedge \boldsymbol{F} = 0. \tag{2.26}$$

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 31.1). We illustrate a curl-free vector field in Figure 2.5.

The curl of a gradient vanishes

$$\begin{aligned}
\nabla \wedge \nabla\psi &= \vec{e}_a\,\partial_a \wedge \vec{e}_b\,\partial_b\psi && (2.27a)\\
&= (\vec{e}_a \wedge \vec{e}_b)\,\partial_a\partial_b\psi && (2.27b)\\
&= 0, && (2.27c)
\end{aligned}$$

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

$$\boldsymbol{F} = -\nabla\psi \qquad \text{scalar potential.} \tag{2.28}$$

The scalar ψ is known as the *scalar potential*. In the specific case of $\boldsymbol{F}$ representing the gravitational force, then ψ is called the gravitational potential (see Section 9.12 and Chapter 43).

[1] Basis vectors corresponding to non-Cartesian coordinates are spatially dependent (see Chapters 5 and 6), thus making this step invalid for general tensors. We will find a "fix" for this step in Section 6.16 by defining the *covariant curl operator*.

Figure 2.5: *A horizontal vector field with a zero curl, where* $\boldsymbol{F} = -\nabla\psi$ *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

$$\nabla \wedge \boldsymbol{F} = 0 \Rightarrow \boldsymbol{F} = -\nabla\psi. \tag{2.29}$$

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

$$\nabla \cdot \nabla\psi = \nabla^2\psi = 0. \tag{2.30}$$

The velocity field arising from a linear non-rotating gravity wave (Section 31.1) in a Boussinesq fluid (Section 25.2) 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 5 and 6, 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

$$\nabla \cdot (\nabla \wedge \boldsymbol{F}) = \partial_a(\epsilon_{abc}\,\partial_b F_c) \tag{2.31a}$$
$$= \epsilon_{abc}\,\partial_a\partial_b F_c \tag{2.31b}$$
$$= 0. \tag{2.31c}$$

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

$$\nabla \cdot (\boldsymbol{F} \wedge \boldsymbol{E}) = \boldsymbol{E} \cdot (\nabla \wedge \boldsymbol{F}) - \boldsymbol{F} \cdot (\nabla \wedge \boldsymbol{E}) \tag{2.32}$$

through the following manipulations

$$\begin{aligned}
\nabla \cdot (\boldsymbol{F} \wedge \boldsymbol{E}) &= \vec{e}_a\, \partial_a\, (F_b\, \vec{e}_b \wedge E_c\, \vec{e}_c) && \text{(2.33a)}\\
&= \vec{e}_a \cdot (\vec{e}_b \wedge \vec{e}_c)\, \partial_a (F_b\, E_c) && \text{(2.33b)}\\
&= \epsilon_{abc}\, \partial_a (F_b\, E_c) && \text{(2.33c)}\\
&= F_b\, \epsilon_{abc}\, \partial_a E_c + E_c\, \epsilon_{abc}\, \partial_a F_b && \text{(2.33d)}\\
&= -\boldsymbol{F} \cdot (\nabla \wedge \boldsymbol{E}) + \boldsymbol{E} \cdot (\nabla \wedge \boldsymbol{F}). && \text{(2.33e)}
\end{aligned}$$

Curl of a scalar times a vector

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

$$\begin{aligned}
\nabla \wedge (\psi\, \boldsymbol{F}) &= \vec{e}_a\, \partial_a \wedge \psi\, \vec{e}_b\, F_b && \text{(2.34a)}\\
&= (\vec{e}_a \wedge \vec{e}_b)\, \partial_a (\psi\, F_b) && \text{(2.34b)}\\
&= \epsilon_{abc}\, \vec{e}_c\, (\psi\, \partial_a F_b + F_b\, \partial_a \psi) && \text{(2.34c)}\\
&= \psi\, \nabla \wedge \boldsymbol{F} + \nabla \psi \wedge \boldsymbol{F}. && \text{(2.34d)}
\end{aligned}$$

Curl of a cross product

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

$$\begin{aligned}
\nabla \wedge (\boldsymbol{F} \wedge \boldsymbol{E}) &= \vec{e}_a\, \partial_a \wedge (\vec{e}_b\, F_b \wedge \vec{e}_c\, E_c) && \text{(2.35a)}\\
&= \vec{e}_a \wedge (\vec{e}_b \wedge \vec{e}_c)\, \partial_a (F_b\, E_c) && \text{(2.35b)}\\
&= \vec{e}_a \wedge (\epsilon_{bcd}\, \vec{e}_d)\, \partial_a (F_b\, E_c) && \text{(2.35c)}\\
&= \epsilon_{ade}\, \epsilon_{bcd}\, \vec{e}_e\, \partial_a (F_b\, E_c) && \text{(2.35d)}\\
&= -\epsilon_{aed}\, \epsilon_{bcd}\, \partial_a (F_b\, E_c) && \text{(2.35e)}\\
&= -(\delta_{ab}\, \delta_{ec} - \delta_{ac}\, \delta_{eb})\, \vec{e}_e\, \partial_a (F_b\, E_c) && \text{(2.35f)}\\
&= (-\delta_{ba}\, \vec{e}_c + \delta_{ca}\, \vec{e}_b)\, \partial_a (F_b\, E_c) && \text{(2.35g)}\\
&= \boldsymbol{F}\, (\nabla \cdot \boldsymbol{E}) + (\boldsymbol{E} \cdot \nabla)\, \boldsymbol{F} - \boldsymbol{E}\, (\nabla \cdot \boldsymbol{F}) - (\boldsymbol{F} \cdot \nabla)\, \boldsymbol{E}. && \text{(2.35h)}
\end{aligned}$$

Curl of a curl

A special case of the identity (2.35h) allows us to write the curl of a curl as

$$\begin{aligned}
\nabla \wedge (\nabla \wedge \boldsymbol{F}) &= \vec{e}_a \wedge (\vec{e}_b \wedge \vec{e}_c)\, \partial_a \partial_b\, F_c && \text{(2.36a)}\\
&= (-\delta_{ba}\, \vec{e}_c + \delta_{ca}\, \vec{e}_b)\, \partial_a \partial_b\, F_c && \text{(2.36b)}\\
&= \nabla (\nabla \cdot \boldsymbol{F}) - \nabla^2 \boldsymbol{F}. && \text{(2.36c)}
\end{aligned}$$

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 34.3.1). Here, we aim to show that

$$(\boldsymbol{v}\cdot\nabla)\boldsymbol{v} = \boldsymbol{\omega}\wedge\boldsymbol{v} + \nabla(\boldsymbol{v}\cdot\boldsymbol{v})/2, \tag{2.37}$$

where

$$\boldsymbol{\omega} = \nabla\wedge\boldsymbol{v} \tag{2.38}$$

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

$$\begin{aligned}
\boldsymbol{\omega}\wedge\boldsymbol{v} &= (\nabla\wedge\boldsymbol{v})\wedge\boldsymbol{v} && \text{insert } \boldsymbol{\omega}=\nabla\wedge\boldsymbol{v} && (2.39a)\\
&= (\vec{e}_a\,\partial_a\wedge\vec{e}_b\,v_b)\wedge\vec{e}_c\,v_c && \text{Cartesian representation of } \boldsymbol{v} \text{ and } \nabla && (2.39b)\\
&= (\vec{e}_a\wedge\vec{e}_b)\wedge\vec{e}_c\,(\partial_a v_b)\,v_c && \text{rearrange} && (2.39c)\\
&= \epsilon_{abd}\,(\vec{e}_d\wedge\vec{e}_c)\,(\partial_a v_b)\,v_c && \text{first cross product expanded} && (2.39d)\\
&= \epsilon_{abd}\,\epsilon_{dce}\,\vec{e}_e\,(\partial_a v_b)\,v_c && \text{second cross product expanded} && (2.39e)\\
&= \epsilon_{abd}\,\epsilon_{ced}\,\vec{e}_e\,(\partial_a v_b)\,v_c && \text{arrange indices to prepare for next step} && (2.39f)\\
&= (\delta_{ac}\,\delta_{be}-\delta_{ae}\,\delta_{bc})\,\vec{e}_e\,(\partial_a v_b)\,v_c && \text{use identity (1.33)} && (2.39g)\\
&= \vec{e}_a\,v_c\,\partial_c v_a - \vec{e}_a\,v_c\,\partial_a v_c && \text{contract the Kronecker deltas} && (2.39h)\\
&= \vec{e}_a\,[(\boldsymbol{v}\cdot\nabla)\,v_a - \partial_a\boldsymbol{v}^2/2] && \text{re-express as Cartesian tensor} && (2.39i)\\
&= (\boldsymbol{v}\cdot\nabla)\,\boldsymbol{v} - \nabla[\boldsymbol{v}\cdot\boldsymbol{v}/2] && \text{rearrange.} && (2.39j)
\end{aligned}$$

Note that Section 4.4.4 of *Griffies* (2004) exhibits these steps making use of a general coordinate framework.

2.4 Path integral of a scalar function

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

$$\mathcal{I} = \int_{\mathcal{C}} \psi(\varphi)\,\mathrm{d}\varphi. \tag{2.40}$$

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 4.1 and 4.2 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

$$\mathcal{C}\mapsto\boldsymbol{x}(\varphi) = \hat{\boldsymbol{x}}\,x(\varphi) + \hat{\boldsymbol{y}}\,y(\varphi) + \hat{\boldsymbol{z}}\,z(\varphi) \tag{2.41}$$

so that the path integral can be written

$$\mathcal{I} = \int_{\mathcal{C}} \psi(\varphi)\,\mathrm{d}\varphi = \int_{\mathcal{C}} \psi[\boldsymbol{x}(\varphi)]\,\mathrm{d}\varphi. \tag{2.42}$$

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

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

$$\mathrm{d}s = \sqrt{\mathrm{d}\boldsymbol{x} \cdot \mathrm{d}\boldsymbol{x}} = \mathrm{d}s\,\sqrt{\frac{\mathrm{d}\boldsymbol{x}}{\mathrm{d}s} \cdot \frac{\mathrm{d}\boldsymbol{x}}{\mathrm{d}s}} = \mathrm{d}s\,|\boldsymbol{x}'(s)|, \tag{2.43}$$

so that the path integral takes the form

$$\mathcal{I} = \int_{\mathcal{C}} \psi[\boldsymbol{x}(s)]\,|\boldsymbol{x}'(s)|\,\mathrm{d}s. \tag{2.44}$$

Linear path example

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

$$\boldsymbol{x}(\varphi) = \boldsymbol{x}_A + (\boldsymbol{x}_B - \boldsymbol{x}_A)\,\varphi \qquad \varphi \in [0,1]. \tag{2.45}$$

Alternatively, we can parameterize using the arc length

$$\boldsymbol{x}(s) = \boldsymbol{x}_A + \hat{\boldsymbol{s}}\,s \qquad s \in [0, L], \tag{2.46}$$

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

$$\hat{\boldsymbol{s}} = \frac{\boldsymbol{x}'(s)}{|\boldsymbol{x}'(s)|} = \frac{\boldsymbol{x}_B - \boldsymbol{x}_A}{|\boldsymbol{x}_B - \boldsymbol{x}_A|}. \tag{2.47}$$

As defined we have $|\boldsymbol{x}'(s)| = |\hat{\boldsymbol{s}}| = 1$, so that the path integral is given by $\mathcal{I} = \int_0^L \psi[\boldsymbol{x}(s)]\,\mathrm{d}s$.

2.5 Path integral of a vector function

Generalizing to a vector field, $\boldsymbol{F}(\boldsymbol{x})$, 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 $\boldsymbol{F}(\boldsymbol{x})$ that projects onto a curve

$$\int_{\mathcal{C}} \boldsymbol{F} \cdot \mathrm{d}\boldsymbol{x} = \int_{\mathcal{C}} \boldsymbol{F} \cdot \frac{\mathrm{d}\boldsymbol{x}}{\mathrm{d}s}\, \mathrm{d}s \tag{2.48}$$

where $\mathrm{d}\boldsymbol{x}/\mathrm{d}s$ is tangent to the curve.

2.5.1 Circulation

For the case of a closed curve or a circuit (see Section 4.1), we refer to the above path integral as the *circulation* and use the convention of putting an arrowed circle on the integral sign

$$\text{circulation of vector field} = \oint_{\mathcal{C}} \boldsymbol{F} \cdot \mathrm{d}\boldsymbol{x}. \tag{2.49}$$

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 $\boldsymbol{F} = -y\,\hat{\boldsymbol{x}} + x\,\hat{\boldsymbol{y}}$ 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 $\boldsymbol{x}(\varphi) = r\,(\hat{\boldsymbol{x}}\,\cos\varphi + \hat{\boldsymbol{y}}\,\sin\varphi)$, and the tangent to the circle is $\mathrm{d}\boldsymbol{x}(\varphi)/\mathrm{d}\varphi = r\,(-\hat{\boldsymbol{x}}\,\sin\varphi + \hat{\boldsymbol{y}}\,\cos\varphi)$. The integrand to the circulation (2.48) thus takes the form

$$\boldsymbol{F} \cdot \frac{\mathrm{d}\boldsymbol{x}(\varphi)}{\mathrm{d}\varphi} = r\,(y\,\sin\varphi + x\cos\varphi) = r^2. \tag{2.50}$$

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

$$\oint_{\mathcal{C}} \boldsymbol{F} \cdot \mathrm{d}\boldsymbol{x} = 2\,\pi\,r^2. \tag{2.51}$$

This result has application for geophysical fluids when computing the vorticity induced by the rotating planet (see Section 34.6.2).

2.5.3 Fundamental theorem of calculus

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

$$\psi(\boldsymbol{x}_A) - \psi(\boldsymbol{x}_B) = \int_{\boldsymbol{x}_A}^{\boldsymbol{x}_B} \mathrm{d}\psi = \int_{\boldsymbol{x}_A}^{\boldsymbol{x}_B} \nabla\psi \cdot \mathrm{d}\boldsymbol{x}. \tag{2.52}$$

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

$$\oint_{\mathcal{C}} \mathrm{d}\psi = \oint_{\mathcal{C}} \nabla\psi \cdot \mathrm{d}\boldsymbol{x} = 0. \tag{2.53}$$

Figure 2.7: *Illustrating the geometry of Stokes' Theorem. The unit normal $\hat{\boldsymbol{n}}$ points outward from the surface, $\mathcal{S}$, with $(\nabla \wedge \boldsymbol{F}) \cdot \hat{\boldsymbol{n}}$ 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 $\hat{\boldsymbol{s}}$ 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

$$\oint_{\partial\mathcal{S}} \boldsymbol{F} \cdot \mathrm{d}\boldsymbol{x} = \int_{\mathcal{S}} (\nabla \wedge \boldsymbol{F}) \cdot \hat{\boldsymbol{n}} \, \mathrm{d}\mathcal{S}. \tag{2.54}$$

In this equation, $\mathrm{d}\boldsymbol{x}$ is the vector line element along the path, $\partial\mathcal{S}$ is the closed path defining the boundary to an oriented two-dimensional surface $\mathcal{S}$, $\hat{\boldsymbol{n}}$ is the outward normal vector on the surface, and $\mathrm{d}\mathcal{S}$ 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 $\hat{\boldsymbol{n}} = \hat{\boldsymbol{z}}$, so that

$$(\nabla \wedge \boldsymbol{F}) \cdot \hat{\boldsymbol{z}} = \frac{\partial F_2}{\partial x} - \frac{\partial F_1}{\partial y}, \tag{2.55}$$

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

$$\int_{\mathcal{S}} (\nabla \wedge \boldsymbol{F}) \cdot \hat{\boldsymbol{n}} \, \mathrm{d}S = \int_0^{L_x} \int_0^{L_y} \left(\frac{\partial F_2}{\partial x} - \frac{\partial F_1}{\partial y} \right) \mathrm{d}x \, \mathrm{d}y. \tag{2.56}$$

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

$$\int_0^{L_x} \int_0^{L_y} \left(\frac{\partial F_2}{\partial x} - \frac{\partial F_1}{\partial y} \right) \mathrm{d}x \, \mathrm{d}y = \int_0^{L_y} F_2(x,y) \Big|_{x=0}^{x=L_x} \mathrm{d}y - \int_0^{L_x} F_1(x,y) \Big|_{y=0}^{y=L_y} \mathrm{d}x \tag{2.57a}$$

$$= \int_0^{L_x} F_1(x,0) \, \mathrm{d}x + \int_0^{L_y} F_2(L_x,y) \, \mathrm{d}y + \int_{L_x}^{0} F_1(x,L_y) \, \mathrm{d}x + \int_{L_y}^{0} F_2(0,y) \, \mathrm{d}y \tag{2.57b}$$

$$= \oint_{\partial\mathcal{S}} \boldsymbol{F} \cdot \mathrm{d}\boldsymbol{x}. \tag{2.57c}$$

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

$$\int_{\mathcal{R}} \nabla \cdot \boldsymbol{F} \, \mathrm{d}V = \oint_{\partial \mathcal{R}} \boldsymbol{F} \cdot \hat{\boldsymbol{n}} \, \mathrm{d}\mathcal{S}, \tag{2.58}$$

where $\hat{\boldsymbol{n}}$ is the outward normal to the boundary surface and $\mathrm{d}\mathcal{S}$ 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 fluid mechanics jargon, we say that the divergence of a vector field, $\nabla \cdot \boldsymbol{F}$, integrated over a volume equals to the flux of that vector field, $\boldsymbol{F} \cdot \hat{\boldsymbol{n}}$, 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 (2.58) gives

$$\int_{\mathcal{R}} \left[\frac{\partial F_1}{\partial x} + \frac{\partial F_2}{\partial y} + \frac{\partial F_3}{\partial z} \right] \mathrm{d}x \, \mathrm{d}y \, \mathrm{d}z. \tag{2.59}$$

Focusing on just the leftmost term, integration in x gives

$$\int_{\mathcal{R}} \frac{\partial F_1}{\partial x}\,\mathrm{d}x\,\mathrm{d}y\,\mathrm{d}z = \int_{y=0}^{y=L_y}\int_{z=0}^{z=L_z} [F_1(L_x,y,z) - F_1(0,y,z)]\,\mathrm{d}y\,\mathrm{d}z \tag{2.60a}$$

$$= \int_{S_1+S_2} \boldsymbol{F}\cdot\hat{\boldsymbol{n}}\,\mathrm{d}\mathcal{S}, \tag{2.60b}$$

where S_1 is the rectangle's face with outward normal $\hat{\boldsymbol{n}} = \hat{\boldsymbol{x}}$ and S_2 is the rectangle's face with normal $\hat{\boldsymbol{n}} = -\hat{\boldsymbol{x}}$. Repeating this procedure on the other terms in equation (2.59) 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 to zero. 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 $\boldsymbol{F} = \phi\,\boldsymbol{c}$ with $\boldsymbol{c}$ an arbitrary *constant* vector. Substitution into the divergence theorem (2.58) yields

$$\oint_{\partial\mathcal{R}} \phi\,\boldsymbol{c}\cdot\hat{\boldsymbol{n}}\,\mathrm{d}\mathcal{S} = \int_{\mathcal{R}} \nabla\cdot(\phi\,\boldsymbol{c})\,\mathrm{d}V. \tag{2.61}$$

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

$$\boldsymbol{c}\cdot\left[\oint_{\partial\mathcal{R}} \phi\,\hat{\boldsymbol{n}}\,\mathrm{d}\mathcal{S} - \int_{\mathcal{R}} \nabla\phi\,\mathrm{d}V\right] = 0. \tag{2.62}$$

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

$$\oint_{\partial\mathcal{R}} \phi\,\hat{\boldsymbol{n}}\,\mathrm{d}\mathcal{S} = \int_{\mathcal{R}} \nabla\phi\,\mathrm{d}V. \tag{2.63}$$

In words, this result says that the integral of a scalar field over a closed surface, when weighted by the outward normal to the surface, equals to the volume integral of the gradient. 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 First and second form of Green's identities

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

$$\boldsymbol{F} = \psi\,\nabla\phi, \tag{2.64}$$

with ψ and ϕ scalar fields. Substitution into the divergence theorem (2.58) leads to

$$\oint_{\partial\mathcal{R}} \psi\,\frac{\partial\phi}{\partial n}\,\mathrm{d}\mathcal{S} = \int_{\mathcal{R}} [\nabla\psi\cdot\nabla\phi + \psi\,\nabla^2\phi]\,\mathrm{d}V \qquad \text{Green's first integral identity.} \tag{2.65}$$

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

$$\oint_{\partial\mathcal{R}} \left[\psi\,\frac{\partial\phi}{\partial n} - \phi\,\frac{\partial\psi}{\partial n}\right]\mathrm{d}\mathcal{S} = \int_{\mathcal{R}} [\psi\,\nabla^2\phi - \phi\,\nabla^2\psi]\,\mathrm{d}V \qquad \text{Green's second integral identity.} \tag{2.66}$$

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.*

Finally, setting $\phi = 1$ renders

$$\oint_{\partial\mathcal{R}} \frac{\partial\psi}{\partial n}\,\mathrm{d}\mathcal{S} = \int_{\mathcal{R}} \nabla^2\psi\,\mathrm{d}V \Longleftrightarrow \oint_{\partial\mathcal{R}} \nabla\psi\cdot\hat{\boldsymbol{n}}\,\mathrm{d}\mathcal{S} = \int_{\mathcal{R}} \nabla\cdot\nabla\psi\,\mathrm{d}V. \tag{2.67}$$

As seen in Section 3.7, these identities are fundamental to the Green's function method for solving linear partial differential equations.

2.7.4 Integral of a curl over a closed surface

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

$$\oint_{\partial\mathcal{R}} (\nabla\wedge\boldsymbol{\Psi})\cdot\hat{\boldsymbol{n}}\,\mathrm{d}\mathcal{S} = \int_{\mathcal{R}} \nabla\cdot(\nabla\wedge\boldsymbol{\Psi})\,\mathrm{d}V = 0, \tag{2.68}$$

where the final equality follows since the divergence of a curl vanishes. Hence, the integral 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)

$$\oint_{\partial\mathcal{R}} (\nabla\wedge\boldsymbol{\Psi})\cdot\hat{\boldsymbol{n}}\,\mathrm{d}\mathcal{S} = \int_{\mathcal{S}_1} (\nabla\wedge\boldsymbol{\Psi})\cdot\hat{\boldsymbol{n}}\,\mathrm{d}\mathcal{S}_1 + \int_{\mathcal{S}_2} (\nabla\wedge\boldsymbol{\Psi})\cdot\hat{\boldsymbol{n}}\,\mathrm{d}\mathcal{S}_2 \tag{2.69a}$$

$$= \oint_{\partial\mathcal{S}_1} \boldsymbol{\Psi}\cdot\mathrm{d}\boldsymbol{x} - \oint_{\partial\mathcal{S}_2} \boldsymbol{\Psi}\cdot\mathrm{d}\boldsymbol{x} \tag{2.69b}$$

$$= 0. \tag{2.69c}$$

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 32), we see that

$$\oint_{\partial\mathcal{R}} \boldsymbol{\omega}\cdot\hat{\boldsymbol{n}}\,\mathrm{d}\mathcal{S} = \int_{\mathcal{R}} \nabla\cdot\boldsymbol{\omega}\,\mathrm{d}V = 0, \tag{2.70}$$

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

2.8 Exact and inexact differentials

Thus far in this chapter all differentials have been exact. However, the thermodynamics discussed in Chapter 22 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, $\Phi(\boldsymbol{x})$. A differential increment for that function, computed between two close points $\boldsymbol{x}$ and $\boldsymbol{x}+\mathrm{d}\boldsymbol{x}$, is given by

$$\mathrm{d}\Phi(\boldsymbol{x}) = \Phi(\boldsymbol{x}+\mathrm{d}\boldsymbol{x}) - \Phi(\boldsymbol{x}) \tag{2.71a}$$

$$= \mathrm{d}\boldsymbol{x}\cdot\nabla\Phi, \tag{2.71b}$$

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

$$\Phi(\boldsymbol{x}_B) - \Phi(\boldsymbol{x}_A) = \int_{\boldsymbol{x}_A}^{\boldsymbol{x}_B} \mathrm{d}\Phi(\boldsymbol{x}) = \int_{\boldsymbol{x}_A}^{\boldsymbol{x}_B} \mathrm{d}\boldsymbol{x}\cdot\nabla\Phi. \tag{2.72}$$

These results are familiar from elementary calculus, with the increment $\mathrm{d}\Phi$ given by equation (2.71b) termed an *exact* differential. Importantly, the finite increment, $\Phi(\boldsymbol{x}_B) - \Phi(\boldsymbol{x}_A)$, depends only on the endpoint values of Φ. It does not depend on the path taken to go from $\boldsymbol{x}_A$ to $\boldsymbol{x}_B$. Correspondingly, the integral of an exact differential around a closed loop vanishes

$$\oint \mathrm{d}\Phi = 0. \tag{2.73}$$

2.8.2 Inexact differentials

Consider now a general differential expression written as

$$\boldsymbol{A}\cdot\mathrm{d}\boldsymbol{x} = A\,\mathrm{d}x + B\,\mathrm{d}y + C\,\mathrm{d}z, \tag{2.74}$$

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

$$\nabla\wedge\boldsymbol{A} = 0 \Longrightarrow \boldsymbol{A} = \nabla\Psi, \tag{2.75}$$

in which case we have an exact differential expression

$$\boldsymbol{A}\cdot\mathrm{d}\boldsymbol{x} = \nabla\Psi\cdot\mathrm{d}\boldsymbol{x} = \mathrm{d}\Psi. \tag{2.76}$$

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

$$\nabla\wedge\boldsymbol{A} = 0 \Longrightarrow \boldsymbol{A}\cdot\mathrm{d}\boldsymbol{x} \text{ exact differential.} \tag{2.77}$$

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

$$đ\Psi = \boldsymbol{A} \cdot \mathrm{d}\boldsymbol{x}. \tag{2.78}$$

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

$$\oint đ\Psi \neq 0. \tag{2.79}$$

2.8.3 Integrating factors

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

$$\nabla \wedge (\boldsymbol{A}\, \tau^{-1}) = 0. \tag{2.80}$$

Consequently, we can write

$$\boldsymbol{A} = \tau\, \nabla\Phi, \tag{2.81}$$

so that

$$đ\Psi = \boldsymbol{A} \cdot \mathrm{d}\boldsymbol{x} = \tau\, \nabla\Phi \cdot \mathrm{d}\boldsymbol{x} = \tau\, \mathrm{d}\Phi. \tag{2.82}$$

The function τ is known as an *integrating factor*. As seen in Section 22.2.3, 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 $\boldsymbol{v} \cdot \mathrm{d}\boldsymbol{x}$, where $\boldsymbol{v}$ is the velocity field for a fluid and $\mathrm{d}\boldsymbol{x}$ is a differential increment in space directed along a path. Furthermore, introduce the curl of the velocity, which defines the vorticity (Section 32.1) $\boldsymbol{\omega} = \nabla \wedge \boldsymbol{v}$. For cases where the vorticity vanishes, $\boldsymbol{\omega} = 0$, then $\mathrm{d}\Psi = \boldsymbol{v} \cdot \mathrm{d}\boldsymbol{x}$ is an exact differential. Consequently, Stokes' theorem means that the circulation of an irrotational velocity field computed around an arbitrary closed loop vanishes (Section 32.3)

$$\mathcal{C} \equiv \oint_{\partial S} \boldsymbol{v} \cdot \mathrm{d}\boldsymbol{x} = \int_S \boldsymbol{\omega} \cdot \hat{\boldsymbol{n}}\, \mathrm{d}S = 0. \tag{2.83}$$

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, $\boldsymbol{v} = \nabla\psi$, so that $\mathrm{d}\Psi = \nabla\psi \cdot \mathrm{d}\boldsymbol{x}$, 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 Helmholz 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, potential vorticity) has a local time change determined by the convergence of a flux, $\boldsymbol{J}$. 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 Helmholz decomposition theorem states that any vector field, $\boldsymbol{J}$, can be decomposed as

$$\boldsymbol{J} = -\nabla\Psi + \nabla \wedge \boldsymbol{A}. \tag{2.84}$$

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

$$\nabla \cdot \boldsymbol{J} = -\nabla^2\Psi \qquad \text{and} \qquad \nabla \wedge \boldsymbol{J} = \nabla \wedge (\nabla \wedge \boldsymbol{A}). \tag{2.85}$$

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 $\boldsymbol{F}$ (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 $\boldsymbol{A}$ are unique only up to a *gauge transformation* whereby

$$\Psi' = \Phi + C \qquad \text{and} \qquad \boldsymbol{A}' = \boldsymbol{A} + \nabla\lambda \tag{2.86}$$

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 *Maddision 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

$$\boldsymbol{x} = x\,\hat{\boldsymbol{x}} + y\,\hat{\boldsymbol{y}} + z\,\hat{\boldsymbol{z}} \tag{2.87}$$

and whose squared distance from the origin is

$$r^2 = \boldsymbol{x} \cdot \boldsymbol{x} = x^2 + y^2 + z^2. \tag{2.88}$$

EXERCISE 2.1: PRACTICE WITH THE GRADIENT OPERATOR
Prove the following identities:

(a) $\nabla(|\boldsymbol{x}|) = \boldsymbol{x}\,|\boldsymbol{x}|^{-1} \equiv \hat{\boldsymbol{r}}$

(b) $\nabla \ln|\boldsymbol{x}| = \boldsymbol{x}\,|\boldsymbol{x}|^{-2} = \hat{\boldsymbol{r}}\,|\boldsymbol{x}|^{-1}$

(c) $\nabla|\boldsymbol{x}|^{-1} = -\boldsymbol{x}\,|\boldsymbol{x}|^{-3} = -\hat{\boldsymbol{r}}\,|\boldsymbol{x}|^{-2}$.

EXERCISE 2.2: PRACTICE WITH THE LAPLACIAN OPERATOR
Show that the Laplacian of the function

$$\psi = \frac{z\,x^2}{r^2} \tag{2.89}$$

is given by

$$\nabla^2\psi = \frac{2\,z\,(r^2 - 5\,x^2)}{r^4}. \tag{2.90}$$

Perform the proof using both Cartesian coordinates as well as spherical coordinates (see Figure 7.1), making use of the following expressions for Laplacian operator acting on a scalar field

$$\nabla^2\psi(x,y,z) = \frac{\partial^2\psi}{\partial x^2} + \frac{\partial^2\psi}{\partial y^2} + \frac{\partial^2\psi}{\partial z^2} \tag{2.91a}$$

$$\nabla^2\psi(\lambda,\phi,r) = \frac{1}{r^2\,\cos\phi}\left[\frac{1}{\cos\phi}\frac{\partial^2\psi}{\partial\lambda^2} + \frac{\partial}{\partial\phi}\left(\cos\phi\frac{\partial\psi}{\partial\phi}\right) + \cos\phi\,\frac{\partial}{\partial r}\left(r^2\,\frac{\partial\phi}{\partial r}\right)\right]. \tag{2.91b}$$

EXERCISE 2.3: MORE PRACTICE WITH OPERATORS
Prove the following identities with $r \neq 0$:

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

(b) $\nabla\cdot(\boldsymbol{x}/r^3) = 0$

(c) $\nabla\cdot(\boldsymbol{A}\wedge\boldsymbol{x}) = \boldsymbol{x}\cdot(\nabla\wedge\boldsymbol{A})$ for an arbitrary vector field $A(\boldsymbol{x})$.

(d) $\nabla\wedge[\boldsymbol{x}\,f(r)] = 0$ for an arbitrary function $f(r) = f(|\boldsymbol{x}|)$.

EXERCISE 2.4: SOLID-BODY ROTATION
Define a velocity field according to

$$\boldsymbol{v} = \boldsymbol{\Omega}\wedge\boldsymbol{x} \tag{2.92}$$

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 32.5. Show that $2\,\boldsymbol{\Omega} = \nabla\wedge\boldsymbol{v}$. See also Exercise 32.1.

EXERCISE 2.5: DIVERGENCE-FREE AND IRROTATIONAL VECTOR
Let Φ be a harmonic function so that $\nabla^2\Phi = 0$. Show that $\boldsymbol{v} = -\nabla\Phi$ satisfies

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

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

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, $\nabla \wedge \boldsymbol{F}$, of the following vector field vanishes

$$\boldsymbol{F} = 2\,x\,z\,\hat{\boldsymbol{x}} + 2\,y\,z^2\,\hat{\boldsymbol{y}} + (x^2 + 2\,y^2\,z - 1)\,\hat{\boldsymbol{z}}. \tag{2.93}$$

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

$$\psi = -[x^2\,z + (y\,z)^2 - z]. \tag{2.94}$$

EXERCISE 2.7: BELTRAMI FLOW
Beltrami flow is defined by velocity and vorticity fields satisfying

$$\nabla \cdot \boldsymbol{v} = 0 \tag{2.95a}$$

$$\boldsymbol{\omega} = \nabla \wedge \boldsymbol{v} = \lambda \boldsymbol{v} \tag{2.95b}$$

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

$$\boldsymbol{v} = (A\,\sin z + C\,\cos y)\,\hat{\boldsymbol{x}} + (B\,\sin x + A\,\cos z)\,\hat{\boldsymbol{y}} + (C\,\sin y + B\,\cos x)\,\hat{\boldsymbol{z}}, \tag{2.96}$$

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

$$u = \frac{\partial w}{\partial y} - \frac{\partial v}{\partial z} \qquad v = \frac{\partial u}{\partial z} - \frac{\partial w}{\partial x} \qquad w = \frac{\partial v}{\partial x} - \frac{\partial u}{\partial y}. \tag{2.97}$$

EXERCISE 2.8: PRACTICE WITH PATH INTEGRALS
Consider the vector field

$$\boldsymbol{F} = x\,y^2\,\hat{\boldsymbol{x}} + 2\,\hat{\boldsymbol{y}} + x\,\hat{\boldsymbol{z}}. \tag{2.98}$$

Let $\mathcal{L}$ be a path parameterized by

$$x = c\,t \qquad y = c/t \qquad z = d \qquad t \in [1, 2], \tag{2.99}$$

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

$$\int_{\mathcal{L}} \boldsymbol{F}\,\mathrm{d}t = c^3 \ln 2\,\hat{\boldsymbol{x}} + 2\,\hat{\boldsymbol{y}} + \frac{3\,c}{2}\,\hat{\boldsymbol{z}} \tag{2.100a}$$

$$\int_{\mathcal{L}} \boldsymbol{F}\,\mathrm{d}y = -\frac{3\,c^4}{8}\,\hat{\boldsymbol{x}} - c\,\hat{\boldsymbol{y}} - c^2\,\ln 2\hat{\boldsymbol{z}} \tag{2.100b}$$

$$\int_{\mathcal{L}} \boldsymbol{F} \cdot \mathrm{d}\boldsymbol{x} = c^4 \ln 2 - c, \tag{2.100c}$$

where $\mathrm{d}\boldsymbol{x} = \hat{\boldsymbol{x}}\,\mathrm{d}x + \hat{\boldsymbol{y}}\,\mathrm{d}y + \hat{\boldsymbol{z}}\,\mathrm{d}z$. 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 2.9: STOKES' THEOREM ON A PLANE
Show that

$$\mathcal{I} = \oint_{\partial\mathcal{S}} \left[y\,(4\,x^2 + y^2)\,\mathrm{d}x + x\,(2\,x^2 + 3\,y^2)\,\mathrm{d}y\right] = \frac{\pi}{2}\,b\,a^3 \tag{2.101}$$

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

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1, \tag{2.102}$$

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

$$\int_{\pi}^{0} \sin^2(2\,\phi)\,\mathrm{d}\phi = -\frac{\pi}{2}. \tag{2.103}$$

EXERCISE 2.10: PRACTICE WITH GAUSS'S DIVERGENCE THEOREM

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

$$\boldsymbol{F} = \frac{\alpha\,\boldsymbol{x}}{(r^2 + a^2)^{3/2}}, \tag{2.104}$$

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

$$\Phi = \oint_{|\boldsymbol{x}|=a\sqrt{3}} \boldsymbol{F}\cdot\hat{\boldsymbol{n}}\,\mathrm{d}\mathcal{S} = \frac{3\,\pi\,\alpha\,\sqrt{3}}{2}. \tag{2.105}$$

With $\boldsymbol{F}$ 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

$$\Phi = \int_{|\boldsymbol{x}|=a\sqrt{3}} \nabla\cdot\boldsymbol{F}\,\mathrm{d}V. \tag{2.106}$$

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

$$\int_{\mathcal{R}} \nabla\cdot\boldsymbol{F}\,\mathrm{d}V = \oint_{\partial\mathcal{R}} \boldsymbol{F}\cdot\hat{\boldsymbol{n}}\,\mathrm{d}\mathcal{S}, \tag{2.107}$$

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

EXERCISE 2.11: MORE PRACTICE WITH GAUSS'S DIVERGENCE THEOREM

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

(a) $\oint_{\partial\mathcal{R}} \boldsymbol{x}\cdot\hat{\boldsymbol{n}}\,\mathrm{d}\mathcal{S} = 3\int_{\mathcal{R}}\mathrm{d}V = 3\,V$, where $\mathcal{R}$ is a closed region bounded by $\partial\mathcal{R}$ and with volume $\int_{\mathcal{R}}\mathrm{d}V = V$.

(b) $\oint_{\partial\mathcal{R}} (\hat{\boldsymbol{n}}\wedge\boldsymbol{F})\,\mathrm{d}\mathcal{S} = \int_{\mathcal{R}} \nabla\wedge\boldsymbol{F}\,\mathrm{d}V$, for an arbitrary vector field $\boldsymbol{F}$ and with $\hat{\boldsymbol{n}}$ 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 $\boldsymbol{A} = \boldsymbol{F}\wedge\boldsymbol{C}$ where $\boldsymbol{C}$ is a constant vector.

(c) Let $\partial\mathcal{R}$ be a closed surface bounding a volume $\mathcal{R}$, and let $\boldsymbol{x}$ denote the position vector of a point measured from an arbitrary origin. Prove the following

$$\oint_{\partial\mathcal{R}} \frac{\hat{\boldsymbol{n}}\cdot\boldsymbol{x}}{r^3}\,\mathrm{d}\mathcal{S} = \begin{cases} 0 & \text{if the origin lies outside of } \partial\mathcal{R} \\ 4\pi & \text{if the origin lies inside of } \partial\mathcal{R}. \end{cases} \tag{2.108}$$

3

Partial differential equations

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

READER'S GUIDE TO THIS CHAPTER

One can penetrate much of this book without reading this chapter. However, the reader will be far less appreciative of 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.

There are many resources devoted to the study and application of PDEs throughout physics, engineering, and applied mathematics. Chapter 8 of *Hildebrand* (1976) offers a pedagogical starting point; chapter 7 of *Morse and Feshbach* (1953) is an authoritative physics discussion of Green's function methods; *Stakgold* (2000a,b) thoroughly develops the theory and methods available for boundary value problems encountered in physics, including Green's function methods; and *Duchateau and Zachmann* (1986) concisely summarize PDEs and provide nearly 300 worked exercises. Much of the presentation in this chapter follows *Duchateau and Zachmann* (1986).

3.1 The advection equation **42**
3.1.1 Constant advection velocity 42
3.1.2 Specifying the arbitrary functions resulting from PDEs 43
3.1.3 Further study 43
3.2 Characteristic curves for first order PDEs **44**
3.2.1 General formulation 44
3.2.2 Examples 45
3.3 Classifying second order PDEs **46**
3.4 Elliptic partial differential equations **46**
3.4.1 Some general properties of Laplace's equation 47
3.4.2 Mean-value property of harmonic functions 47
3.4.3 Laplace's boundary value problem 48

3.4.4 Poisson's equation . . . 49
3.4.5 Extended max-min principle for Poisson's equation . . . 49
3.4.6 Poisson's boundary value problem . . . 49
3.5 Parabolic partial differential equations . . . **50**
3.5.1 Initial and initial-boundary value problems . . . 50
3.5.2 Smoothing property . . . 51
3.5.3 Duhamel's superposition integral for the heat equation . . . 52
3.5.4 Further study . . . 53
3.6 Properties of hyperbolic equations . . . **53**
3.6.1 Initial value problem for the infinite-domain wave equation . . . 53
3.6.2 Domain of influence for wave signals . . . 55
3.6.3 Helmholz equation . . . 56
3.6.4 Duhamel's superposition integral for the wave equation . . . 56
3.6.5 Further study . . . 57
3.7 Green's function method . . . **57**
3.7.1 Dirac delta (generalized) function . . . 58
3.7.2 Poisson's equation . . . 58
3.7.3 An example Green's function for the heat equation . . . 59
3.7.4 Comments and further study . . . 60
3.8 Evolution of time averages . . . **60**
3.8.1 Time averages . . . 61
3.8.2 Massaging the double time integrals . . . 62
3.8.3 Making use of a double integral identity . . . 62

3.1 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 47.4, the tracer concentration in the absence of diffusion satisfies the advection equation

$$\frac{\partial C}{\partial t} + \boldsymbol{v} \cdot \nabla C = 0. \tag{3.1}$$

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., dust), which are tracers that do not significantly alter velocity (Section 15.2). We limit the present discussion to passive tracers.

3.1.1 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,

$$\frac{\partial C}{\partial t} + U \frac{\partial C}{\partial x} = 0 \tag{3.2}$$

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

$$C(x,t) = \Gamma(x - Ut) \tag{3.3}$$

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

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

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

The functional dependence $x - Ut$ reveals that as time progresses with $U > 0$, an observer that moves in the positive $\hat{\boldsymbol{x}}$ 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 $\boldsymbol{v} = U\hat{\boldsymbol{x}}$ with $U > 0$. The initially square pulse of tracer concentration is translated, unchanged, by the constant advection velocity.*

3.1.2 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

$$C(x, t = 0) = C_0 \sin x, \tag{3.5}$$

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 $\hat{\boldsymbol{x}}$ direction with speed U

$$C(x, t) = C_0 \sin(x - Ut). \tag{3.6}$$

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.

3.1.3 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 12, 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 47.4 and 47.5.

3.2 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 by

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

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 allows us to determine the functional dependence of the solutions to the 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.

3.2.1 General formulation

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

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

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

$$\frac{\mathrm{d}\Upsilon}{\mathrm{d}t} = 0 = \frac{\partial\Upsilon}{\partial\psi}\frac{\partial\psi}{\partial t} + \frac{\partial\Upsilon}{\partial t} \tag{3.9a}$$

$$\frac{\mathrm{d}\Upsilon}{\mathrm{d}x} = 0 = \frac{\partial\Upsilon}{\partial\psi}\frac{\partial\psi}{\partial x} + \frac{\partial\Upsilon}{\partial x}. \tag{3.9b}$$

So long as $\partial\Upsilon/\partial\psi \neq 0$ then the first order PDE (3.7) takes on the equivalent form

$$P\,\frac{\partial\Upsilon}{\partial x} + Q\,\frac{\partial\Upsilon}{\partial t} + R\,\frac{\partial\Upsilon}{\partial\psi} = 0. \tag{3.10}$$

Considering the ordered triplet, (P,Q,R), to be components to a vector in (x,t,ψ) space, then equation (3.10) reveals that (P,Q,R) is perpendicular to the direction in (x,t,ψ) space that is normal to the integral surface $\Upsilon(x,t,\psi) = \text{constant}$. That is, (P,Q,R) lives in the plane tangent to the integral surface. The solution space fills out a curve on the tangent plane known as the *characteristic curve*. This interpretation takes on a somewhat less abstract form if we consider the function ψ to measure the vertical position z of a surface $\psi(x,t) = z$, so that the integral surface is given by

$$\Upsilon(x,t,z) = \text{constant}. \tag{3.11}$$

Let us parameterize the characteristic curve by its arc-length s and let $\boldsymbol{r}$ be the position on a characteristic curve so that

$$\frac{\mathrm{d}\boldsymbol{r}}{\mathrm{d}s} = \hat{\boldsymbol{x}}\,\frac{\mathrm{d}x}{\mathrm{d}s} + \hat{\boldsymbol{t}}\,\frac{\mathrm{d}t}{\mathrm{d}s} + \hat{\boldsymbol{z}}\,\frac{\mathrm{d}\psi}{\mathrm{d}s}, \tag{3.12}$$

where $\hat{\boldsymbol{t}}$ points in the direction of increasing time. In order for $\mathrm{d}\boldsymbol{r}/\mathrm{d}s$ to point in the direction of the tangent to a characteristic curve requires

$$P = \mu\,\frac{\mathrm{d}x}{\mathrm{d}s} \qquad Q = \mu\,\frac{\mathrm{d}t}{\mathrm{d}s} \qquad R = \mu\,\frac{\mathrm{d}\psi}{\mathrm{d}s} \tag{3.13}$$

for μ an arbitrary function. These relations in turn imply the following ordinary differential equations for the characteristics

$$\frac{\mathrm{d}x}{P} = \frac{\mathrm{d}t}{Q} = \frac{\mathrm{d}\psi}{R}. \tag{3.14}$$

If any one of the functions P, Q, or R vanish, then we merely remove that piece of the above relations.

3.2.2 Examples

Let us ground the discussion by considering the linear homogeneous advection equation

$$U\,\frac{\partial\psi}{\partial x} + \frac{\partial\psi}{\partial t} = 0, \tag{3.15}$$

in which we identify $P = U$ and $Q = 1$. The single ODE defining the characteristic curve is given by

$$\frac{\mathrm{d}x}{U} = \frac{\mathrm{d}t}{1}, \tag{3.16}$$

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

$$x - U\,t = \alpha \tag{3.17}$$

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

$$U\,\frac{\partial\psi}{\partial x} + \frac{\partial\psi}{\partial t} = R. \tag{3.18}$$

The two ODEs defining the characteristic curve are

$$\frac{\mathrm{d}x}{U} = \frac{\mathrm{d}t}{1} = \frac{\mathrm{d}\psi}{R}. \tag{3.19}$$

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

$$\Gamma[x - U\,t, \psi - R\,t] = 0, \tag{3.20}$$

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

$$\psi = f(x - U\,t) + R\,t \tag{3.21}$$

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

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

$$x\,\frac{\partial\psi}{\partial x} + t\,\frac{\partial\psi}{\partial t} = \psi, \tag{3.22}$$

in which the ODEs determining the characteristics are given by

$$\frac{\mathrm{d}x}{x} = \frac{\mathrm{d}t}{t} = \frac{\mathrm{d}\psi}{\psi}. \tag{3.23}$$

We are thus led to the relations

$$\frac{t}{x} = \alpha_1 \qquad \frac{\psi}{x} = \alpha_2. \tag{3.24}$$

Hence, the general solution of the PDE (3.22) is given by

$$\Gamma(t/x, \psi/x) = 0 \Rightarrow \psi = x\, F(t/x) \tag{3.25}$$

for an arbitrary smooth function F.

3.3 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

$$A\,\frac{\partial^2\psi}{\partial x^2} + B\frac{\partial^2\psi}{\partial x\,\partial t} + C\frac{\partial^2\psi}{\partial t^2} = \Lambda. \tag{3.26}$$

For linear PDEs, A, B, C are arbitrary functions of space and time that are independent of ψ and Λ is a function of space and time and at most a linear function of ψ and its derivatives. Furthermore, for linear PDEs the most general solution 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 (3.26) are of principle importance for determining the character of the solutions, with importance placed on the sign of the discriminant $B^2 - 4\,A\,C$. By analogy with conic sections we classify 2nd order PDEs as follows

$$\text{PDE form} = \begin{cases} \text{hyperbolic} & B^2 - 4\,A\,C > 0 \\ \text{elliptic} & B^2 - 4\,A\,C < 0 \\ \text{parabolic} & B^2 - 4\,A\,C = 0. \end{cases} \tag{3.27}$$

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

$$\psi(x,t) = f(m\,x + t). \tag{3.28}$$

Plugging into the 2nd order PDE (3.26) with $\Lambda = 0$ leads to

$$A\,m^2 + B\,m + C = 0. \tag{3.29}$$

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.

3.4 Elliptic partial differential equations

The elliptic case from Section 3.3 has discriminant $B^2 - 4\,A\,C < 0$, in which case there are two imaginary characteristics. The canonical example elliptic equation is Laplace's equation

$$\frac{\partial^2\psi}{\partial x^2} + \frac{\partial^2\psi}{\partial y^2} = 0, \tag{3.30}$$

where we converted from t as an independent variable to the space coordinate y. Formally, this transition is realized by setting $t = \mathrm{i}\,y$, where $\mathrm{i} = \sqrt{-1}$. Laplace's equation is satisfied by time-independent (i.e., *steady*) solutions to the heat/diffusion equation (Section 3.5). Another common elliptic equation is Poisson's equation, which results from adding a source to Laplace's equation

$$\frac{\partial^2\psi}{\partial x^2} + \frac{\partial^2\psi}{\partial y^2} = \Lambda. \tag{3.31}$$

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. However, it 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.[1] For many purposes it is thus suitable to assume acoustic speeds are infinite, and in so doing to *filter* them out of the dynamical equations. In the process, the hyperbolic equation describing acoustic signals is converted into an elliptic equation.

3.4.1 Some general properties of Laplace's equation

Solutions to Laplace's equation, $\nabla^2\psi = 0$, are known as *harmonic functions*. Example harmonic functions are

$$\psi = x^3 - 3\,x\,y^2 \qquad \psi = \ln(x^2+y^2) \qquad \psi = e^{\gamma\,x}\cos(\gamma\,y) \qquad \psi = a\,x + b\,y, \tag{3.32}$$

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

$$\nabla^2(\psi\,\phi) = \psi\,\nabla^2\phi + 2\,\nabla\phi\cdot\nabla\psi + \phi\,\nabla^2\psi, \tag{3.33}$$

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 (3.36a)-(3.36b).

3.4.2 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 $\boldsymbol{x}_o$ 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 $\boldsymbol{x}_o$. In equations this property states that

$$\psi(\boldsymbol{x}_o) = \frac{\oint_{\mathcal{S}_R}\psi(\boldsymbol{x})\,\mathrm{d}\mathcal{S}}{\oint_{\mathcal{S}_R}\mathrm{d}\mathcal{S}}, \tag{3.34}$$

where $\mathcal{S}_R$ is a sphere with radius R centered at $\boldsymbol{x}_o$ with "area" given by

$$\oint_{\mathcal{S}_R}\mathrm{d}\mathcal{S} = \begin{cases} 2\pi R & n = 2 \text{ space dimensions} \\ 4\pi R^2 & n = 3 \text{ space dimensions.} \end{cases} \tag{3.35}$$

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, $T(\boldsymbol{x})$, in a region with zero heat sources and zero fluid flow. As shown in Section 47.2.4, the steady state temperature satisfies $\nabla^2 T = 0$, and as such it is harmonic and hence has no extrema within the domain.

[1] A 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 $\boldsymbol{x}_o$ equals to the area average of the function over a sphere centered at $\boldsymbol{x}_o$. We here illustrate this property for 3-dimensions, but it holds for arbitrary space dimensions.*

3.4.3 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

$$\nabla^2\psi = 0 \qquad \boldsymbol{x} \in \mathcal{R} \tag{3.36a}$$

$$\alpha\,\psi + \beta\,\hat{\boldsymbol{n}}\cdot\nabla\psi = f \qquad \boldsymbol{x} \in \partial\mathcal{R}, \tag{3.36b}$$

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

We can establish constraints on the boundary conditions that lead to a self-consistent Laplacian boundary value problem (3.36a)-(3.36b). We do so through the use of Gauss's divergence theorem (Section 2.7) in which integration over the full domain leads to

$$0 = \int_{\mathcal{R}} \nabla^2\psi\,\mathrm{d}V = \oint_{\partial\mathcal{R}} \hat{\boldsymbol{n}}\cdot\nabla\psi\,\mathrm{d}\mathcal{S}. \tag{3.37}$$

In physical applications the boundary condition (3.36b) usually appear 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

$$\psi = f \qquad \boldsymbol{x} \in \partial\mathcal{R}, \tag{3.38}$$

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 f. Thinking again about temperature, this boundary condition specifies the temperature on the domain boundary to equal $T = f$, with all boundary functions f 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

$$\oint_{\partial\mathcal{R}} f\,\mathrm{d}\mathcal{S} = 0. \tag{3.39}$$

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 (3.39) is not satisfied, then the interior temperature field cannot be harmonic so that it will not be in a steady state.

3.4.4 Poisson's equation

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

$$\nabla^2\psi = \Lambda \qquad \boldsymbol{x} \in \mathcal{R} \tag{3.40a}$$

$$\alpha\,\psi + \beta\,\hat{\boldsymbol{n}}\cdot\nabla\psi = f \qquad \boldsymbol{x} \in \partial\mathcal{R}, \tag{3.40b}$$

where $\Lambda(\boldsymbol{x})$ 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 (3.40b).

3.4.5 Extended max-min principle for Poisson's equation

A subharmonic function is one where

$$\nabla^2\psi = \Lambda \geq 0 \qquad \boldsymbol{x} \in \mathcal{R}. \tag{3.41}$$

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

$$\psi(\boldsymbol{x}_o) \leq \frac{\oint_{\mathcal{S}_R}\psi(\boldsymbol{x})\,\mathrm{d}\mathcal{S}}{\oint_{\mathcal{S}_R}\mathrm{d}\mathcal{S}}, \tag{3.42}$$

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

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 $\nabla^2 T = \Lambda \geq 0$ for regions with the heat source. The minimum principle (3.42) 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 $\nabla^2 T = 0$, that we recover the mean-value property of harmonic functions given by equation (3.34).

3.4.6 Poisson's boundary value problem

We follow the method in Section 3.4.3 to develop constraints on the boundary conditions applied as part of the Poisson boundary value problem (3.40a)-(3.40b). Use of Gauss's divergence theorem leads to the constraint

$$\oint_{\partial\mathcal{R}}\hat{\boldsymbol{n}}\cdot\nabla\psi\,\mathrm{d}\mathcal{S} = \int_{\mathcal{R}}\Lambda\,\mathrm{d}V. \tag{3.43}$$

We separately consider the Dirichlet and Neumann cases.

Dirichlet boundary condition

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

$$\psi = f \qquad \boldsymbol{x} \in \partial\mathcal{R}. \tag{3.44}$$

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 f. Thinking again about temperature, this boundary condition specifies the temperature on the domain boundary to equal $T = f$, with all boundary functions f 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

$$\oint_{\partial\mathcal{R}} f \, \mathrm{d}\mathcal{S} = \int_{\mathcal{R}} \Lambda \, \mathrm{d}V. \tag{3.45}$$

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 and not realize a steady state.

3.5 Parabolic partial differential equations

The parabolic case from Section 3.3, $B^2 - 4\,A\,C = 0$, contains a single real characteristic. The canonical example is the *heat equation*, which is also known as the *diffusion equation*

$$\frac{\partial \psi}{\partial t} = \kappa \, \frac{\partial^2 \psi}{\partial x^2}, \tag{3.46}$$

where $\kappa > 0$ is the kinematic diffusivity.

3.5.1 Initial and initial-boundary value problems

The *Cauchy Problem* is the name given to the initial value problem for the heat equation in the full space of reals, $\mathbb{R}^n$

$$\frac{\partial \psi}{\partial t} = \kappa \, \nabla^2 \psi \qquad \boldsymbol{x} \in \mathbb{R}^n, t > 0 \tag{3.47a}$$

$$\psi(\boldsymbol{x}, t = 0) = f(\boldsymbol{x}) \qquad \boldsymbol{x} \in \mathbb{R}^n \tag{3.47b}$$

$$|\psi(\boldsymbol{x}, t)| < \infty \qquad \boldsymbol{x} \in \mathbb{R}^n, t > 0. \tag{3.47c}$$

The general initial-boundary value problem over a finite domain $\mathcal{R}$ takes the form

$$\frac{\partial \psi}{\partial t} = \kappa \, \nabla^2 \psi \qquad \boldsymbol{x} \in \mathcal{R}, t > 0 \tag{3.48a}$$

$$\psi(\boldsymbol{x}, t = 0) = f(\boldsymbol{x}) \qquad \boldsymbol{x} \in \mathcal{R} \tag{3.48b}$$

$$\alpha(\boldsymbol{x}) \, \psi(\boldsymbol{x}, t) + \beta(\boldsymbol{x}) \, \hat{\boldsymbol{n}} \cdot \nabla \psi(\boldsymbol{x}, t) = g(\boldsymbol{x}, t) \qquad \boldsymbol{x} \in \mathcal{S}, t > 0, \alpha \, \beta \geq 0. \tag{3.48c}$$

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 less common in geophysical fluid applications.

3.5.2 Smoothing property

The extended max-min principle from Section 3.4.5 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 solution to Laplace's equation whereby the mean-value property holds (Section 3.4.2). 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 Dirichlet boundary conditions

$$\frac{\partial \psi}{\partial t} = \kappa \frac{\partial^2 \psi}{\partial x^2} \qquad 0 < x < \pi,\ t > 0 \tag{3.49a}$$

$$\psi(x,t) = f(x) \qquad 0 < x < \pi,\ t = 0 \tag{3.49b}$$

$$\psi(0,t) = \psi(\pi,t) = 0 \qquad t > 0, \tag{3.49c}$$

where x and t are non-dimensional space and time variables. A variety of methods can be used to construct the following Fourier series solution

$$\psi(x,t) = \sum_{n=1}^{\infty} f_n\, e^{-n^2 t} \sin(n\,x) \qquad \text{with} \qquad f_n = \frac{2}{\pi} \int_0^{\pi} f(x) \sin(n\,x)\, \mathrm{d}x. \tag{3.50}$$

As per the smoothing property, 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

$$\frac{\partial \psi}{\partial t} = \kappa \frac{\partial^2 \psi}{\partial x^2} \qquad -\infty < x < \infty,\ t > 0 \tag{3.51a}$$

$$\psi(x,t) = f(x) \qquad -\infty < x < \infty,\ t = 0. \tag{3.51b}$$

One can show by direct differentiation that the following Gaussian is a solution

$$\psi(x,t) = \frac{1}{\sqrt{4\,\pi\,\kappa\,t}} \int_{-\infty}^{\infty} f(\xi)\, \exp\left[-\frac{(x-\xi)^2}{4\,\kappa\,t}\right] \mathrm{d}\xi. \tag{3.52}$$

Again, this function smooths/damps the initial condition function $f(x)$ as time increases.

3.5.3 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

$$\frac{\partial \Psi}{\partial t} = \kappa \, \nabla^2 \Psi + f(\boldsymbol{x}) \qquad \boldsymbol{x} \in \mathbb{R}^n, \; t > 0 \tag{3.53a}$$

$$\Psi(\boldsymbol{x}, t) = 0 \qquad \boldsymbol{x} \in \mathbb{R}^n, \; t = 0. \tag{3.53b}$$

Now consider the converse, in which another scalar field evolves without a source and yet is initialized according to the source

$$\frac{\partial \psi}{\partial t} = \kappa \, \nabla^2 \psi \qquad \boldsymbol{x} \in \mathbb{R}^n, \; t > 0 \tag{3.54a}$$

$$\psi(\boldsymbol{x}, t) = f(\boldsymbol{x}) \qquad \boldsymbol{x} \in \mathbb{R}^n, \; t = 0. \tag{3.54b}$$

The two scalar fields are related by *Duhamel's superposition integral*

$$\Psi(\boldsymbol{x}, t) = \int_0^t \psi(\boldsymbol{x}, t - \tau) \, \mathrm{d}\tau. \tag{3.55}$$

We verify the connection by direct differentiation

$$\frac{\partial \Psi(\boldsymbol{x}, t)}{\partial t} = \psi(\boldsymbol{x}, 0) + \int_0^t \frac{\partial \psi(\boldsymbol{x}, t - \tau)}{\partial t} \, \mathrm{d}\tau = f(\boldsymbol{x}) + \kappa \, \nabla^2 \, \Psi(\boldsymbol{x}, t). \tag{3.56}$$

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 $\Psi(\boldsymbol{x}, t)$ 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, $f(\boldsymbol{x}, t)$, in which case we must develop a family of solutions, $\psi_f(\boldsymbol{x}, t; \tau)$, generated by reinitializing $\psi_f(\boldsymbol{x}, t = \tau; \tau) = f(\boldsymbol{x}, \tau)$ and then superposing the members of this family to generate $\Psi(\boldsymbol{x}, t)$.

As an example, consider the initial value problem for the heat equation on a line as given by equations (3.51a)-(3.51b)

$$\frac{\partial \psi}{\partial t} = \kappa \, \frac{\partial^2 \psi}{\partial x^2} \qquad -\infty < x < \infty, \; t > 0 \tag{3.57a}$$

$$\psi(x, t) = f(x) \qquad -\infty < x < \infty, \; t = 0, \tag{3.57b}$$

whose solution is given by the Gaussian function in equation (3.52). Duhamel's superposition integral (3.55) says that

$$\Psi(x, t) = \int_0^t \psi(x, t - \tau) \, \mathrm{d}\tau = \int_0^t \frac{1}{\sqrt{4 \, \pi \, \kappa \, (t - \tau)}} \int_{-\infty}^{\infty} f(\xi) \, \exp\left[-\frac{(x - \xi)^2}{4 \, \kappa \, (t - \tau)} \right] \mathrm{d}\tau \, \mathrm{d}\xi \tag{3.58}$$

satisfies the forced (inhomogeneous) initial value problem with zero initial condition

$$\frac{\partial \Psi}{\partial t} = \kappa \, \frac{\partial^2 \Psi}{\partial x^2} + f(x) \qquad -\infty < x < \infty, \; t > 0 \tag{3.59a}$$

$$\Psi(x, t) = 0 \qquad -\infty < x < \infty, \; t = 0. \tag{3.59b}$$

We make use of this result in our discussion of Green's functions in Section 3.7.

3.5.4 Further study

We examine physical and mathematical properties of the heat/diffusion equation in Sections 47.2 and 47.3. Section 9.11 of *Hildebrand* (1976) offers a lucid discussion of Duhamel's superposition integral.

3.6 Properties of hyperbolic equations

The hyperbolic case from Section 3.3 has $B^2 - 4\,A\,C > 0$ and thus contains two real characteristics. The canonical example of a hyperbolic PDE is the linear homogeneous wave equation

$$\frac{\partial^2 \psi}{\partial t^2} - U^2\,\frac{\partial^2 \psi}{\partial x^2} = 0. \tag{3.60}$$

Solutions have the form of a moving wave in both directions (the two wave characteristics)

$$\psi(x,t) = \mathcal{F}(x - Ut) + \mathcal{G}(x + U\,t), \tag{3.61}$$

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

$$(\partial_t - U\,\partial_x)\,(\partial_t + U\,\partial_x)\,\psi = 0. \tag{3.62}$$

Consequently, if either one of the linear first-order PDEs are satisfied

$$(\partial_t - U\,\partial_x)\psi = 0 \tag{3.63a}$$
$$(\partial_t + U\,\partial_x)\psi = 0 \tag{3.63b}$$

then ψ will satisfy the full wave equation. These first-order PDEs are the one-dimensional advection equations considered in Section 3.2 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*.

3.6.1 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 (no boundary conditions)

$$\frac{\partial^2 \psi}{\partial t^2} = c^2\,\frac{\partial^2 \psi}{\partial x^2} \qquad -\infty < x < \infty,\ t > 0 \qquad \text{wave equation on a line} \tag{3.64a}$$
$$\psi = F(x) \qquad -\infty < x < \infty,\ t = 0 \qquad \text{initial condition} \tag{3.64b}$$
$$\frac{\partial \psi}{\partial t} = G(x) \qquad -\infty < x < \infty,\ t = 0 \qquad \text{initial tendency,} \tag{3.64c}$$

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 3.2, we are motivated to transform the wave equation into wave characteristic coordinates

$$\xi = x + c\,t \qquad \text{and} \qquad \eta = x - c\,t \Longrightarrow \frac{\xi + \eta}{2} = x \qquad \text{and} \qquad \frac{\xi - \eta}{2c} = t. \tag{3.65}$$

Wave signals propagate in directions defined by constant ξ and η, so that these coordinates isolate the signal tranmission. 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 (3.65) as a matrix-vector equation

$$\begin{bmatrix} \xi \\ \eta \end{bmatrix} = \begin{bmatrix} c & 1 \\ -c & 1 \end{bmatrix} \begin{bmatrix} t \\ x \end{bmatrix} \Longleftrightarrow \begin{bmatrix} t \\ x \end{bmatrix} = \frac{1}{2} \begin{bmatrix} c^{-1} & -c^{-1} \\ 1 & 1 \end{bmatrix} \begin{bmatrix} \xi \\ \eta \end{bmatrix}. \tag{3.66}$$

Furthermore, define

$$x^n = (x^0, x^1) = (x, t) \qquad \text{and} \qquad x^{\overline{n}} = (x^{\overline{0}}, x^{\overline{1}}) = (\xi, \eta) \tag{3.67}$$

and use index notation from Chapter 1 so that the transformation (3.66) takes the tidy form

$$x^{\overline{n}} = \Lambda^{\overline{n}}{}_m \, x^m \Longleftrightarrow x^n = \Lambda^n{}_{\overline{m}} \, x^{\overline{m}}, \tag{3.68}$$

where the transformation matrices are given by

$$\Lambda^{\overline{n}}{}_m = \begin{bmatrix} c & 1 \\ -c & 1 \end{bmatrix} \qquad \text{and} \qquad \Lambda^n{}_{\overline{m}} = \frac{1}{2} \begin{bmatrix} c^{-1} & -c^{-1} \\ 1 & 1 \end{bmatrix}. \tag{3.69}$$

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 5. 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 (3.68) and the transformation matrices (3.69) then lead to the partial derivative relationship

$$\partial_{\overline{n}} = \Lambda^m{}_{\overline{n}} \, \partial_m \Longleftrightarrow \partial_n = \Lambda^{\overline{m}}{}_n \, x_{\overline{m}}, \tag{3.70}$$

so that

$$\frac{\partial^2}{\partial x^2} = \left[\frac{\partial}{\partial \xi} + \frac{\partial}{\partial \eta} \right]^2 = \frac{\partial^2}{\partial \xi^2} + 2 \, \frac{\partial}{\partial \xi} \, \frac{\partial}{\partial \eta} + \frac{\partial^2}{\partial \eta^2} \tag{3.71a}$$

$$c^{-2} \, \frac{\partial^2}{\partial t^2} = \left[\frac{\partial}{\partial \xi} - \frac{\partial}{\partial \eta} \right]^2 = \frac{\partial^2}{\partial \xi^2} - 2 \, \frac{\partial}{\partial \xi} \, \frac{\partial}{\partial \eta} + \frac{\partial^2}{\partial \eta^2}. \tag{3.71b}$$

General solution for the initial value problem

The operator transformations (3.71a) and (3.71b) bring the initial value problem (3.64a)-(3.64c) into

$$\frac{\partial^2 \psi}{\partial \xi \partial \eta} = 0 \qquad -\infty < \eta < \xi < \infty \tag{3.72a}$$

$$\psi(\xi, \eta) = F(\xi) \qquad -\infty < \xi < \infty, \; \xi = \eta \tag{3.72b}$$

$$\frac{\partial \psi}{\partial \xi} - \frac{\partial \psi}{\partial \eta} = c^{-1} \, G(\xi) \qquad -\infty < \xi < \infty, \; \xi = \eta. \tag{3.72c}$$

Integrating equation (3.72a) in two steps leads to $\partial_\xi \psi = \theta(\xi)$ so that

$$\psi(\xi, \eta) = \Phi(\eta) + \int^{\xi} \theta(s)\, \mathrm{d}s \equiv \Phi(\eta) + \Theta(\xi), \tag{3.73}$$

for two functions $\Phi(\eta)$ and $\Theta(\xi)$. The initial conditions (3.72b) and (3.72c) determine relations between $\Phi(\eta)$ and $\Theta(\xi)$ and the initial data

$$\Theta(\xi) = \frac{1}{2}\left[F(\xi) + \frac{1}{c}\int^{\xi} G(s)\, \mathrm{d}s\right] \tag{3.74a}$$

$$\Phi(\eta) = \frac{1}{2}\left[F(\eta) - \frac{1}{c}\int^{\eta} G(s)\, \mathrm{d}s\right], \tag{3.74b}$$

in which case the general solution to the initial value problem (3.64a)-(3.64c) takes the form

$$\psi(x, t) = \frac{1}{2}\left[F(x + c\,t) + F(x - c\,t)\right] + \frac{1}{2c}\int_{x-c\,t}^{x+c\,t} G(s)\, \mathrm{d}s, \tag{3.75}$$

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, $\xi = x + c\,t$ and $\eta = x - c\,t$, 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 3.5, with its single time derivative resulting in a smoothing of the full solution.

Figure 3.3: *According to the wave solution (3.75) 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, $\xi_o = x_o + c\,t_o$ and $\eta_o = x_o - c\,t_o$, 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 $x_o - c\,t_o \leq x \leq x_o + c\,t_o$. Points outside the domain of influence are causally disconnected from the point (x_o, t_o).*

3.6.2 Domain of influence for wave signals

The wave solution (3.75) 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, $\xi_o = x_o + c\,t_o$ and $\eta_o = x_o - c\,t_o$. 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).

3.6.3 Helmholz equation

Consider the wave equation with a constant wave speed

$$\frac{\partial^2 \psi}{\partial t^2} - c^2 \nabla^2 \psi = 0. \tag{3.76}$$

Assuming a wave ansatz[2] of the form

$$\psi(\boldsymbol{x}, t) = e^{\mathrm{i}\,\omega\, t}\, \Psi(\boldsymbol{x}) \tag{3.77}$$

results in the *Helmholz equation* for the amplitude function

$$(\nabla^2 + k^2)\, \Psi = 0, \tag{3.78}$$

where $k = \omega/c$ is the wavenumber. The Helmholz equation thus plays a central role in the wave theory.

3.6.4 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 3.5.3, For this purpose, consider the forced wave equation with time independent forcing

$$\frac{\partial^2 \Psi}{\partial t^2} = c^2\, \nabla^2 \Psi + G(\boldsymbol{x}) \qquad \boldsymbol{x} \in \mathbb{R}^n,\ t > 0 \tag{3.79a}$$

$$\Psi(\boldsymbol{x}, t) = 0 \qquad \boldsymbol{x} \in \mathbb{R}^n,\ t = 0, \tag{3.79b}$$

and the corresponding unforced wave equation with inhomogenous initial time tendency

$$\frac{\partial^2 \psi}{\partial t^2} = c^2\, \nabla^2 \psi \qquad \boldsymbol{x} \in \mathbb{R}^n, t > 0 \tag{3.80a}$$

$$\psi(\boldsymbol{x}, t) = 0 \qquad \boldsymbol{x} \in \mathbb{R}^n,\ t = 0. \tag{3.80b}$$

$$\frac{\partial \psi(\boldsymbol{x}, t)}{\partial t} = G(\boldsymbol{x}) \qquad \boldsymbol{x} \in \mathbb{R}^n,\ t = 0. \tag{3.80c}$$

The two scalar fields are related by Duhamel's superposition integral

$$\Psi(\boldsymbol{x}, t) = \int_0^t \psi(\boldsymbol{x}, t - \tau)\, \mathrm{d}\tau. \tag{3.81}$$

We can verify this formula through direct differentiation

$$\frac{\partial \Psi(\boldsymbol{x}, t)}{\partial t} = \int_0^t \frac{\partial \psi(\boldsymbol{x}, t - \tau)}{\partial t}\, \mathrm{d}\tau \tag{3.82a}$$

$$\frac{\partial^2 \Psi(\boldsymbol{x}, t)}{\partial t^2} = \frac{\partial \psi(\boldsymbol{x}, 0)}{\partial t} + \int_0^t \frac{\partial^2 \psi(\boldsymbol{x}, t - \tau)}{\partial t^2}\, \mathrm{d}\tau = G(\boldsymbol{x}) + c^2\, \nabla^2 \Psi(\boldsymbol{x}, t). \tag{3.82b}$$

[2] *Ansatz* is a German word meaning "educated guess".

As an example, consider the initial value problem for wave equation on a line with time independent forcing

$$\frac{\partial^2 \Psi}{\partial t^2} = c^2 \frac{\partial^2 \Psi}{\partial x^2} + G(x) \qquad -\infty < x < \infty,\ t > 0 \tag{3.83a}$$

$$\Psi(x,t) = 0 \qquad -\infty < x < \infty,\ t = 0 \tag{3.83b}$$

$$\frac{\partial \Psi(x,t)}{\partial t} = 0 \qquad -\infty < x < \infty,\ t = 0. \tag{3.83c}$$

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

$$\frac{\partial^2 \psi}{\partial t^2} = c^2 \frac{\partial^2 \psi}{\partial x^2} \qquad -\infty < x < \infty,\ t > 0 \tag{3.84a}$$

$$\psi(x,t) = 0 \qquad -\infty < x < \infty, t = \tau \tag{3.84b}$$

$$\frac{\partial \psi(x,t)}{\partial t} = G(x) \qquad -\infty < x < \infty,\ t = \tau > 0. \tag{3.84c}$$

We know from Section 3.6.1 that the solution ψ is given by the D'Alembert formula in equation (3.75), only here with the initial condition function set to zero

$$\psi(x,t) = \frac{1}{2c} \int_{x-ct}^{x+ct} G(s)\,\mathrm{d}s. \tag{3.85}$$

Hence, D'Alembert's formula says that the solution to the forced wave equation (3.83a)-(3.83c) is given by the superposition integral

$$\Psi(x,t) = \frac{1}{2c} \int_0^t \int_{x-c(t-\tau)}^{x+c(t-\tau)} G(s)\,\mathrm{d}s. \tag{3.86}$$

Introducing the *antiderivative* function via

$$\mathcal{G}(s) = \int^s G(s')\,\mathrm{d}s' \Longleftrightarrow \frac{\partial \mathcal{G}(s)}{\partial s} = G(s) \tag{3.87}$$

allows us to interpret the solution (3.86) as the superposition of two oppositely traveling waves

$$\Psi(x,t) = \frac{1}{2c} \int_0^t \Big[\mathcal{G}[x + c(t-\tau)] - \mathcal{G}[x - c(t-\tau)]\Big] \mathrm{d}\tau. \tag{3.88}$$

3.6.5 Further study

Stakgold (2000a,b) provides a thorough discussion of the wave equation and the related Helmholz equation.

3.7 Green's function method

There are many methods to solve linear PDEs either analytically or numerically. We here introduce the method of Green's functions. This method offers a very powerful conceptual and analytical framework for solving linear PDEs, and it is encountered in many places throughout physics.

The Green's function method exploits the superposition property of linear PDEs by finding a particular solution to a PDE (the Green's function) and convolving this solution with boundary and source terms to determine the solution of the original PDE. The Green's function provides a formal means to invert the linear PDE in a manner reminscient of how one inverts to find the solution to a matrix-vector problem. Our treatment is terse, offering a mere taste of a vast literature.

3.7.1 Dirac delta (generalized) function

Consider a still pool of water subject to a point forcing, or consider an electrostatic potential resulting from a point charge. The Green's function is the resulting solution to the PDEs describing these systems with the Dirac delta (generalized) function (or Dirac distribution) the mathematical expression for the point source. The delta function is written $\delta(\boldsymbol{x}-\boldsymbol{x}')$, and it is the mathematical limit of a point source at $\boldsymbol{x}=\boldsymbol{x}'$ with infinite strength. The delta function satisfies the following properties

$$\int_{\mathcal{R}} \delta(\boldsymbol{x}'-\boldsymbol{x})\,\mathrm{d}V' = \int_{\mathcal{R}} \delta(\boldsymbol{x}-\boldsymbol{x}')\,\mathrm{d}V' \qquad \text{symmetry} \tag{3.89a}$$

$$\int_{\mathcal{R}} \delta(\boldsymbol{x}'-\boldsymbol{x})\,\mathrm{d}V' = 1 \qquad \text{normalization} \tag{3.89b}$$

$$\int_{\mathcal{R}} \delta(\boldsymbol{x}'-\boldsymbol{x})\,\psi(\boldsymbol{x}')\,\mathrm{d}V' = \psi(\boldsymbol{x}) \qquad \text{sifting property,} \tag{3.89c}$$

so long as the source location, $\boldsymbol{x}$, is within the integration region: $\boldsymbol{x}\in\mathcal{R}$. Notably, these two relations imply that the delta function has dimensions of inverse volume (if considering three-dimensional space). If instead we are working on a two-dimensional area then the delta function has dimensions of inverse area, or when on a line it has dimensions of inverse length.

We find occasion to introduce a temporal delta function, $\delta(t-t')$, which turns on only at a specified time $t'=t$. It satisfies similar properties to the spatial delta function

$$\int_{-\infty}^{\infty} \delta(t'-t)\,\mathrm{d}t' = 1 \qquad \text{normalization} \tag{3.90a}$$

$$\int_{-\infty}^{\infty} \delta(t'-t)\,\psi(t')\,\mathrm{d}t' = \psi(t) \qquad \text{sifting property.} \tag{3.90b}$$

The temporal delta function has dimensions of inverse time, as required by these two properties.

3.7.2 Poisson's equation

We illustrate the Green's function method by considering the three-dimensional Poisson equation with Neumann boundary conditions

$$\nabla^2\psi = \Lambda \qquad \boldsymbol{x}\in\mathcal{R} \tag{3.91a}$$

$$\hat{\boldsymbol{n}}\cdot\nabla\psi = f \qquad \boldsymbol{x}\in\partial\mathcal{R}. \tag{3.91b}$$

The goal is to determine an integral expression for ψ in terms of the known source function, Λ, and the boundary data, f. To do so we introduce the Green's function as the solution to the Poisson equation with a delta function source and a homogeneous boundary condition

$$\nabla^2 G(\boldsymbol{x},\boldsymbol{x}') = \delta(\boldsymbol{x}-\boldsymbol{x}') \qquad \boldsymbol{x}\in\mathcal{R} \tag{3.92a}$$

$$\hat{\boldsymbol{n}}\cdot\nabla G = 0 \qquad \boldsymbol{x}\in\partial\mathcal{R}. \tag{3.92b}$$

The Green's function, $G(\boldsymbol{x},\boldsymbol{x}')$, is harmonic everywhere except at the source point, $\boldsymbol{x}=\boldsymbol{x}'$, where it equals to the inverse Laplacian acting on the delta function. Furthermore, the Green's function satisfies the homogeneous version of the boundary condition (3.91b); i.e., it does not feel the boundary function f. In some treatments, the Green's function is known as the *singularity solution* to the linear partial differential operator.

Now assume we have a method to determine the Green's function (there are many methods). How does knowing G help to find ψ? To answer this question requires the second form of Green's integral identity (2.66) using ψ and G

$$\oint_{\partial\mathcal{R}} [\psi\,\hat{\boldsymbol{n}}\cdot\nabla G - G\,\hat{\boldsymbol{n}}\cdot\nabla\psi]\,\mathrm{d}\mathcal{S}' = \int_{\mathcal{R}} [\psi\,\nabla^2 G - G\,\nabla^2\psi]\,\mathrm{d}V'. \tag{3.93}$$

Integration is taken over all positions $\boldsymbol{x}'$, as denoted for the volume integral by $\mathrm{d}V'$ and corresponding surface integral $\mathrm{d}\mathcal{S}'$. Making use of the Green's function equation (3.92a) and the sifting property (3.89c) means that the first term on the right hand side of equation (3.93) renders

$$\int_{\mathcal{R}} \psi(\boldsymbol{x}')\,\nabla^2 G(\boldsymbol{x},\boldsymbol{x}')\,\mathrm{d}V' = \int_{\mathcal{R}} \psi(\boldsymbol{x}')\,\delta(\boldsymbol{x}-\boldsymbol{x}')\,\mathrm{d}V' = \psi(\boldsymbol{x}). \tag{3.94}$$

Use of the Poission equation (3.91a) for the second term on the right hand side of equation (3.93) leads to

$$\int_{\mathcal{R}} G(\boldsymbol{x},\boldsymbol{x}')\,\nabla^2\psi(\boldsymbol{x}')\,\mathrm{d}V' = \int_{\mathcal{R}} G(\boldsymbol{x},\boldsymbol{x}')\,\Lambda(\boldsymbol{x}')\,\mathrm{d}V'. \tag{3.95}$$

The boundary conditions (3.91b) and (3.92b) allow us to write the boundary integral in the form

$$\oint_{\partial\mathcal{R}} [\psi\,\hat{\boldsymbol{n}}\cdot\nabla G - G\,\hat{\boldsymbol{n}}\cdot\nabla\psi]\,\mathrm{d}\mathcal{S}' = -\oint_{\partial\mathcal{R}} G(\boldsymbol{x},\boldsymbol{x}')\,f(\boldsymbol{x}')\,\mathrm{d}\mathcal{S}'. \tag{3.96}$$

Bringing results together leads to

$$\psi(\boldsymbol{x}) = \underbrace{\int_{\mathcal{R}} G(\boldsymbol{x},\boldsymbol{x}')\,\Lambda(\boldsymbol{x}')\,\mathrm{d}V'}_{\text{volume convolution}} - \underbrace{\oint_{\partial\mathcal{R}} G(\boldsymbol{x},\boldsymbol{x}')\,f(\boldsymbol{x}')\,\mathrm{d}\mathcal{S}.'}_{\text{boundary convolution}} \tag{3.97}$$

This solution for ψ consists of a volume convolution of the Green's function with the source Λ plus a boundary convolution of the Green's function with the boundary function f.

3.7.3 An example Green's function for the heat equation

Recall our discussion of Duhamel's superposition integral in Section 3.5.3 where we considered the one-dimensional unforced heat equation with nonhomogeneous initial conditions

$$\frac{\partial\psi}{\partial t} - \kappa\,\frac{\partial^2\psi}{\partial x^2} = 0 \qquad -\infty < x < \infty,\ t > 0 \tag{3.98a}$$

$$\psi(x,t) = f(x) \qquad -\infty < x < \infty,\ t = 0, \tag{3.98b}$$

and noted that the solution is

$$\psi(x,t) = \frac{1}{\sqrt{4\,\pi\,\kappa\,t}}\int_{-\infty}^{\infty} f(y)\,\exp\left[-\frac{(x-\xi)^2}{4\,\kappa\,t}\right]\mathrm{d}\xi. \tag{3.99}$$

Duhamel's superposition integral was then used to show that

$$\Psi(x,t) = \int_0^t \psi(x,t-\tau)\,\mathrm{d}\tau \tag{3.100a}$$

$$= \int_0^t \frac{1}{\sqrt{4\,\pi\,\kappa\,(t-\tau)}}\int_{-\infty}^{\infty} f(\xi)\,\exp\left[-\frac{(x-\xi)^2}{4\,\kappa\,(t-\tau)}\right]\mathrm{d}\tau\,\mathrm{d}\xi \tag{3.100b}$$

$$\equiv \int_0^t \int_{-\infty}^{\infty} G(x,t;\xi,\tau)\,f(\xi)\,\mathrm{d}\tau\,\mathrm{d}\xi \tag{3.100c}$$

satisfies the forced (inhomogeneous) initial value problem with zero initial condition

$$\frac{\partial \Psi}{\partial t} = \kappa \frac{\partial^2 \Psi}{\partial x^2} + f(x) \qquad -\infty < x < \infty,\ t > 0 \tag{3.101a}$$

$$\Psi(x,t) = 0 \qquad -\infty < x < \infty,\ t = 0. \tag{3.101b}$$

In equation (3.100c) we identified the Green's function for the forced heat equation with zero initial conditions

$$G(x,t;\xi,\tau) = \frac{1}{\sqrt{4\,\pi\,\kappa\,(t-\tau)}}\,\exp\left[-\frac{(x-\xi)^2}{4\,\kappa\,(t-\tau)}\right], \tag{3.102}$$

with the formula (3.100c) expressing Ψ as a space-time convolution of the source function with the Green's function.

3.7.4 Comments and further study

The Green's function allows us to invert the linear differential operator taking into account the initial and boundary conditions. Hence, if we have the Green's function then we have a direct integral solution to the initial-boundary value problem. There are many methods available to determine the Green's function (e.g., Laplace transforms, Fourier transforms, separation of variables, method of images), depending on the linear operator, the space-time domain, and the initial and/or boundary conditions. For more details, the math savvy reader may wish to penetrate the presentations offered in chapter 7 of *Morse and Feshbach* (1953); the two volume development of boundary value problems by *Stakgold* (2000a) and *Stakgold* (2000b), and/or work through the concise summary of Green's functions given in Chapter 8 of *Duchateau and Zachmann* (1986).

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

$$\frac{\mathrm{d}A}{\mathrm{d}t} = \mathcal{B}. \tag{3.103}$$

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 $\mathrm{d}A/\mathrm{d}t$ 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 (3.103) leads to

$$A(t) = A(t_0) + \int_{t_0}^{t} \mathcal{B}(s)\,\mathrm{d}s, \tag{3.104}$$

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 $\Delta T_{n,n+1} = t_{n+1} - t_n > 0$

$$\overline{A}_{n,n+1} = \frac{1}{\Delta T_{n,n+1}} \int_{t_n}^{t_{n+1}} A(t)\,\mathrm{d}t \tag{3.105a}$$

$$\overline{\mathcal{B}}_{n,n+1} = \frac{1}{\Delta T_{n,n+1}} \int_{t_n}^{t_{n+1}} \mathcal{B}(t)\,\mathrm{d}t. \tag{3.105b}$$

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 (3.104) into the time mean (3.105a) renders

$$\overline{A}_{n,n+1} - A(t_0) = \frac{1}{\Delta T_{n,n+1}} \int_{t_n}^{t_{n+1}} \left[\int_{t_0}^{t} \mathcal{B}(s)\,\mathrm{d}s\right] \mathrm{d}t, \tag{3.106}$$

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

$$\overline{A}_{p,p+1} - \overline{A}_{n,n+1} = \frac{1}{\Delta T_{p,p+1}} \int_{t_p}^{t_{p+1}} \left[\int_{t_0}^{t} \mathcal{B}(s)\,\mathrm{d}s\right] \mathrm{d}t - \frac{1}{\Delta T_{n,n+1}} \int_{t_n}^{t_{n+1}} \left[\int_{t_0}^{t} \mathcal{B}(s)\,\mathrm{d}s\right] \mathrm{d}t. \tag{3.107}$$

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 (3.107).

3.8.2 Massaging the double time integrals

The double time integrals in equation (3.107) can be massaged into a simpler form. We start by making the following decomposition and noting that $t_n \leq t \leq t_{n+1}$

$$\int_{t_n}^{t_{n+1}} \left[\int_{t_0}^{t} \mathcal{B}(s)\,\mathrm{d}s \right] \mathrm{d}t = \int_{t_n}^{t_{n+1}} \left[\int_{t_0}^{t_n} \mathcal{B}(s)\,\mathrm{d}s + \int_{t_n}^{t} \mathcal{B}(s)\,\mathrm{d}s \right] \mathrm{d}t \tag{3.108a}$$

$$= \Delta T_{n,n+1} \int_{t_0}^{t_n} \mathcal{B}(s)\,\mathrm{d}s + \int_{t_n}^{t_{n+1}} \left[\int_{t_n}^{t} \mathcal{B}(s)\,\mathrm{d}s \right] \mathrm{d}t. \tag{3.108b}$$

We are thus led to the difference

$$\overline{A}_{p,p+1} - \overline{A}_{n,n+1} = \left[\int_{t_0}^{t_p} \mathcal{B}(s)\,\mathrm{d}s - \int_{t_0}^{t_n} \mathcal{B}(s)\,\mathrm{d}s \right] \tag{3.109a}$$

$$+ \frac{1}{\Delta T_{p,p+1}} \int_{t_p}^{t_{p+1}} \left[\int_{t_p}^{t} \mathcal{B}(s)\,\mathrm{d}s \right] \mathrm{d}t - \frac{1}{\Delta T_{n,n+1}} \int_{t_n}^{t_{n+1}} \left[\int_{t_n}^{t} \mathcal{B}(s)\,\mathrm{d}s \right] \mathrm{d}t.$$

$$= \int_{t_n}^{t_p} \mathcal{B}(t)\,\mathrm{d}t + \frac{1}{\Delta T_{p,p+1}} \int_{t_p}^{t_{p+1}} \left[\int_{t_p}^{t} \mathcal{B}(s)\,\mathrm{d}s \right] \mathrm{d}t - \frac{1}{\Delta T_{n,n+1}} \int_{t_n}^{t_{n+1}} \left[\int_{t_n}^{t} \mathcal{B}(s)\,\mathrm{d}s \right] \mathrm{d}t. \tag{3.109b}$$

The double integrals in equation (3.109b) take place over triangular time domains, such as shown in Figure 3.5. In Section 3.8.3 we make use of an identity that reduces the double integral to a single integral to expose the underlying geometry of the time windowing.

Figure 3.5: *Gold region depicts the time integration domain used in one of the double integrals from equation (3.109b) for the special case of $n = 2$. Note that the gray triangular region generally leads to a distinct integral.*

3.8.3 Making use of a double integral identity

We can further massage the double integrals in equation (3.109b) through the following identity

$$\int_{t_n}^{t_{n+1}} \left[\int_{t_n}^{t} \mathcal{B}(s)\,\mathrm{d}s \right] \mathrm{d}t = \int_{t_n}^{t_{n+1}} (t_{n+1} - t)\,\mathcal{B}(t)\,\mathrm{d}t. \tag{3.110}$$

To prove this identity we make the substitution $\mathcal{B}(s) = \mathrm{d}A/\mathrm{d}s$ from equation (3.103) and then show that both sides to equation (3.110) yield the same result. For the left hand side we have

$$\int_{t_n}^{t_{n+1}} \left[\int_{t_n}^{t} \mathcal{B}(s)\,\mathrm{d}s\right] \mathrm{d}t = \int_{t_n}^{t_{n+1}} \left[\int_{t_n}^{t} \frac{\mathrm{d}A(s)}{\mathrm{d}s}\,\mathrm{d}s\right] \mathrm{d}t \tag{3.111a}$$

$$= \int_{t_n}^{t_{n+1}} \left[\int_{t_n}^{t} \mathrm{d}A(s)\right] \mathrm{d}t \tag{3.111b}$$

$$= \int_{t_n}^{t_{n+1}} [A(t) - A(t_n)]\mathrm{d}t \tag{3.111c}$$

$$= \int_{t_n}^{t_{n+1}} A(t)\,\mathrm{d}t - (t_{n+1} - t_n)\,A(t_n), \tag{3.111d}$$

whereas the right hand side is

$$\int_{t_n}^{t_{n+1}} (t_{n+1} - t)\,\mathcal{B}(t)\,\mathrm{d}t = \int_{t_n}^{t_{n+1}} (t_{n+1} - t)\,\frac{\mathrm{d}A(t)}{\mathrm{d}t})\,\mathrm{d}t \tag{3.112a}$$

$$= \int_{t_n}^{t_{n+1}} (t_{n+1} - t)\,\mathrm{d}A(t) \tag{3.112b}$$

$$= \int_{t_n}^{t_{n+1}} \mathrm{d}[A(t)\,(t_{n+1} - t)] + A(t)\,\mathrm{d}t \tag{3.112c}$$

$$= -A(t_n)\,(t_{n+1} - t_n) + \int_{t_n}^{t_{n+1}} A(t)\,\mathrm{d}t, \tag{3.112d}$$

which is identical to the left hand side given by equation (3.111d). We have thus proven the double integral formula (3.110).

The right hand side of the double integral formula (3.110) can be written

$$\int_{t_n}^{t_{n+1}} (t_{n+1} - t)\,\mathcal{B}(t)\,\mathrm{d}t = t_{n+1}\,\Delta T_{n,n+1}\,\overline{\mathcal{B}}_{n,n+1} - \int_{t_n}^{t_{n+1}} t\,\mathcal{B}(t)\,\mathrm{d}t, \tag{3.113}$$

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 (3.110) we bring equation (3.109b) into the form

$$\overline{A}_{p,p+1} - \overline{A}_{n,n+1} = \int_{t_n}^{t_p} \mathcal{B}(t)\,\mathrm{d}t + \int_{t_p}^{t_{p+1}} \frac{(t_{p+1} - t)\,\mathcal{B}(t)}{t_{p+1} - t_p}\,\mathrm{d}t - \int_{t_n}^{t_{n+1}} \frac{(t_{n+1} - t)\,\mathcal{B}(t)}{t_{n+1} - t_n}\,\mathrm{d}t \tag{3.114a}$$

$$= \int_{t_n}^{t_{n+1}} \frac{(t - t_n)\,\mathcal{B}(t)}{t_{n+1} - t_n}\,\mathrm{d}t + \int_{t_{n+1}}^{t_p} \mathcal{B}(t)\,\mathrm{d}t + \int_{t_p}^{t_{p+1}} \frac{(t_{p+1} - t)\,\mathcal{B}(t)}{t_{p+1} - t_p}\,\mathrm{d}t. \tag{3.114b}$$

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 descreasing 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. We illustrate the time windowing used for equation (3.114b) in Figure 3.6.

As a final means to write equation (3.114b), extend the middle term to the end of the time period and then subtract the extra piece and recombine to render

$$\overline{A}_{p,p+1} - \overline{A}_{n,n+1} = \int_{t_{n+1}}^{t_{p+1}} \mathcal{B}(t)\,\mathrm{d}t + \int_{t_n}^{t_{n+1}} \frac{(t - t_n)\,\mathcal{B}(t)}{t_{n+1} - t_n}\,\mathrm{d}t - \int_{t_p}^{t_{p+1}} \frac{(t - t_p)\,\mathcal{B}(t)}{t_{p+1} - t_p}\,\mathrm{d}t. \tag{3.115}$$

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 one only needs 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 (3.114b). Note that if $p = n + 1$ then there is no plateau region of unit weight; instead there are two adjacent triangular regions.*

4

Geometry of 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 whet the appetite for these notions by introducing some of the basics.

Although the curves and surfaces of geophysical fluid mechanics are commonly moving as part of the fluid flow, we are concerned in this chapter in 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 4.1). This constraint is satisfied by the surfaces of constant generalized vertical coordinates (e.g., isopycnal surfaces) considered in Chapter 8 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.

4.1 Definitions and notation **66**
4.1.1 Definitions 66
4.1.2 Notation 66
4.2 Curves in 2D Euclidan space (planar curves) **67**
4.2.1 Differential increments along the curve 67
4.2.2 Length along the curve 67
4.2.3 Curvature of a curve 69
4.3 Surfaces embedded in 3D Euclidean space **70**
4.3.1 Area on the surface 71
4.3.2 Curvature of a surface 71

4.1 Definitions and notation

The basic notions of curves and surfaces are rather intuitive. Nonetheless, it is important to be precise in our usage. For that purpose we here offer some notation and defintions.

4.1.1 Definitions

We assume the notion of a curve and surface to be self-evident, offering analytical expressions for curves in Section 4.2 and surfaces in Section 4.3. Given such, we here define some related notions used in our study of fluid mechanics.

- ORIENTED: An *orientable curve* is a curve that allows for normal and tangent directions to specify directions and sides to the curve. Likewise, and *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 single non-orientable curve. Although very interesting mathematically, we do not encounter non-orientable geometric objects in this book.

- PATH: A path is a continuous piecewise smooth oriented curve. A *simple path* is a path that does not cross itself.

- 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 leave the ocean.

4.1.2 Notation

In this chapter we write the Cartesian position of a point on a surface as

$$\boldsymbol{S} = x\,\hat{\boldsymbol{x}} + y\,\hat{\boldsymbol{y}} + \eta(x,y)\,\hat{\boldsymbol{z}} \qquad \text{position on surface,} \tag{4.1}$$

with the vertical position written as

$$z = \eta(x,y) \qquad \text{vertical position on surface.} \tag{4.2}$$

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

$$\boldsymbol{S} = x\,\hat{\boldsymbol{x}} + \eta(x)\,\hat{\boldsymbol{z}} \qquad \text{position on planar curve.} \tag{4.3}$$

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 4.1. 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 is

$$\hat{\boldsymbol{n}} = \frac{\nabla(z-\eta)}{|\nabla(z-\eta)|} = \frac{\hat{\boldsymbol{z}} - \nabla\eta}{[1+(\nabla\eta)^2\,]^{1/2}}. \tag{4.4}$$

Figure 4.2 provides an example surface along with the notation.

Figure 4.1: *Two sample curves on the $x-z$ plane. The left panel shows a curve whose outward normal, $\hat{\boldsymbol{n}}$, encounters points where $\hat{\boldsymbol{n}} \cdot \hat{\boldsymbol{z}} = 0$ and where $\hat{\boldsymbol{n}} \cdot \hat{\boldsymbol{z}}$ 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 $\hat{\boldsymbol{n}} \cdot \hat{\boldsymbol{z}} \neq 0$ everywhere, and thus where $\hat{\boldsymbol{n}} \cdot \hat{\boldsymbol{z}}$ is single signed. For curves such as the right panel, we can express the vertical position as a one-to-one 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 8.*

4.2 Curves in 2D Euclidan space (planar curves)

We here describe the geometry of a curve on the $x-z$-plane (a *planar curve*) as depicted in Figure 4.3. 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.

4.2.1 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 $\boldsymbol{S}(\varphi)$ specify the position of a point along the curve. Correspondingly, the differential increment between two infinitesimally close points on the curve is given by

$$\boldsymbol{S}(\varphi + \mathrm{d}\varphi) - \boldsymbol{S}(\varphi) = \mathrm{d}\boldsymbol{S} = \frac{\mathrm{d}\boldsymbol{S}}{\mathrm{d}\varphi}\,\mathrm{d}\varphi \equiv \boldsymbol{t}\,\mathrm{d}\varphi, \tag{4.5}$$

where

$$\boldsymbol{t} = \frac{\mathrm{d}\boldsymbol{S}}{\mathrm{d}\varphi} \tag{4.6}$$

is tangent to the curve. If $\varphi = s$ is the arc length along the curve, then $\boldsymbol{t} = \hat{\boldsymbol{t}}$ is a unt vector

$$\hat{\boldsymbol{t}} \cdot \hat{\boldsymbol{t}} = \frac{\mathrm{d}\boldsymbol{S}}{\mathrm{d}s} \cdot \frac{\mathrm{d}\boldsymbol{S}}{\mathrm{d}s} = 1. \tag{4.7}$$

Recall we made use of the arc-length along a curve in Section 2.4 when describing path integration.

4.2.2 Length along the curve

As in equation (4.3) we can represent the position of a point along the curve using Cartesian coordinates

$$\boldsymbol{S} = x\,\hat{\boldsymbol{x}} + \eta(x)\,\hat{\boldsymbol{z}}. \tag{4.8}$$

Figure 4.2: *An example of a surface considered in this chapter. The position of a point on the surface is given by the Cartesian position vector* $\boldsymbol{S} = x\,\hat{\boldsymbol{x}} + y\,\hat{\boldsymbol{y}} + \eta(x,y)\,\hat{\boldsymbol{z}}$. *The relation* $z = \eta(x,y)$ *provides a one-to-one 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.*

Hence, letting $\varphi = x$ parameterize the curve leads to the representation of the tangent direction

$$\boldsymbol{t} = \frac{\mathrm{d}\boldsymbol{S}}{\mathrm{d}x} = \hat{\boldsymbol{x}} + \frac{\mathrm{d}\eta}{\mathrm{d}x}\,\hat{\boldsymbol{z}}, \tag{4.9}$$

which has the magnitude

$$\boldsymbol{t}\cdot\boldsymbol{t} = 1 + (\mathrm{d}\eta/\mathrm{d}x)^2, \tag{4.10}$$

so that the normalized tangent vector is

$$\hat{\boldsymbol{t}} = \frac{\hat{\boldsymbol{x}} + (\mathrm{d}\eta/\mathrm{d}x)\,\hat{\boldsymbol{z}}}{[1 + (\mathrm{d}\eta/\mathrm{d}x)^2\,]^{1/2}}. \tag{4.11}$$

Likewise, the curve's normal vector is given by

$$\hat{\boldsymbol{n}} = \frac{\nabla(z-\eta)}{|\nabla(z-\eta)|} = \frac{\hat{\boldsymbol{z}} - (\mathrm{d}\eta/\mathrm{d}x)\,\hat{\boldsymbol{x}}}{[1 + (\mathrm{d}\eta/\mathrm{d}x)^2\,]^{1/2}}, \tag{4.12}$$

with orthogonality simple to show

$$\hat{\boldsymbol{t}}\cdot\hat{\boldsymbol{n}} = 0. \tag{4.13}$$

The squared length of an infinitesimal segment along the curve is given by

$$(\mathrm{d}s)^2 = \mathrm{d}\boldsymbol{S}\cdot\mathrm{d}\boldsymbol{S} = \left[\frac{\mathrm{d}\boldsymbol{S}}{\mathrm{d}x}\cdot\frac{\mathrm{d}\boldsymbol{S}}{\mathrm{d}x}\right]\mathrm{d}x\,\mathrm{d}x, \tag{4.14}$$

so that the finite length of the curve is determined by the integral

$$L = \int_0^L \mathrm{d}s = \int_{x_A}^{x_B} |\mathrm{d}\boldsymbol{S}/\mathrm{d}x|\,\mathrm{d}x = \int_{x_A}^{x_B} [1 + (\mathrm{d}\eta/\mathrm{d}x)^2]^{1/2}\,\mathrm{d}x, \tag{4.15}$$

where $x_A \le x \le x_B$ is the range over which x runs for the projection of the curve onto the x-axis (see Figure 4.3).

Figure 4.3: *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 $\boldsymbol{S} = x\,\hat{\boldsymbol{x}} + \eta(x)\,\hat{\boldsymbol{z}}$, where $z = \eta(x)$ is the vertical position of the point. The projection of the curve onto the horizontal x-axis occupies a range $x_A \le x \le x_B$. 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 $\hat{\boldsymbol{n}} = \nabla(z-\eta)/|\nabla(z-\eta)| = [\hat{\boldsymbol{z}} - (\mathrm{d}\eta/\mathrm{d}x)\,\hat{\boldsymbol{x}}]\,[1 + (\mathrm{d}\eta/\mathrm{d}x)^2\,]^{-1/2}$, and the normalized tangent direction $\hat{\boldsymbol{t}} = [\hat{\boldsymbol{x}} + (\mathrm{d}\eta/\mathrm{d}x)\,\hat{\boldsymbol{z}}]\,[1 + (\mathrm{d}\eta/\mathrm{d}x)^2\,]^{-1/2}$.*

4.2.3 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 4.4). 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 4.4. Consequently, the outward normal, $\hat{\boldsymbol{n}}$, is parallel to the $\hat{\boldsymbol{z}}$ 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

$$\eta(x) = \eta(0) + x\left[\frac{\mathrm{d}\eta}{\mathrm{d}x}\right]_{x=0} + \frac{x^2}{2}\left[\frac{\mathrm{d}^2\eta}{\mathrm{d}x^2}\right]_{x=0} + \mathcal{O}(x^3) \tag{4.16a}$$

$$= \frac{x^2}{2}\left[\frac{\mathrm{d}^2\eta}{\mathrm{d}x^2}\right]_{x=0} + \mathcal{O}(x^3). \tag{4.16b}$$

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 $\mathrm{d}\eta/\mathrm{d}x = 0$ 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 4.4. 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

$$h(x) \approx R\,[1 - 1 + \vartheta^2/2] = x^2/(2R), \tag{4.17}$$

where $\vartheta = x/R$ near the origin. For the height of a point on the curve (equation (4.16b)) to match the height along the circle, to second order accuracy, requires us to set the circle's radius to

$$\frac{1}{R} = \frac{\mathrm{d}^2\eta}{\mathrm{d}x^2}. \tag{4.18}$$

Figure 4.4: *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 $R^{-1} = \mathrm{d}^2\eta/\mathrm{d}x^2$ provides a second order accurate fit of the curvature circle to the curve at the point P.*

Equation (4.18) thus provides an expression for the radius of curvature, R, whose inverse is the curvature

$$\text{curvature} = \frac{1}{R}. \tag{4.19}$$

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 $\hat{\boldsymbol{n}}$), whereas $R < 0$ when the normal points towards the convex side (side where the curve falls away from $\hat{\boldsymbol{n}}$).

In closing this section we note that

$$-\nabla\cdot\hat{\boldsymbol{n}} = \frac{\mathrm{d}^2\eta/\mathrm{d}x^2}{[1+(\mathrm{d}\eta/\mathrm{d}x)^2\,]^{3/2}}. \tag{4.20}$$

When evaluated at the point of interest along the curve, we set $\mathrm{d}\eta/\mathrm{d}x = 0$ so that

$$-\nabla\cdot\hat{\boldsymbol{n}} = \frac{\mathrm{d}^2\eta}{\mathrm{d}x^2} = \frac{1}{R}. \tag{4.21}$$

This result supports our earlier statement that curvature measures the change in the normal direction along the curve. In fact, the identity

$$-\nabla\cdot\hat{\boldsymbol{n}} = \frac{1}{R} \tag{4.22}$$

holds for an arbitrary point along the curve since it is a coordinate invariant statement.

4.3 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 4.2. 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

$$\mathrm{d}\boldsymbol{S} = \frac{\partial\boldsymbol{S}}{\partial\varphi^1}\,\mathrm{d}\varphi_1 + \frac{\partial\boldsymbol{S}}{\partial\varphi^2}\,\mathrm{d}\varphi_2 = \boldsymbol{t}_1\,\mathrm{d}\varphi_1 + \boldsymbol{t}_2\,\mathrm{d}\varphi_2. \tag{4.23}$$

The vectors $\boldsymbol{t}_1$ and $\boldsymbol{t}_2$ 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 (4.1)

$$\boldsymbol{S} = x\,\hat{\boldsymbol{x}} + y\,\hat{\boldsymbol{y}} + \eta(x,y)\,\hat{\boldsymbol{z}} \tag{4.24}$$

brings the two tangent directions into the form

$$\boldsymbol{t}_1 = \frac{\partial \boldsymbol{S}}{\partial x} = \hat{\boldsymbol{x}} + \frac{\partial \eta}{\partial x}\,\hat{\boldsymbol{z}} \tag{4.25a}$$

$$\boldsymbol{t}_2 = \frac{\partial \boldsymbol{S}}{\partial y} = \hat{\boldsymbol{y}} + \frac{\partial \eta}{\partial y}\,\hat{\boldsymbol{z}}. \tag{4.25b}$$

Likewise, the surface normal vector is given by

$$\hat{\boldsymbol{n}} = \frac{\nabla(z-\eta)}{|\nabla(z-\eta)|} = \frac{\hat{\boldsymbol{z}} - \nabla\eta}{[1+(\nabla\eta)^2\,]^{1/2}}, \tag{4.26}$$

and it is straightforward to show orthogonality with the two tangent vectors

$$\hat{\boldsymbol{t}}_1 \cdot \hat{\boldsymbol{n}} = \hat{\boldsymbol{t}}_2 \cdot \hat{\boldsymbol{n}} = 0. \tag{4.27}$$

4.3.1 Area on the surface

Recall from Section 1.4.4 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 $\mathrm{d}\varphi_1$ and $\mathrm{d}\varphi_2$ is given by

$$\mathrm{d}\mathcal{S} = \left| \frac{\partial \boldsymbol{S}}{\partial \varphi^1} \wedge \frac{\partial \boldsymbol{S}}{\partial \varphi^2} \right| \mathrm{d}\varphi_1\,\mathrm{d}\varphi_2. \tag{4.28}$$

Making use of Cartesian coordinates brings the area element to

$$\mathrm{d}\mathcal{S} = [1+(\nabla\eta)^2\,]^{1/2}\,\mathrm{d}x\,\mathrm{d}y = [1+(\nabla\eta)^2\,]^{1/2}\,\mathrm{d}A, \tag{4.29}$$

where

$$\mathrm{d}A = \mathrm{d}x\,\mathrm{d}y \tag{4.30}$$

is the area of the surface projected onto the horizontal plane. Hence, the area of a finite region is given by the integral

$$\mathcal{S} = \int \mathrm{d}\mathcal{S} = \int [1+(\nabla\eta)^2\,]^{1/2}\,\mathrm{d}x\,\mathrm{d}y, \tag{4.31}$$

where the second integral extends over the region defined by the projection of the surface onto the horizontal (see Figure 4.2).

4.3.2 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 of a curve discussed in Section 4.2.3. The method for developing the curvature is analogous to that used for a curve, yet with a bit more mathematics needed to allow for the extra dimension. Figure 4.5 depicts the situation.

Figure 4.5: *Depicting the elements needed to construct the curvature of a surface at an arbitrary point, P. The local normal direction is given by $\hat{\boldsymbol{n}}$, along with the two tangent vectors $\boldsymbol{t}_1$ and $\boldsymbol{t}_2$. 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.*

Let $\boldsymbol{x} = (x_1, x_2) = (x, y)$ 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

$$\eta \approx \frac{1}{2}\, x_m\, \mathbb{K}_{mn}\, x_n, \tag{4.32}$$

where $\mathbb{K}$ is the second order tensor of second partial derivatives evaluated at the point

$$\mathbb{K} = \begin{bmatrix} \dfrac{\partial^2 \eta}{\partial x_1^2} & \dfrac{\partial^2 \eta}{\partial x_1 \partial x_2} \\ \dfrac{\partial^2 \eta}{\partial x_1 \partial x_2} & \dfrac{\partial^2 \eta}{\partial x_2^2} \end{bmatrix}. \tag{4.33}$$

As a symmetric tensor, $\mathbb{K}$ is diagonalizable and it has two eigenvalues, R_1^{-1} and R_2^{-1}, along with its associated eigenvectors, $\boldsymbol{e}_1$ and $\boldsymbol{e}_2$. The quadratic form (4.32) can thus be written as

$$\eta \approx \frac{1}{2}\, R_1^{-1}\, (\boldsymbol{x} \cdot \boldsymbol{e}_1)^2 + \frac{1}{2}\, R_2^{-1}\, (\boldsymbol{x} \cdot \boldsymbol{e}_2)^2. \tag{4.34}$$

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 $\hat{\boldsymbol{n}} - \boldsymbol{e}_1$ and $\hat{\boldsymbol{n}} - \boldsymbol{e}_2$ planes. If the radius of curvature R_m is positive, then the surface curves towards $\hat{\boldsymbol{n}}$ along the $\hat{\boldsymbol{n}} - \boldsymbol{e}_i$ 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 $\mathbb{K}$ that commonly arise in applications.

- $\mathrm{Tr}(\mathbb{K}) = R_1^{-1} + R_2^{-1}$, which is twice the mean curvature for the surface. With the normal vector given by equation (4.4), one can show that

$$-\nabla \cdot \hat{\boldsymbol{n}} = \frac{\nabla^2 \eta}{[1 + (\nabla \eta)^2\,]^{3/2}}. \tag{4.35}$$

A bit of algebra leads us to conclude that

$$-\nabla \cdot \hat{\boldsymbol{n}} = \frac{1}{R_1} + \frac{1}{R_2}, \tag{4.36}$$

for any point along the surface, thus generalizing the result (4.21) found for a curve.

- $\det(\mathbb{K}) = 1/(R_1\, R_2)$ is known as the *Gaussian curvature*, which is the product of the two curvatures.

5

General tensors in brief

In physics we aim to uncover objective statements about how physical systems operate. That aim is supported by mathematical tools that reflect the underlying objective nature of physical relationships, while also allowing for the quantitative realization of those relationships in particular situations. In this chapter we provide a conceptual platform for general tensor analysis, with details of the machinery presented in Chapter 6. This brief chapter offers an overview sufficient to appreciate the power of general tensors yet without going into details.

READER'S GUIDE TO THIS CHAPTER

This chapter builds from the Cartesian tensor analysis of Chapter 1 and introduces mathematical concepts that are central to this book.

5.1 General tensors and geophysical fluid mechanics **76**
5.2 Covariance of physical relations **76**
5.2.1 Tensor operations retaining covariance 76
5.2.2 Comments 77
5.3 Points, trajectories, and coordinates **77**
5.3.1 Time as a parameter and time as a coordinate 78
5.3.2 The importance of index placement 78
5.4 Example coordinate descriptions **78**
5.4.1 Eulerian coordinates 78
5.4.2 Isopycnal coordinates 79
5.4.3 Lagrangian or material coordinates 79
5.4.4 Tracer coordinates 80
5.5 The velocity vector and basis vectors **80**
5.5.1 Coordinate representation 80
5.5.2 Basis vectors 81
5.6 Notational conventions **81**
5.6.1 Placement of tensor labels 81
5.6.2 Einstein summation convention 81
5.6.3 The boldface notation 81
5.6.4 Space plus time 82

5.1 General tensors and geophysical fluid mechanics

The Cartesian tensors described in Chapters 1 and 2 are sufficient for many areas of geophysical fluid mechanics. However, there are a number of applications where general tensors prove of great use to ensure that the physics shines through the maths. We thus go beyond Cartesian tensor analysis to enable a more versatile, and precise, mathematical framework for the study of geophysical fluid mechanics.

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 fluid systems are embedded within Euclidean space, $\mathbb{R}^3$. Euclidean space is the familiar space of classical Newtonian mechanics. Notably, $\mathbb{R}^3$ is flat in that it has zero intrinsic curvature. 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 geometry local to the surface inherits features from the background flat Euclidean space such as how to measure distance between points. A further simplification arises since we make use of universal Newtonian time. In summary, our mathematical needs are simpler than the general relativist.

5.2 Covariance of 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. This is the essence of *covariance*. In practice, covariance means that a mathematical expression of a physical relation is form invariant 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, in which they exemplify covariance, then the equations are form invariant. That is, covariance 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.

5.2.1 Tensor operations retaining covariance

Extending a mathematical equation to all coordinates requires the equations to respect covariance. Operationally, general covariance means that all tensor indices are properly matched and each derivative is covariant as specified in Section 6.9. In chapter 6, we provide the details needed to understand general covariance. In this chapter we outline the procedure. The elegance and power rendered by general covariance is the key reason that tensor analysis is ubiquitous in theoretical physics.

To ensure an equation respects covariance requires us to understand certain properties of tensors and operations with tensors that produces components of new tensors. We here summarize the specific properties characterizing covariance (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* and is denoted by the operator $\otimes$.

   (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 6.9 and 6.10) 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 general covariance.

5.2.2 Comments

The remainder of this chapter, as well as Chapter 6, provide the needed details for supporting the above notions of covariance. Even without penetrating these details, the reader should be able to appreciate why covariance is so central to physics.

5.3 Points, trajectories, and coordinates

Consider a point in space, $\mathscr{P}$, 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 12.2). 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 4.2.1). We choose the time measured by an observer on the trajectory for this parameter, in which case the trajectory is written $\mathscr{P}(\tau)$.

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 underlyng physics and geometry.

5.3.1 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 $\tau = t$ 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 5.3.2), then $\partial/\partial t$ can be a mixture of Lagrangian and Eulerian or perhaps neither.

5.3.2 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

$$\xi^\alpha = (\xi^0, \xi^1, \xi^2, \xi^3). \tag{5.1}$$

We emphasize that the index is a label for the coordinate; it is not an exponent. 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

$$\xi^\alpha = (\xi^0, \xi^1, \xi^2, \xi^3) = (\xi^0, \xi^a) = (\xi^0, \boldsymbol{\xi}). \tag{5.2}$$

5.4 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, $a = 1, 2, 3$.

5.4.1 Eulerian coordinates

The Cartesian coordinates for a point are written

$$\xi^a = (x, y, z) \qquad \text{Cartesian}, \tag{5.3}$$

whereas for spherical coordinates we write (see Figure 7.1)

$$\xi^a = (r, \lambda, \phi) \qquad \text{spherical} \tag{5.4}$$

and polar cylindrical coordinates

$$\xi^a = (r, \lambda, z) \qquad \text{cylindrical.} \tag{5.5}$$

These coordinates identify fixed positions in space. We can use these *Eulerian* coordinates to mark the trajectory $\mathscr{P}(\tau)$ as it crosses the spatial point ξ^a at time t. We provide a more complete discussion of Eulerian coordinates in Section 12.3.

As shown in Section 5.5, 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.

5.4.2 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

$$\xi^a = (x, y, b) \qquad \text{with} \qquad b = b(x, y, z, t) \qquad \text{isopycnal coordinates,} \tag{5.6}$$

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.

Isopycnal coordinate is 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.

5.4.3 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 one-to-one 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 12.3 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 an perfect fluid parcel whereas its horizontal coordinates are Eulerian.

5.4.4 Tracer coordinates

Consider a triplet of linearly independent tracer concentrations $C^a = C^a(x, y, z, t)$. Linear independence means that for 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

$$\xi^a = (C^1, C^2, C^3). \tag{5.7}$$

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.

5.5 The velocity vector and basis vectors

Consider two points in space that sit along a particular trajectory separated by an infinitesimal time increment, $\mathrm{d}\tau$. The velocity vector for this trajectory is defined by

$$\vec{v}(\tau) = \lim_{\Delta\tau\to 0} \frac{\mathscr{P}(\tau + \Delta\tau/2) - \mathscr{P}(\tau - \Delta\tau/2)}{\Delta\tau} = \frac{\mathrm{d}\mathscr{P}(\tau)}{\mathrm{d}\tau}. \tag{5.8}$$

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; i.e., it is a geometric object and this fact is emphasized by exposing the arrow on the symbol for the velocity.

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.

5.5.1 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

$$\mathscr{P}(\tau) = \mathscr{P}[\xi^a(\tau)], \tag{5.9}$$

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

$$\vec{v}(\tau) = \frac{\mathrm{d}\mathscr{P}(\tau)}{\mathrm{d}\tau} = \frac{\partial\mathscr{P}}{\partial\xi^a}\frac{\mathrm{d}\xi^a}{\mathrm{d}\tau} \equiv \vec{e}_a\, v^a. \tag{5.10}$$

The expansion coefficients

$$v^a = \frac{\mathrm{d}\xi^a}{\mathrm{d}\tau} \tag{5.11}$$

provide a representation the velocity vector $\vec{v}(\tau)$ within the coordinate system ξ^a.

5.5.2 Basis vectors

For each number v^a there is a corresponding basis vector $\vec{e}_a$ defined by

$$\vec{e}_a = \frac{\partial \mathscr{P}}{\partial \xi^a}. \tag{5.12}$$

The basis vectors are generally a function of space and time, with Cartesian coordinates a notable exception whereby they are space-time constants.

5.6 Notational conventions

We here introduce notational conventions that help to simplify many of the manipulations ("index gymnastics") encountered with general tensors.

5.6.1 Placement of tensor labels

As indicated in Section 5.3.2, the placement of tensor labels has specific meaning with general tensor analysis. It is therefore critical to maintain proper usage to ensure "conservation of labels" across an equals sign. As a first example of this usage, notice how the basis vectors in equation (5.12) 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.

5.6.2 Einstein summation convention

For general tensors, the Einstein summation convention assumes that labels are summed over their range when a lower (covariant) label matches an upper (contra-variant) label. In this way we have an arbitrary vector in space written as

$$\vec{F}(\tau) = \sum_{a=1}^{3} \vec{e}_a \, F^a = \vec{e}_a \, F^a. \tag{5.13}$$

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 6.2, where we show that the contraction of a vector and a one-form renders a scalar.

5.6.3 The boldface notation

In Cartesian tensors we have no need to distinguish $\vec{F}$ from the boldface $\boldsymbol{F}$. One means to extend this notation to general tensors is to organize the velocity vector representation F^a into an ordered list (F^1, F^2, F^3) and to use the boldface notation

$$\boldsymbol{F} = (F^1, F^2, F^3). \tag{5.14}$$

Likewise, we can organize the basis vectors according to

$$\vec{\boldsymbol{e}} = (\vec{e}_1, \vec{e}_2, \vec{e}_3) \tag{5.15}$$

and the coordinates as

$$\boldsymbol{\xi} = (\xi^1, \xi^2, \xi^3). \tag{5.16}$$

With this notation the velocity vector representation from equation (5.10) takes on the form

$$\vec{v} = \vec{e}_a \, v^a = \vec{\boldsymbol{e}} \cdot \boldsymbol{v}. \tag{5.17}$$

Likewise, a trajectory can be represented in terms of a chosen set of coordinates according to

$$\mathscr{P} = \vec{e}_a \, \xi^a = \vec{\boldsymbol{e}} \cdot \boldsymbol{\xi}. \tag{5.18}$$

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

$$\boldsymbol{F} = \vec{F} = \vec{e}_a \, F^a. \tag{5.19}$$

Though corrupting the convention in equation (5.17), 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.

5.6.4 Space plus time

As introduced in Section 5.3.2, 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 (5.2) used for coordinates, we make use of the following index notation and ordered list for contravariant components of 4-vectors

$$F^\alpha = (F^0, F^1, F^2, F^3) = (F^0, F^a) = (F^0, \boldsymbol{F}). \tag{5.20}$$

In Cartesian coordinates, the time component of the velocity 4-vector is unity

$$v^\alpha = (1, v^1, v^2, v^3) = (1, v^a) = (1, \boldsymbol{v}) \qquad \text{velocity in Cartesian coordinates.} \tag{5.21}$$

In subsequent chapters, we develop the formalism necessary to represent these components in arbitrary coordinates.

6

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, since the time derivative cares about the motion of space coordinates.

READER'S GUIDE TO THIS CHAPTER

This chapter is the most specialized of the book's math chapters. It is necessary for understanding the mathematics of generalized vertical coordinates in Chapters 8, 17, and 38. Otherwise, it can be skimmed or returned to later if the need arises. Material here is a streamlined and updated version of Chapters 20 and 21 from *Griffies* (2004). Other resources include the treatment of tensors for fluid mechanics given by *Aris* (1962) and the physics treatment by *Thorne and Blandford* (2017).

6.1 The metric tensor and coordinate transformations **84**
6.1.1 Cartesian coordinates in Euclidean space 84
6.1.2 The metric as a symmetric second order tensor 85
6.1.3 Coordinate representation of the metric tensor 85
6.1.4 Transforming the metric tensor representation 86
6.1.5 Basis vectors 86
6.1.6 Finite distance between points 87
6.2 One-forms **87**
6.2.1 Coordinate representation of a one-form 87
6.2.2 Basis one-forms and the bi-orthogonality relation 87
6.2.3 Metric as a mapping between vectors and one-forms 88
6.2.4 Transformation of the coordinate representation 89
6.2.5 Arbitrary coordinate representation of inverse metric 89
6.3 Scalar product **89**
6.4 The volume element and Jacobian of transformation **90**
6.4.1 Jacobian of transformation 90
6.4.2 Relating the Jacobian to the determinant of the metric 90
6.5 The permutation symbol and the determinant **91**
6.5.1 Connecting the permutation symbol to the determinant 91
6.5.2 Further identities satisfied by the determinant 92
6.6 The Levi-Civita tensor and the volume element **92**

6.6.1 General coordinate representation of the Levi-Civita tensor 92
6.6.2 The Levi-Civita tensor and the volume element 93
6.7 Vector cross product **93**
6.8 Coordinate transformation of partial derivatives **93**
6.8.1 The space-time transformation matrix 94
6.8.2 Determinant of the transformation matrix 94
6.8.3 Multiplying the transformation matrix and its inverse 95
6.8.4 Transformation of space and time partial derivatives 95
6.9 Covariant derivative acting on a scalar **96**
6.10 Covariant derivative acting on a vector **96**
6.10.1 Derivative of a vector 96
6.10.2 An alternative derivation 97
6.10.3 Christoffel symbols are not components of a tensor 97
6.11 Covariant derivative of a one-form **97**
6.12 Covariant derivative of the metric tensor **98**
6.13 Christoffel symbols in terms of the metric **98**
6.14 Covariant divergence of a vector **98**
6.14.1 Contraction of the Christoffel symbols 98
6.14.2 Exponential of the determinant 99
6.15 Covariant Laplacian of a scalar **99**
6.16 Covariant curl of a vector **99**
6.17 Gauss's divergence theorem **100**
6.18 Stokes' theorem **100**

6.1 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.

6.1.1 Cartesian coordinates in Euclidean space

Consider a Cartesian coordinate representation for the spatial position of two points, with point $\mathscr{P}$ having space coordinates $\xi^a = x^a$ and the other point $\mathscr{Q}$ an infinitesimal distance away at $x^a + \mathrm{d}x^a$. Furthermore, let

$$\mathrm{d}\boldsymbol{x} = \mathrm{d}x^a\,\vec{e}_a \tag{6.1}$$

be the infinitesimal space vector pointing from $\mathscr{P}$ to $\mathscr{Q}$. 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)

$$\mathrm{d}s^2 = \mathrm{d}\boldsymbol{x}\cdot\mathrm{d}\boldsymbol{x} = \vec{e}_a\cdot\vec{e}_b\,\mathrm{d}x^a\mathrm{d}x^b = \delta_{ab}\,\mathrm{d}x^a\mathrm{d}x^b. \tag{6.2}$$

In this expression,

$$(\mathrm{d}s)^2 \equiv \mathrm{d}s^2 \tag{6.3}$$

is the squared infinitesimal arc-length separating the two points. The Kronecker symbol, δ_{ab}, is symmetric

$$\delta_{ab} = \delta_{ba}, \tag{6.4}$$

and vanishes when $a \neq b$ and is unity when $a = b$

$$\delta_{ab} = \begin{cases} 0 & \text{if } a \neq b \\ 1 & \text{if } a = b. \end{cases} \tag{6.5}$$

That is, the Kronecker symbol is a representation of the identity tensor.

6.1.2 The metric as a symmetric second order tensor

As defined by equation (6.2), 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

$$\text{distance}(\vec{P}, \vec{Q}) = \sqrt{\mathcal{G}(\vec{P}, \vec{Q})}. \tag{6.6}$$

Here, $\mathcal{G}$ is the metric tensor with coordinate representation g_{ab} and $\vec{P}$, $\vec{Q}$ are inifinitesimally close vectors with coordinate representations

$$\vec{P} = \xi^a \, \vec{e}_a \qquad \vec{Q} = \vec{P} + \mathrm{d}\xi^a \, \vec{e}_a. \tag{6.7}$$

Equation (6.6) indicates that the metric tensor takes two vectors as argument and produces a scalar. Furthermore, since

$$\text{distance}(\vec{P}, \vec{Q}) = \text{distance}(\vec{Q}, \vec{P}) \geq 0, \tag{6.8}$$

the metric tensor is a symmetric and positive tensor that produces zero only when $\vec{P} = \vec{Q}$.

6.1.3 Coordinate representation of the metric tensor

Given the geometric expression (6.6) for the metric, we determine its representation in an arbitrary coordinate system by considering the squared distance between the coordinate basis vectors

$$\text{distance}(\vec{e}_a, \vec{e}_b) = \sqrt{\mathcal{G}(\vec{e}_a, \vec{e}_b)}. \tag{6.9}$$

This relation determines the coordinate components of the metric tensor

$$\mathcal{G}(\vec{e}_a, \vec{e}_b) \equiv g_{ab}. \tag{6.10}$$

Furthermore, for a manifold embedded in Euclidean space this relation is written

$$g_{ab} = \vec{e}_a \cdot \vec{e}_b. \tag{6.11}$$

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 $a = b$.

6.1.4 Transforming the metric tensor representation

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 $\xi^{\overline{a}}$. Use of the chain rule along with some index gymnastics leads to the equivalent expression for the squared infinitesimal length between two points

$$\mathrm{d}s^2 = \delta_{ab}\,\mathrm{d}\xi^a\,\mathrm{d}\xi^b = \delta_{ab}\,\frac{\partial\xi^a}{\partial\xi^{\overline{a}}}\frac{\partial\xi^b}{\partial\xi^{\overline{b}}}\,\mathrm{d}\xi^{\overline{a}}\,\mathrm{d}\xi^{\overline{b}} \equiv \delta_{ab}\Lambda^a_{\ \overline{a}}\,\Lambda^b_{\ \overline{b}}\,\mathrm{d}\xi^{\overline{a}}\,\mathrm{d}\xi^{\overline{b}} \equiv g_{\overline{a}\overline{b}}\,\mathrm{d}\xi^{\overline{a}}\,\mathrm{d}\xi^{\overline{b}}, \tag{6.12}$$

where

$$g_{\overline{a}\overline{b}} = \delta_{ab}\Lambda^a_{\ \overline{a}}\,\Lambda^b_{\ \overline{b}} \tag{6.13}$$

defines the components to the metric tensor as represented by the new set of coordinates $\xi^{\overline{a}}$. We also introduced elements to the *transformation matrix*

$$\Lambda^a_{\ \overline{a}} = \frac{\partial\xi^a}{\partial\xi^{\overline{a}}}. \tag{6.14}$$

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 $\Lambda^a_{\ \overline{a}}$ 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 $\overline{a}$

$$\Lambda^a_{\overline{a}} = \begin{bmatrix} (\partial\xi^1/\partial\xi^{\overline{1}})_{\overline{2},\overline{3}} & (\partial\xi^1/\partial\xi^{\overline{2}})_{\overline{1},\overline{3}} & (\partial\xi^1/\partial\xi^{\overline{3}})_{\overline{1},\overline{2}} \\ (\partial\xi^2/\partial\xi^{\overline{1}})_{\overline{2},\overline{3}} & (\partial\xi^2/\partial\xi^{\overline{2}})_{\overline{1},\overline{3}} & (\partial\xi^2/\partial\xi^{\overline{3}})_{\overline{1},\overline{2}} \\ (\partial\xi^3/\partial\xi^{\overline{1}})_{\overline{2},\overline{3}} & (\partial\xi^3/\partial\xi^{\overline{2}})_{\overline{1},\overline{3}} & (\partial\xi^3/\partial\xi^{\overline{3}})_{\overline{1},\overline{2}} \end{bmatrix}. \tag{6.15}$$

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

$$\Lambda^a_{\overline{a}} = \begin{bmatrix} (\partial\boldsymbol{\xi}/\partial\xi^{\overline{1}})_{\overline{2},\overline{3}} & (\partial\boldsymbol{\xi}/\partial\xi^{\overline{2}})_{\overline{1},\overline{3}} & (\partial\boldsymbol{\xi}/\partial\xi^{\overline{3}})_{\overline{1},\overline{2}} \end{bmatrix}. \tag{6.16}$$

Use of this transformation matrix leads to the arbitrary coordinate representation of the metric tensor

$$g_{\overline{a}\overline{b}} = \vec{e}_{\overline{a}}\cdot\vec{e}_{\overline{b}} = \begin{bmatrix} \frac{\partial\boldsymbol{\xi}}{\partial\xi^{\overline{1}}}\cdot\frac{\partial\boldsymbol{\xi}}{\partial\xi^{\overline{1}}} & \frac{\partial\boldsymbol{\xi}}{\partial\xi^{\overline{1}}}\cdot\frac{\partial\boldsymbol{\xi}}{\partial\xi^{\overline{2}}} & \frac{\partial\boldsymbol{\xi}}{\partial\xi^{\overline{1}}}\cdot\frac{\partial\boldsymbol{\xi}}{\partial\xi^{\overline{3}}} \\ \frac{\partial\boldsymbol{\xi}}{\partial\xi^{\overline{2}}}\cdot\frac{\partial\boldsymbol{\xi}}{\partial\xi^{\overline{1}}} & \frac{\partial\boldsymbol{\xi}}{\partial\xi^{\overline{2}}}\cdot\frac{\partial\boldsymbol{\xi}}{\partial\xi^{\overline{2}}} & \frac{\partial\boldsymbol{\xi}}{\partial\xi^{\overline{2}}}\cdot\frac{\partial\boldsymbol{\xi}}{\partial\xi^{\overline{3}}} \\ \frac{\partial\boldsymbol{\xi}}{\partial\xi^{\overline{3}}}\cdot\frac{\partial\boldsymbol{\xi}}{\partial\xi^{\overline{1}}} & \frac{\partial\boldsymbol{\xi}}{\partial\xi^{\overline{3}}}\cdot\frac{\partial\boldsymbol{\xi}}{\partial\xi^{\overline{2}}} & \frac{\partial\boldsymbol{\xi}}{\partial\xi^{\overline{3}}}\cdot\frac{\partial\boldsymbol{\xi}}{\partial\xi^{\overline{3}}} \end{bmatrix}. \tag{6.17}$$

6.1.5 Basis vectors

We transform the basis vectors from Cartesian into arbitrary coordinates through the transformation

$$\vec{e}_{\overline{a}} = \Lambda^a_{\ \overline{a}}\,\vec{e}_a. \tag{6.18}$$

Use of the transformation matrix (6.16) renders the arbitrary coordinate basis vectors

$$\vec{e}_{\overline{1}} = \frac{\partial \boldsymbol{\xi}}{\partial \xi^{\overline{1}}} \qquad \text{and} \qquad \vec{e}_{\overline{2}} = \frac{\partial \boldsymbol{\xi}}{\partial \xi^{\overline{2}}} \qquad \text{and} \qquad \vec{e}_{\overline{3}} = \frac{\partial \boldsymbol{\xi}}{\partial \xi^{\overline{3}}}. \tag{6.19}$$

6.1.6 Finite distance between points

Once the metric is determined, the distance along a curve between two finitely separated points is given by integration

$$L = \int \sqrt{\mathrm{d}s^2} = \int_{\varphi_1}^{\varphi_2} \left| g_{ab} \frac{\mathrm{d}\xi^a}{\mathrm{d}\varphi} \frac{\mathrm{d}\xi^b}{\mathrm{d}\varphi} \right|^{1/2} \mathrm{d}\varphi, \tag{6.20}$$

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.

6.2 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 (6.6))

$$\text{distance}(\vec{A}, \vec{B}) = \sqrt{\mathcal{G}(\vec{A}, \vec{B})}. \tag{6.21}$$

What if the metric only eats one vector? The resulting geometric object is known as a one-form

$$\widetilde{A} \equiv \mathcal{G}(\vec{A}, \ ), \tag{6.22}$$

with the tilde used to distinguish a one-form from a vector.

6.2.1 Coordinate representation of a one-form

We can determine the coordinate representation of a one-form by eating a basis vector

$$\begin{aligned} \widetilde{A}(\vec{e}_b) &= \mathcal{G}(\vec{A}, \vec{e}_b) && \text{(6.23a)} \\ &= \mathcal{G}(A^a \, \vec{e}_a, \vec{e}_b) && \text{(6.23b)} \\ &= \mathcal{G}(\vec{e}_a, \vec{e}_b) \, A^a && \text{(6.23c)} \\ &= g_{ab} A^a. && \text{(6.23d)} \end{aligned}$$

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 $\widetilde{A}$ in terms of its *dual* vector $\vec{A}$ and the metric tensor

$$A_b = g_{ab} A^a. \tag{6.24}$$

6.2.2 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, $\widetilde{e}^a$, are defined through the bi-orthogonality relation

$$\mathcal{G}(\widetilde{e}^a, \vec{e}_b) = \widetilde{e}^a \cdot \vec{e}_b = \delta^a_{\ b}, \tag{6.25}$$

where

$$\delta^a_{\ b} = g^{ac}\, g_{cb} \tag{6.26}$$

are components to the Kronecker delta tensor, taking the value of unity when $a = b$ and zero otherwise

$$\delta^a_{\ b} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}. \tag{6.27}$$

It is only for Cartesian coordinates that we have

$$\delta^a_{\ c} = g^{ab}\, \delta_{bc} \qquad \text{Cartesian coordinates,} \tag{6.28}$$

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

$$\widetilde{e}^{\overline{a}} = \Lambda^{\overline{a}}_{\ a}\, \widetilde{e}^a, \tag{6.29}$$

which renders

$$\widetilde{e}^{\overline{1}} = \hat{\boldsymbol{x}}\, \frac{\partial \xi^{\overline{1}}}{\partial x} + \hat{\boldsymbol{y}}\, \frac{\partial \xi^{\overline{1}}}{\partial y} + \hat{\boldsymbol{z}}\, \frac{\partial \xi^{\overline{1}}}{\partial z} = \nabla \xi^{\overline{1}} \tag{6.30a}$$

$$\widetilde{e}^{\overline{2}} = \hat{\boldsymbol{x}}\, \frac{\partial \xi^{\overline{2}}}{\partial x} + \hat{\boldsymbol{y}}\, \frac{\partial \xi^{\overline{2}}}{\partial y} + \hat{\boldsymbol{z}}\, \frac{\partial \xi^{\overline{2}}}{\partial z} = \nabla \xi^{\overline{2}} \tag{6.30b}$$

$$\widetilde{e}^{\overline{3}} = \hat{\boldsymbol{x}}\, \frac{\partial \xi^{\overline{3}}}{\partial x} + \hat{\boldsymbol{y}}\, \frac{\partial \xi^{\overline{3}}}{\partial y} + \hat{\boldsymbol{z}}\, \frac{\partial \xi^{\overline{3}}}{\partial z} = \nabla \xi^{\overline{3}}. \tag{6.30c}$$

In Section 6.8.3 we verify that the basis one-forms satisfy the orthogonality relation (6.25) with the basis vectors

$$\widetilde{e}^{\overline{a}} \cdot \vec{e}_{\overline{b}} = \delta^{\overline{a}}_{\ \overline{b}}. \tag{6.31}$$

6.2.3 Metric as a mapping between vectors and one-forms

We can contract the expression (6.24) with components of the inverse metric tensor, g^{ab}, to render

$$g^{ab} A_b = g^{ab}\, g_{bc} A^c = \delta^a_{\ c} A^c = A^a. \tag{6.32}$$

This identity, as well as equation (6.24), show that the metric provides a map between coordinate representations of one-forms and vectors.

In general, to every vector $\vec{A}$ there is a corresponding one-form $\widetilde{A}$. We say that the one-forms and vectors are *dual*, with mapping between one-forms and vectors rendered by the metric tensor. In Cartesian tensor analysis, duality between one-forms and vectors becomes 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 (6.24) 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.

6.2.4 Transformation of the coordinate representation

The transformation matrix (6.14) 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

$$\vec{F}(\widetilde{e}^a) = F^a. \tag{6.33}$$

Now consider another coordinate system with basis one-forms $\widetilde{e}^{\overline{a}}$, so that the vector has a representation

$$\vec{F}(\widetilde{e}^{\overline{a}}) = F^{\overline{a}}. \tag{6.34}$$

Transforming the basis one-form using the transformation matrix leads to

$$F^{\overline{a}} = \vec{F}(\widetilde{e}^{\overline{a}}) = \vec{F}(\Lambda^{\overline{a}}{}_a\, \widetilde{e}^a) = \Lambda^{\overline{a}}{}_a\, \vec{F}(\widetilde{e}^a) = \Lambda^{\overline{a}}{}_a\, F^a. \tag{6.35}$$

Transformation of an arbitrary one-form representation takes place with the inverse transformation matrix

$$F_{\overline{a}} = \widetilde{F}(\vec{e}_{\overline{a}}) = \widetilde{F}(\Lambda^a{}_{\overline{a}}\, \vec{e}_a) = \Lambda^a{}_{\overline{a}}\, \widetilde{F}(\vec{e}_a) = \Lambda^a{}_{\overline{a}}\, F_a. \tag{6.36}$$

6.2.5 Arbitrary coordinate representation of inverse metric

The inverse metric tensor has an arbitrary coordinate representation given by

$$g^{\overline{a}\overline{b}} = \widetilde{e}^{\overline{a}} \cdot \widetilde{e}^{\overline{b}} = \begin{bmatrix} \nabla\xi^{\overline{1}} \cdot \nabla\xi^{\overline{1}} & \nabla\xi^{\overline{1}} \cdot \nabla\xi^{\overline{2}} & \nabla\xi^{\overline{1}} \cdot \nabla\xi^{\overline{3}} \\ \nabla\xi^{\overline{2}} \cdot \nabla\xi^{\overline{1}} & \nabla\xi^{\overline{2}} \cdot \nabla\xi^{\overline{2}} & \nabla\xi^{\overline{2}} \cdot \nabla\xi^{\overline{3}} \\ \nabla\xi^{\overline{3}} \cdot \nabla\xi^{\overline{1}} & \nabla\xi^{\overline{3}} \cdot \nabla\xi^{\overline{2}} & \nabla\xi^{\overline{3}} \cdot \nabla\xi^{\overline{3}} \end{bmatrix}. \tag{6.37}$$

Proof that $g^{\overline{a}\overline{b}}\, g_{\overline{b}\overline{c}} = \delta^{\overline{a}}_{\overline{c}}$ requires use of the chain rule relations derived in Section 6.8.3.

6.3 Scalar product

In Section 1.3.2 we defined the scalar product between two Cartesian vectors. The natural generalization is given by

$$\vec{P} \cdot \vec{Q} = P^a\, Q^b\, \vec{e}_a \cdot \vec{e}_b = P^a\, Q^b\, g_{ab} = P^a\, Q_a = P_b\, Q^b, \tag{6.38}$$

where the second equality made use of the metric tensor coordinate representation given by equation (6.11). We can conceive of the scalar product in a somewhat more general manner by recalling that a one-form operates on a vector, $\widetilde{P}(\vec{Q})$. Conversely, a vector operates on a one-form, $\vec{Q}(\widetilde{P})$. Exposing components leads to

$$\widetilde{P}(\vec{Q}) = \widetilde{P}(Q^a\, \vec{e}_a) = Q^a\, \widetilde{P}(\vec{e}_a) = Q^a\, P_a, \tag{6.39}$$

which equals to

$$\vec{Q}(\widetilde{P}) = \vec{Q}(P_a\, \widetilde{e}^a) = P_a\, \vec{Q}(\widetilde{e}^a) = P_a\, Q^a. \tag{6.40}$$

The scalar product is invariant to coordinate changes, as seen through

$$\vec{Q}(\widetilde{P}) = \vec{Q}(P_a\, \widetilde{e}^a) = P_a\, Q^a = \vec{Q}(P_{\overline{a}}\, \widetilde{e}^{\overline{a}}) = P_{\overline{a}}\, Q^{\overline{a}}. \tag{6.41}$$

The invariance is also revealed by working just with the coordinate representations and introducing the transformation matrix elements

$$P_a\, Q^a = (\Lambda^{\overline{a}}{}_a\, P_{\overline{a}})\, (\Lambda^a{}_{\overline{b}}\, Q^{\overline{b}}) = \Lambda^{\overline{a}}{}_a\, \Lambda^a{}_{\overline{b}}\, P_{\overline{a}}\, Q^{\overline{b}} = \delta^{\overline{a}}_{\overline{b}}\, P_{\overline{a}}\, Q^{\overline{b}} = P_{\overline{a}}\, Q^{\overline{a}}. \tag{6.42}$$

6.4 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

$$\mathrm{d}V = \mathrm{d}x\,\mathrm{d}y\,\mathrm{d}z\,(\hat{\boldsymbol{x}} \wedge \hat{\boldsymbol{y}}) \cdot \hat{\boldsymbol{z}} = \mathrm{d}x\,\mathrm{d}y\,\mathrm{d}z. \tag{6.43}$$

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.

6.4.1 Jacobian of transformation

From multi-variate calculus, the relation between $\mathrm{d}\xi^1\,\mathrm{d}\xi^2\,\mathrm{d}\xi^3$ and $\mathrm{d}\xi^{\bar{1}}\,\mathrm{d}\xi^{\bar{2}}\,\mathrm{d}\xi^{\bar{3}}$ for two sets of coordinates is given by

$$\mathrm{d}\xi^1\,\mathrm{d}\xi^2\,\mathrm{d}\xi^3 = \frac{\partial(\xi^1,\xi^2,\xi^3)}{\partial(\xi^{\bar{1}},\xi^{\bar{2}},\xi^{\bar{3}})}\,\mathrm{d}\xi^{\bar{1}}\,\mathrm{d}\xi^{\bar{2}}\,\mathrm{d}\xi^{\bar{3}} \tag{6.44a}$$

$$= \left[\frac{\partial\boldsymbol{\xi}}{\partial\xi^{\bar{1}}} \wedge \frac{\partial\boldsymbol{\xi}}{\partial\xi^{\bar{2}}}\right] \cdot \frac{\partial\boldsymbol{\xi}}{\partial\xi^{\bar{3}}}\,\mathrm{d}\xi^{\bar{1}}\,\mathrm{d}\xi^{\bar{2}}\,\mathrm{d}\xi^{\bar{3}} \tag{6.44b}$$

$$= \det(\Lambda^a{}_{\bar{a}})\,\mathrm{d}\xi^{\bar{1}}\,\mathrm{d}\xi^{\bar{2}}\,\mathrm{d}\xi^{\bar{3}}, \tag{6.44c}$$

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 tranformation. This notation also helps maintain proper conservation of tensor indices.

6.4.2 Relating the Jacobian to the determinant of the metric

Recall the expression (6.12) for the transformation of the metric

$$g_{\bar{a}\bar{b}} = \Lambda^a{}_{\bar{a}}\Lambda^b{}_{\bar{b}}\,g_{ab}. \tag{6.45}$$

We can write this expression as a matrix equation

$$\overline{\mathcal{G}} = \Lambda^T\,\mathcal{G}\,\Lambda \tag{6.46}$$

where Λ^T is the transposed matrix. Taking determinants of both sides yields

$$\det(\overline{\mathcal{G}}) = \det(\Lambda^T\,\mathcal{G}\,\Lambda) = \det(\Lambda^T)\,\det(\mathcal{G})\,\det(\Lambda) = [\det(\Lambda)]^2\,\det(\mathcal{G}). \tag{6.47}$$

To reach this result we used the property of determinants that $\det(A\,B) = \det(A)\det(B)$ for any two matrices, and $\det(\Lambda^T) = \det(\Lambda)$. Consequently,

$$\det(\Lambda^a{}_{\bar{a}}) = \frac{\sqrt{\det(g_{\bar{a}\bar{b}})}}{\sqrt{\det(g_{ab})}} = \frac{\sqrt{\det(\overline{\mathcal{G}})}}{\sqrt{\det(\mathcal{G})}}. \tag{6.48}$$

We are thus led to the equivalent expressions for the volume element

$$\mathrm{d}V \equiv \sqrt{\det(\mathcal{G})}\,\mathrm{d}\xi^1\,\mathrm{d}\xi^2\,\mathrm{d}\xi^3 = \sqrt{\det(\overline{\mathcal{G}})}\,\mathrm{d}\xi^{\bar{1}}\,\mathrm{d}\xi^{\bar{2}}\,\mathrm{d}\xi^{\bar{3}}. \tag{6.49}$$

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

$$\det(\Lambda^a{}_{\bar{a}}) = \sqrt{\det(g_{\bar{a}\bar{b}})} \qquad \text{unbarred coordinates are Cartesian.} \tag{6.50}$$

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.

6.5 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.

6.5.1 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 with $a = 1, 2$. The determinant of the transformation is given by

$$\det(\Lambda^a{}_{\bar{a}}) = \Lambda^1{}_{\bar{1}}\,\Lambda^2{}_{\bar{2}} - \Lambda^1{}_{\bar{2}}\,\Lambda^2{}_{\bar{1}}. \tag{6.51}$$

Introducing the permutation symbol ϵ_{ab} allows us to write this expression in a more tidy manner

$$\det(\Lambda^a{}_{\bar{a}}) = \epsilon_{ab}\,\Lambda^a{}_{\bar{1}}\,\Lambda^b{}_{\bar{2}} \tag{6.52}$$

with

$$\epsilon_{12} = 1 \qquad \text{and} \qquad \epsilon_{21} = -1. \tag{6.53}$$

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

$$\epsilon_{123} = 1 \tag{6.54a}$$

$$\epsilon_{abc} = \begin{cases} 0 & \text{if any two labels are the same,} \\ 1 & \text{if } a, b, c \text{ is an even permutation of } 1, 2, 3, \\ -1 & \text{if } a, b, c \text{ is an odd permutation of } 1, 2, 3. \end{cases} \tag{6.54b}$$

Likewise, the determinant of the transformation matrix takes the form

$$\det(\Lambda^a{}_{\bar{a}}) = \frac{\partial(\xi^1, \xi^2, \xi^3)}{\partial(\xi^{\bar{1}}, \xi^{\bar{2}}, \xi^{\bar{3}})} = \epsilon_{abc}\,\Lambda^a{}_{\bar{1}}\,\Lambda^b{}_{\bar{2}}\,\Lambda^c{}_{\bar{3}}. \tag{6.55}$$

6.5.2 Further identities satisfied by the determinant

The following identity in two dimensions can be readily verified through enumeration

$$\epsilon_{ab}\,\Lambda^a{}_{\bar{a}}\,\Lambda^b{}_{\bar{b}} = \epsilon_{\bar{a}\bar{b}}\,\det(\Lambda^a{}_{\bar{a}}). \tag{6.56}$$

It follows directly from the definition of the determinant and can be explicity 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

$$\frac{1}{2}\,\epsilon^{\bar{a}\bar{b}}\,\epsilon_{ab}\,\Lambda^a{}_{\bar{a}}\,\Lambda^b{}_{\bar{b}} = \det(\Lambda^a{}_{\bar{a}}), \tag{6.57}$$

where we used

$$\epsilon^{\bar{a}\bar{b}}\,\epsilon_{\bar{a}\bar{b}} = \epsilon^{\overline{12}}\,\epsilon_{\overline{12}} + \epsilon^{\overline{21}}\,\epsilon_{\overline{21}} = 2. \tag{6.58}$$

The three dimensional version takes the form

$$\epsilon_{abc}\,\Lambda^a{}_{\bar{a}}\,\Lambda^b{}_{\bar{b}}\,\Lambda^c{}_{\bar{c}} = \epsilon_{\bar{a}\bar{b}\bar{c}}\,\det(\Lambda^a{}_{\bar{a}}), \tag{6.59}$$

so that

$$\frac{1}{3!}\,\epsilon^{\bar{a}\bar{b}\bar{c}}\,\epsilon_{abc}\,\Lambda^a{}_{\bar{a}}\,\Lambda^b{}_{\bar{b}}\,\Lambda^c{}_{\bar{c}} = \det(\Lambda^a{}_{\bar{a}}). \tag{6.60}$$

6.6 The Levi-Civita tensor and the volume element

The metric tensor introduced in Section 6.1 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.

6.6.1 General coordinate representation of the Levi-Civita tensor

The relations (6.56) and (6.59) 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. However, they are not unity in general, which motivates us to introduce the general coordinate form of the *Levi-Civita tensor*

$$\varepsilon_{abc} = \sqrt{\det(\mathcal{G})}\,\epsilon_{abc}. \tag{6.61}$$

We highlight the distinct symbols in this definition, with ε the Levi-Citia tensor and ϵ the permutation symbol. By construction, the Levi-Civita tensor components transform as

$$\begin{aligned}
\Lambda^a{}_{\bar{a}}\,\Lambda^b{}_{\bar{b}}\,\Lambda^c{}_{\bar{c}}\,\varepsilon_{abc} &= \Lambda^a{}_{\bar{a}}\,\Lambda^b{}_{\bar{b}}\,\Lambda^c{}_{\bar{c}}\,\sqrt{\det(\mathcal{G})}\,\epsilon_{abc} && \text{(6.62a)}\\
&= \sqrt{\det(\mathcal{G})}\,\epsilon_{\bar{a}\bar{b}\bar{c}}\,\det(\Lambda^a{}_{\bar{a}}) && \text{(6.62b)}\\
&= \sqrt{\det(\overline{\mathcal{G}})}\,\epsilon_{\bar{a}\bar{b}\bar{c}} && \text{(6.62c)}\\
&= \varepsilon_{\bar{a}\bar{b}\bar{c}}, && \text{(6.62d)}
\end{aligned}$$

where equations (6.48) and (6.56) were used. Therefore, ε_{abc} transforms as components to a third order covariant tensor. Likewise,

$$\varepsilon^{abc} = \frac{\epsilon^{abc}}{\sqrt{\det(\mathcal{G})}} \tag{6.63}$$

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.

6.6.2 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

$$\begin{aligned}
\varepsilon(\vec{e}_1\,\mathrm{d}\xi^1, \vec{e}_2\,\mathrm{d}\xi^2, \vec{e}_3\,\mathrm{d}\xi^3) &= \mathrm{d}\xi^1\,\mathrm{d}\xi^2\,\mathrm{d}\xi^3\,\varepsilon(\vec{e}_1, \vec{e}_2, \vec{e}_3) && \text{(6.64a)}\\
&= \mathrm{d}\xi^1\,\mathrm{d}\xi^2\,\mathrm{d}\xi^3\,\varepsilon_{123} && \text{(6.64b)}\\
&= \mathrm{d}\xi^1\,\mathrm{d}\xi^2\,\mathrm{d}\xi^3\,\sqrt{\det(\mathcal{G})}\,\epsilon_{123} && \text{(6.64c)}\\
&= \mathrm{d}V, && \text{(6.64d)}
\end{aligned}$$

where we used equation (6.49) for the final step. This result means that geometrically, the Levi-Civita tensor measures the volume defined by three vectors

$$\varepsilon(\vec{A}, \vec{B}, \vec{C}) = \mathrm{volume}(\vec{A}, \vec{B}, \vec{C}). \tag{6.65}$$

This interpretation accords with the Cartesian coordinate discussion of the Levi-Civita tensor in Section 1.5.3.

6.7 Vector cross product

The vector cross product of two Cartesian basis vectors yields the third, so that

$$\hat{\boldsymbol{x}} \wedge \hat{\boldsymbol{y}} = \hat{\boldsymbol{z}} \qquad \text{and cyclic permutations.} \tag{6.66}$$

The coordinate invariant generalization of this relation is given by

$$\vec{e}_a \wedge \vec{e}_b \equiv \varepsilon_{abc}\,\widetilde{e}^c. \tag{6.67}$$

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

$$\begin{aligned}
\vec{P} \wedge \vec{Q} &= P^a\,Q^b\,\vec{e}_a \wedge \vec{e}_b && \text{(6.68a)}\\
&= P^a\,Q^b\,\varepsilon_{abc}\,\widetilde{e}^c. && \text{(6.68b)}
\end{aligned}$$

6.8 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, at least 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 12.6 and 12.7, also for considering the transformation of partial derivatives.

6.8.1 The space-time transformation matrix

As discussed in Section 6.1.4, 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 time takes on the form

$$\Lambda^{\alpha}{}_{\overline{\alpha}} = \frac{\partial \xi^{\alpha}}{\partial \xi^{\overline{\alpha}}} = \begin{bmatrix} \frac{\partial \xi^0}{\partial \xi^{\overline{0}}} & \frac{\partial \xi^0}{\partial \xi^{\overline{1}}} & \frac{\partial \xi^0}{\partial \xi^{\overline{2}}} & \frac{\partial \xi^0}{\partial \xi^{\overline{3}}} \\ \frac{\partial \xi^1}{\partial \xi^{\overline{0}}} & \frac{\partial \xi^1}{\partial \xi^{\overline{1}}} & \frac{\partial \xi^1}{\partial \xi^{\overline{2}}} & \frac{\partial \xi^1}{\partial \xi^{\overline{3}}} \\ \frac{\partial \xi^2}{\partial \xi^{\overline{0}}} & \frac{\partial \xi^2}{\partial \xi^{\overline{1}}} & \frac{\partial \xi^2}{\partial \xi^{\overline{2}}} & \frac{\partial \xi^2}{\partial \xi^{\overline{3}}} \\ \frac{\partial \xi^3}{\partial \xi^{\overline{0}}} & \frac{\partial \xi^3}{\partial \xi^{\overline{1}}} & \frac{\partial \xi^3}{\partial \xi^{\overline{2}}} & \frac{\partial \xi^3}{\partial \xi^{\overline{3}}} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ \frac{\partial \xi^1}{\partial \xi^{\overline{0}}} & \frac{\partial \xi^1}{\partial \xi^{\overline{1}}} & \frac{\partial \xi^1}{\partial \xi^{\overline{2}}} & \frac{\partial \xi^1}{\partial \xi^{\overline{3}}} \\ \frac{\partial \xi^2}{\partial \xi^{\overline{0}}} & \frac{\partial \xi^2}{\partial \xi^{\overline{1}}} & \frac{\partial \xi^2}{\partial \xi^{\overline{2}}} & \frac{\partial \xi^2}{\partial \xi^{\overline{3}}} \\ \frac{\partial \xi^3}{\partial \xi^{\overline{0}}} & \frac{\partial \xi^3}{\partial \xi^{\overline{1}}} & \frac{\partial \xi^3}{\partial \xi^{\overline{2}}} & \frac{\partial \xi^3}{\partial \xi^{\overline{3}}} \end{bmatrix}. \tag{6.69}$$

The final equality made use of our assumption that $\xi^{\overline{0}} = \xi^0$ since the time coordinate remains universal. Hence, when computing $\partial \xi^0 / \partial \xi^{\overline{\alpha}}$ we keep $\xi^{\overline{0}}$ fixed so that the derivative vanishes as in the specific case of

$$\left[\frac{\partial \xi^0}{\partial \xi^{\overline{1}}}\right]_{\xi^{\overline{0}},\xi^{\overline{2}},\xi^{\overline{3}}} = \left[\frac{\partial \xi^0}{\partial \xi^{\overline{1}}}\right]_{\xi^{0},\xi^{\overline{2}},\xi^{\overline{3}}} = 0. \tag{6.70}$$

Zero elements in the first row of the transformation matrix (6.69) reveals that time is not a function of space

$$\frac{\partial \xi^0}{\partial \xi^{\overline{a}}} = 0. \tag{6.71}$$

In contrast, nonzero elements in the first column indicate that our description of space is a function of time

$$\frac{\partial \xi^a}{\partial \xi^{\overline{0}}} \neq 0. \tag{6.72}$$

We see the same overall structure in the inverse space-time transformation matrix

$$\Lambda^{\overline{\alpha}}{}_{\alpha} = \frac{\partial \xi^{\overline{\alpha}}}{\partial \xi^{\alpha}} = \begin{bmatrix} \frac{\partial \xi^{\overline{0}}}{\partial \xi^0} & \frac{\partial \xi^{\overline{0}}}{\partial \xi^1} & \frac{\partial \xi^{\overline{0}}}{\partial \xi^2} & \frac{\partial \xi^{\overline{0}}}{\partial \xi^3} \\ \frac{\partial \xi^{\overline{1}}}{\partial \xi^0} & \frac{\partial \xi^{\overline{1}}}{\partial \xi^1} & \frac{\partial \xi^{\overline{1}}}{\partial \xi^2} & \frac{\partial \xi^{\overline{1}}}{\partial \xi^3} \\ \frac{\partial \xi^{\overline{2}}}{\partial \xi^0} & \frac{\partial \xi^{\overline{2}}}{\partial \xi^1} & \frac{\partial \xi^{\overline{2}}}{\partial \xi^2} & \frac{\partial \xi^{\overline{2}}}{\partial \xi^3} \\ \frac{\partial \xi^{\overline{3}}}{\partial \xi^0} & \frac{\partial \xi^{\overline{3}}}{\partial \xi^1} & \frac{\partial \xi^{\overline{3}}}{\partial \xi^2} & \frac{\partial \xi^{\overline{3}}}{\partial \xi^3} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ \frac{\partial \xi^{\overline{1}}}{\partial \xi^0} & \frac{\partial \xi^{\overline{1}}}{\partial \xi^1} & \frac{\partial \xi^{\overline{1}}}{\partial \xi^2} & \frac{\partial \xi^{\overline{1}}}{\partial \xi^3} \\ \frac{\partial \xi^{\overline{2}}}{\partial \xi^0} & \frac{\partial \xi^{\overline{2}}}{\partial \xi^1} & \frac{\partial \xi^{\overline{2}}}{\partial \xi^2} & \frac{\partial \xi^{\overline{2}}}{\partial \xi^3} \\ \frac{\partial \xi^{\overline{3}}}{\partial \xi^0} & \frac{\partial \xi^{\overline{3}}}{\partial \xi^1} & \frac{\partial \xi^{\overline{3}}}{\partial \xi^2} & \frac{\partial \xi^{\overline{3}}}{\partial \xi^3} \end{bmatrix}. \tag{6.73}$$

6.8.2 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

$$\det(\Lambda^{\alpha}{}_{\overline{\alpha}}) = \det(\Lambda^{a}{}_{\overline{a}}) \qquad \text{and} \qquad \det(\Lambda^{\overline{\alpha}}{}_{\alpha}) = \det(\Lambda^{\overline{a}}{}_{a}), \tag{6.74}$$

which follows since the first row in both transformations has only a single non-zero value, $\Lambda^{0}{}_{\overline{0}} = 1$ and $\Lambda^{\overline{0}}{}_{0} = 1$. Hence, the relations developed in Sections 6.4 and 6.5 for the volume element and Jacobian of transformation remain unchanged when adding the universal time coordinate.

6.8.3 Multiplying the transformation matrix and its inverse

We here verify that the transformation matrix (6.69) indeed has its inverse given by (6.73). For this purpose we must prove that

$$\delta^{\alpha}{}_{\beta} = \Lambda^{\alpha}{}_{\overline{\beta}}\,\Lambda^{\overline{\beta}}{}_{\beta} \qquad \text{and} \qquad \delta^{\overline{\alpha}}{}_{\overline{\beta}} = \Lambda^{\overline{\alpha}}{}_{\beta}\,\Lambda^{\beta}{}_{\overline{\beta}}, \tag{6.75}$$

where $\delta^{\alpha}{}_{\beta}$ and $\delta^{\overline{\alpha}}{}_{\overline{\beta}}$ are components to the identity tensor. The proof relies on writing the space-time coordinate transformation as a composite function

$$\xi^{\overline{\alpha}} = \xi^{\overline{\alpha}}(\xi^{\alpha}) = \xi^{\overline{\alpha}}[\xi^{\alpha}(\xi^{\overline{\beta}})]. \tag{6.76}$$

Taking partial derivatives and using the chain rule renders

$$\delta^{\overline{\alpha}}{}_{\overline{\beta}} = \frac{\partial \xi^{\overline{\alpha}}}{\partial \xi^{\overline{\beta}}} = \frac{\partial \xi^{\overline{\alpha}}}{\partial \xi^{\alpha}}\frac{\partial \xi^{\alpha}}{\partial \xi^{\overline{\beta}}} = \Lambda^{\overline{\alpha}}{}_{\alpha}\,\Lambda^{\alpha}{}_{\overline{\beta}} \qquad \text{and} \qquad \delta^{\alpha}{}_{\beta} = \frac{\partial \xi^{\alpha}}{\partial \xi^{\beta}} = \frac{\partial \xi^{\alpha}}{\partial \xi^{\overline{\alpha}}}\frac{\partial \xi^{\overline{\alpha}}}{\partial \xi^{\beta}} = \Lambda^{\alpha}{}_{\overline{\alpha}}\,\Lambda^{\overline{\alpha}}{}_{\beta}. \tag{6.77}$$

Furthermore, the space subcomponents decouple from time, which can be seen by considering a few representative cases

$$1 = \delta^{0}{}_{0} = \Lambda^{0}{}_{\overline{\alpha}}\,\Lambda^{\overline{\alpha}}{}_{0} = \Lambda^{0}{}_{\overline{0}}\,\Lambda^{\overline{0}}{}_{0} \tag{6.78a}$$
$$1 = \delta^{1}{}_{1} = \Lambda^{1}{}_{\overline{\alpha}}\,\Lambda^{\overline{\alpha}}{}_{1} = \Lambda^{1}{}_{\overline{a}}\,\Lambda^{\overline{a}}{}_{1} \tag{6.78b}$$
$$0 = \delta^{0}{}_{1} = \Lambda^{0}{}_{\overline{\alpha}}\,\Lambda^{\overline{\alpha}}{}_{1} = \Lambda^{0}{}_{\overline{0}}\,\Lambda^{\overline{0}}{}_{1} \tag{6.78c}$$
$$0 = \delta^{1}{}_{2} = \Lambda^{1}{}_{\overline{\alpha}}\,\Lambda^{\overline{\alpha}}{}_{2} = \Lambda^{1}{}_{\overline{a}}\,\Lambda^{\overline{a}}{}_{2}. \tag{6.78d}$$

Consequently, the spatial components satisfy

$$\delta^{a}{}_{b} = \Lambda^{a}{}_{\overline{b}}\,\Lambda^{\overline{b}}{}_{b} \qquad \text{and} \qquad \delta^{\overline{a}}{}_{\overline{b}} = \Lambda^{\overline{a}}{}_{b}\,\Lambda^{b}{}_{\overline{b}}, \tag{6.79}$$

which allows for a splitting of the spatial components from the time component.

6.8.4 Transformation of space and time partial derivatives

Application of the chain rule leads to the transformation of the partial derivative operator

$$\partial_{\overline{\alpha}} = \frac{\partial}{\partial \xi^{\overline{\alpha}}} = \frac{\partial \xi^{\alpha}}{\partial \xi^{\overline{\alpha}}}\frac{\partial}{\partial \xi^{\alpha}} = \Lambda^{\alpha}{}_{\overline{\alpha}}\,\partial_{\alpha}. \tag{6.80}$$

Extracting the time and space components from the transformation matrix (6.69) yields

$$\partial_{\overline{0}} = \Lambda^{\alpha}{}_{\overline{0}}\,\partial_{\alpha} = \partial_{0} + \Lambda^{a}{}_{\overline{0}}\,\partial_{a} \tag{6.81a}$$
$$\partial_{\overline{a}} = \Lambda^{\alpha}{}_{\overline{a}}\,\partial_{\alpha} = \Lambda^{a}{}_{\overline{a}}\,\partial_{a}. \tag{6.81b}$$

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.

6.9 Covariant derivative acting on 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 form of the gradient operator acting on a scalar

$$\text{grad}(\psi) = \nabla\psi = \widetilde{e}^a\,\partial_a\psi = \widetilde{e}^{\overline{a}}\,\partial_{\overline{a}}\psi. \tag{6.82}$$

From this expression we define the *covariant derivative operator*

$$\nabla = \widetilde{e}^a\,\partial_a \tag{6.83}$$

so that we refer to equation (6.82) as either the gradient acting on a scalar or the covariant derivative acting on a scalar.

6.10 Covariant derivative acting on 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

$$\nabla\vec{F} = (\widetilde{e}^b\,\partial_b)\,(F^a\,\vec{e}_a). \tag{6.84}$$

6.10.1 Derivative of a vector

The chain rule leads to the expression for the partial derivative operator acting on a vector field

$$\begin{aligned}
\partial_b\vec{F} &= \partial_b(\vec{e}_a\,F^a) && \text{coordinate representation of the vector } \vec{F} && (6.85a)\\
&= (\partial_b\,F^a)\,\vec{e}_a + F^a\,\partial_b\,\vec{e}_a && \text{chain rule} && (6.85b)\\
&= (\partial_b\,F^a)\,\vec{e}_a + F^a\,\Gamma^c_{ba}\,\vec{e}_c && \text{define Christoffel symbols} && (6.85c)\\
&= (\partial_b\,F^a + F^c\,\Gamma^a_{bc})\,\vec{e}_a && \text{reorganize} && (6.85d)\\
&= (\nabla_b F^a)\,\vec{e}_a && \text{define covariant derivative acting on vector component.} && (6.85e)
\end{aligned}$$

In the third equality we introduced the *Christoffel symbols*

$$\partial_b\,\vec{e}_a = \Gamma^c_{ba}\,\vec{e}_c. \tag{6.86}$$

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

$$\nabla_b F^a = \partial_b\,F^a + \Gamma^a_{bc}\,F^c. \tag{6.87}$$

Contracting $\partial_b\vec{F}$ with the basis one-form $\tilde{e}^b$ leads to

$$\nabla\vec{F} = (\tilde{e}^b\,\partial_b)\,\vec{F} = (\tilde{e}^b\,\nabla_b\,F^a)\,\vec{e}_a. \tag{6.88}$$

6.10.2 An alternative derivation

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. Now apply this operation to a vector field $\vec{F}$ represented by arbitrary coordinates ξ^a, in which case

$$\partial_b \vec{F} = \lim_{\Delta \to 0} \frac{\vec{F}(\vec{P} + \Delta\, \vec{e}_b) - \vec{F}(\vec{P})}{\Delta}, \tag{6.89}$$

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

$$\partial_b \vec{F} = (\partial_b F^a)\, \vec{e}_a \qquad \text{Cartesian coordinates.} \tag{6.90}$$

However, for general coordinates both the vector components and the basis vectors are spatially dependent, in which case

$$\vec{F}(\vec{P} + \Delta\, \vec{e}_b) - \vec{F}(\vec{P}) = [F^a + \Delta\, \partial_b F^b]\, [\vec{e}_a + \Delta\, \partial_b\, \vec{e}_a] - F^a\, \vec{e}_a \tag{6.91a}$$

$$= \Delta\, \partial_b (F^a\, \vec{e}_a) + \mathcal{O}(\Delta^2). \tag{6.91b}$$

This is the same result as found in the first step of the chain rule used in equation (6.85a). Following through that derivation then leads to the same coordinate expression for the covariant derivative acting on a vector field.

6.10.3 Christoffel symbols are not components of a tensor

The Christoffel symbols vanish in Euclidean space when using Cartesian coordinates whereas they are nonzero with other coordinates. As discussed in Section 5.2, 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.

6.11 Covariant derivative of a one-form

The gradient acting on the product of a one-form and a vector is given by

$$\nabla(\widetilde{E} \cdot \vec{F}) = \widetilde{e}^b\, \partial_b\, (E_a\, F^a). \tag{6.92}$$

Expanding the partial derivative yields

$$\partial_b (E_a\, F^a) = F^a\, \partial_b E_a + E_a\, \partial_b F^a \tag{6.93a}$$

$$= F^a\, \partial_b E_a + E_a\, (\nabla_b F^a - \Gamma^a_{bc} F^c) \tag{6.93b}$$

$$= F^a (\partial_b E_a - \Gamma^c_{ba} E_c) + E_a\, \nabla_b F^a \tag{6.93c}$$

$$\equiv F^a\, \nabla_b E_a + E_a\, \nabla_b F^a. \tag{6.93d}$$

The last equality defines the covariant derivative when acting on the components to a one form

$$\nabla_b E_a = \partial_b E_a - \Gamma^c_{ba} E_c, \tag{6.94}$$

which leads to

$$\nabla \widetilde{E} = (\widetilde{e}^b\, \partial_b)\, \widetilde{E} = (\widetilde{e}^b\, \nabla_b E_a)\, \widetilde{e}^a. \tag{6.95}$$

6.12 Covariant derivative of the metric tensor

When written in Cartesian coordinates, the covariant derivative of components to the metric tensor for Euclidean space vanishes,

$$\nabla g_{ab} = \nabla \delta_{ab} = 0, \tag{6.96}$$

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.

6.13 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 (6.96)) then leads to the identity

$$0 = \nabla_c \, g_{ab} = \partial_c \, g_{ab} - \Gamma^d_{ca} \, g_{db} - \Gamma^d_{cb} \, g_{ad}. \tag{6.97}$$

We can solve this equation for the Christoffel symbols

$$\Gamma^c_{ab} = \frac{1}{2} \, g^{cd} \, (\partial_b \, g_{da} + \partial_a \, g_{db} - \partial_d \, g_{ab}). \tag{6.98}$$

This expression exhibits the symmetry property of the lower two indices on the Christoffel symbols

$$\Gamma^c_{ab} = \Gamma^c_{ba}. \tag{6.99}$$

6.14 Covariant divergence of a vector

The covariant divergence of the components to a vector results in a scalar

$$\nabla_a F^a = \partial_a F^a + \Gamma^a_{ab} F^b. \tag{6.100}$$

We now bring this expression into a form more convenient for practical calculations.

6.14.1 Contraction of the Christoffel symbols

Expression (6.98) for the Christoffel symbols yields for the contraction

$$\Gamma^a_{ab} = \frac{1}{2} \, g^{ad} (\partial_b \, g_{da} + \partial_a \, g_{db} - \partial_d \, g_{ab}) = \frac{1}{2} \, g^{ad} \, \partial_b \, g_{ad} \tag{6.101}$$

where symmetry of the metric tensor and its inverse was used.

6.14.2 Exponential of the determinant

For any symmetric positive definite matrix such as the metric tensor we can write

$$\det(A) = e^{\ln\det(A)} \qquad \text{simple identity} \tag{6.102a}$$

$$= e^{\ln(\Pi_i \Lambda_i)} \qquad \text{determinant related to product of eigenvalues} \tag{6.102b}$$

$$= e^{\Sigma_i \ln \Lambda_i} \qquad \text{simple identity} \tag{6.102c}$$

$$= e^{\mathrm{Tr}(\ln A)} \qquad \text{sum of eigenvalues related to trace of matrix.} \tag{6.102d}$$

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

$$\partial_c \ln\det(A) = \partial_c[\mathrm{Tr}(\ln A)] = \mathrm{Tr}(\partial_c \ln A) = \mathrm{Tr}(A^{-1}\partial_c A). \tag{6.103}$$

With A now set equal to the metric tensor $\mathcal{G}$ with components g_{ab}, this result yields

$$\partial_c \ln\det(\mathcal{G}) = g^{ab}\,\partial_c\, g_{ab} \tag{6.104}$$

which in turn yields for the contracted Christoffel symbol

$$\Gamma^a_{ac} = \partial_c \ln\sqrt{\det(\mathcal{G})}. \tag{6.105}$$

This result brings the covariant divergence of a vector to the form

$$\nabla_a F^a = \partial_a F^a + F^a\,\partial_a \ln\sqrt{\det(\mathcal{G})} = \frac{1}{\sqrt{\mathcal{G}}}\,\partial_a\,[\sqrt{\mathcal{G}}\;F^a]. \tag{6.106}$$

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 $\sqrt{\det(\mathcal{G})}$.

6.15 Covariant Laplacian of a scalar

Making use of equation (6.106) with

$$F^a = g^{ab}\,\partial_b\psi \tag{6.107}$$

leads to the covariant Laplacian of a scalar field

$$\nabla_a\,(g^{ab}\,\partial_b\,\psi) = \frac{1}{\sqrt{\mathcal{G}}}\,\partial_a\,[\sqrt{\mathcal{G}}\;g^{ab}\,\partial_b\psi]. \tag{6.108}$$

This expression is fundamental to the evolution of scalar fields under the impacts from diffusion (Chapter 47).

6.16 Covariant curl of a vector

The Levi-Civita tensor

$$\varepsilon_{abc} = \sqrt{\det(\mathcal{G})}\,\epsilon_{abc} \tag{6.109}$$

from Section 6.6 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

$$\mathrm{curl}(\vec{F}) = \varepsilon^{abc}\,(\nabla_b F_c)\,\vec{e}_a = \varepsilon_{abc}\,(\nabla^b F^c)\,\tilde{e}^a. \tag{6.110}$$

This expression simplifies by making use of equation (6.94) for the covariant derivative $\nabla_b F_c = \partial_b F_c - \Gamma^a_{cb} F_a$. Conveniently, the contraction $\varepsilon^{abc}\,\Gamma^a_{cb}$ vanishes identically since $\varepsilon^{abc} = -\varepsilon^{acb}$, whereas $\Gamma^a_{cb} = \Gamma^a_{bc}$. Hence, one is left with the general expression for the covariant curl that involves just the partial derivatives

$$\mathrm{curl}(\vec{F}) = \varepsilon^{abc}\,(\partial_b F_c)\,\vec{e}_a = \varepsilon^{abc}\,[\partial_b(g_{cd}F^d)]\,\vec{e}_a. \tag{6.111}$$

The second equality made use of the identity $F_c = g_{cd}F^d$.

6.17 Gauss's divergence theorem

The integral theorems from Cartesian vector analysis transform in a straightforward manner to arbitrary coordinates in arbitrary smooth spaces. An easy way to prove the theorems is to invoke the ideas of general covariance from Section 5.2, 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 (6.49)

$$\mathrm{d}V = \sqrt{\det(\mathcal{G})}\,\mathrm{d}\xi^1\,\mathrm{d}\xi^2\,\mathrm{d}\xi^3, \tag{6.112}$$

multiplied by the covariant divergence (6.106). Hence, the volume integral of the divergence is given by

$$\int_{\mathcal{R}} (\nabla_a F^a)\,\mathrm{d}V = \int_{\mathcal{R}} \partial_a[\sqrt{\det(\mathcal{G})}\,F^a]\,\mathrm{d}\xi^1\,\mathrm{d}\xi^2\,\mathrm{d}\xi^3 = \oint_{\partial\mathcal{R}} F^a\,\hat{n}_a\,\mathrm{d}\mathcal{S}, \tag{6.113}$$

where $\hat{\boldsymbol{n}}$ is the outward normal one-form for the boundary, $\partial\mathcal{R}$, and $\hat{n}_a$ are its covariant components.

6.18 Stokes' theorem

The Cartesian form of Stokes' Theorem from Section 2.6 is generalized in a manner similar to the divergence theorem

$$\oint_{\partial\mathcal{S}} \vec{F}\cdot\mathrm{d}\vec{x} = \int_{\mathcal{S}} \mathrm{curl}(\vec{F})\cdot\hat{\boldsymbol{n}}\,\mathrm{d}\mathcal{S}, \tag{6.114}$$

where $\mathrm{d}\vec{x}$ 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

$$\vec{F}\cdot\mathrm{d}\vec{x} = F^a\,\vec{e}_a\cdot\vec{e}_b\,\mathrm{d}x^b = F_b\,\mathrm{d}x^b = F_{\overline{b}}\,\mathrm{d}\xi^{\overline{b}}. \tag{6.115}$$

For the curl on the right hand side we have

$$\mathrm{curl}(\vec{F})\cdot\hat{\boldsymbol{n}} = \varepsilon^{abc}\,(\partial_b F_c)\,\vec{e}_a\cdot\hat{\boldsymbol{n}} = \varepsilon^{abc}\,(\partial_b F_c)\,\hat{n}_a = \varepsilon^{\overline{a}\overline{b}\overline{c}}\,(\partial_{\overline{b}} F_{\overline{c}})\,\hat{n}_{\overline{a}}, \tag{6.116}$$

thus leading to the expression of Stokes' theorem in arbitrary coordinates

$$\oint_{\partial\mathcal{S}} F_{\overline{b}}\,\mathrm{d}\xi^{\overline{b}} = \int_{\mathcal{S}} \varepsilon^{\overline{a}\overline{b}\overline{c}}\,(\partial_{\overline{b}} F_{\overline{c}})\,\hat{n}_{\overline{a}}\,\mathrm{d}\mathcal{S}. \tag{6.117}$$

7 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 6. These coordinates are all time independent and orthogonal. They are hence simpler than the non-orthogonal generalized vertical coordinates detailed in Chapter 8. This chapter can be referred to when needed.

7.1 Cartesian coordinates **102**
7.1.1 The basics 102
7.1.2 Summary of Cartesian coordinate expressions 103
7.2 Spherical coordinates **103**
7.2.1 Transforming between Cartesian and spherical coordinates 104
7.2.2 Basis vectors 105
7.2.3 Basis one-forms 105
7.2.4 Position and velocity 106
7.2.5 Metric tensor 106
7.2.6 Components of a vector field 107
7.2.7 Differential operators 107
7.2.8 Summary of spherical coordinate expressions 108
7.3 Cylindrical-polar coordinates **108**
7.3.1 Transforming between Cartesian and cylindrical-polar coordinates . 109
7.3.2 Basis vectors 110
7.3.3 Basis one-forms 110
7.3.4 Position and velocity 110
7.3.5 Metric tensor 111
7.3.6 Components of a vector field 111
7.3.7 Differential operators 112
7.3.8 Summary of cylindrical coordinate expressions 113
7.4 General orthogonal coordinates **113**

7.1 Cartesian coordinates

Whenever developing a general tensor relation it is useful to check its validity by considering Cartesian coordinates. We here summarize some results from our discussion of Cartesian tensors in Chapters 1 and 2.

7.1.1 The basics

We start by expressing the trajectory of a point through space in the following equivalent forms

$$\begin{aligned}
\mathscr{P}(\tau) &= \vec{e}_1\, x(\tau) + \vec{e}_2\, y(\tau) + \vec{e}_3\, z(\tau) && (7.1a)\\
&= \hat{\boldsymbol{x}}\, x(\tau) + \hat{\boldsymbol{y}}\, y(\tau) + \hat{\boldsymbol{z}}\, z(\tau) && (7.1b)\\
&= \vec{x}(\tau) && (7.1c)\\
&= \boldsymbol{x}(\tau), && (7.1d)
\end{aligned}$$

with the basis vectors written

$$\vec{e}_1 = \hat{\boldsymbol{x}} \qquad \text{and} \qquad \vec{e}_2 = \hat{\boldsymbol{y}} \qquad \text{and} \qquad \vec{e}_3 = \hat{\boldsymbol{z}}. \tag{7.2}$$

The boldface notation is used for the position vector in the final equality of equation (7.1d), with the boldface commonly used throughout this book. Notably, the orthogonal unit vectors for Cartesian coordinates are normalized so that

$$\vec{e}_1 \cdot \vec{e}_1 = \vec{e}_2 \cdot \vec{e}_2 = \vec{e}_3 \cdot \vec{e}_3 = 1. \tag{7.3}$$

Furthermore, the basis vectors are identical to the basis one-forms

$$\vec{e}_1 = \widetilde{e}^1 = \hat{\boldsymbol{x}} \qquad \text{and} \qquad \vec{e}_2 = \widetilde{e}^2 = \hat{\boldsymbol{y}} \qquad \text{and} \qquad \vec{e}_3 = \widetilde{e}^3 = \hat{\boldsymbol{z}}. \tag{7.4}$$

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

$$\vec{v}(\tau) = \boldsymbol{v}(\tau) = \frac{\mathrm{d}\mathscr{P}}{\mathrm{d}\tau} = \frac{\mathrm{d}\boldsymbol{x}}{\mathrm{d}\tau}, \tag{7.5}$$

which take on the expanded forms

$$\begin{aligned}
\vec{v}(\tau) &= \vec{e}_1\, \frac{\mathrm{d}x(\tau)}{\mathrm{d}\tau} + \vec{e}_2\, \frac{\mathrm{d}y(\tau)}{\mathrm{d}\tau} + \vec{e}_3\, \frac{\mathrm{d}z(\tau)}{\mathrm{d}\tau} && (7.6a)\\
&= \hat{\boldsymbol{x}}\, v^1(\tau) + \hat{\boldsymbol{y}}\, v^2(\tau) + \hat{\boldsymbol{z}}\, v^3(\tau). && (7.6b)\\
&= \hat{\boldsymbol{x}}\, u(\tau) + \hat{\boldsymbol{y}}\, v(\tau) + \hat{\boldsymbol{z}}\, w(\tau). && (7.6c)
\end{aligned}$$

7.1.2 Summary of Cartesian coordinate expressions

In Cartesian coordinates, mathematical operators and integral theorems take their familiar form. We here list the key ones in forms that are encountered throughout this book.

$$\boldsymbol{x} = (x^1, x^2, x^3) = (x, y, z) \qquad \text{Cartesian coordinates} \tag{7.7}$$

$$\boldsymbol{F} = \hat{\boldsymbol{x}} F^1 + \hat{\boldsymbol{y}} F^2 + \hat{\boldsymbol{z}} F^3 = \hat{\boldsymbol{x}} F_1 + \hat{\boldsymbol{y}} F_2 + \hat{\boldsymbol{z}} F_3 \qquad \text{covariant} = \text{contravariant} \tag{7.8}$$

$$\frac{\partial}{\partial x^a} = \partial_a \ \text{ or } \ (\partial_x, \partial_y, \partial_z) \qquad \text{partial derivative operator} \tag{7.9}$$

$$\nabla = \hat{\boldsymbol{x}}\, \partial_x + \hat{\boldsymbol{y}}\, \partial_y + \hat{\boldsymbol{z}}\, \partial_z \qquad \text{gradient operator} \tag{7.10}$$

$$\nabla_z = \hat{\boldsymbol{x}}\, \partial_x + \hat{\boldsymbol{y}}\, \partial_y \qquad \text{horizontal gradient operator} \tag{7.11}$$

$$\nabla \cdot \boldsymbol{F} = \partial_x F_x + \partial_y F_y + \partial_z F_z \qquad \text{divergence of a vector} \tag{7.12}$$

$$\nabla_z \cdot \boldsymbol{F} = \partial_x F_x + \partial_y F_y \qquad \text{horizontal divergence of a vector} \tag{7.13}$$

$$(\nabla \wedge \boldsymbol{F})_a = \epsilon_{abc} \partial_b F_c \qquad \text{components to curl of a vector} \tag{7.14}$$

$$\nabla \cdot \nabla \psi = \nabla^2 \psi = (\partial_{xx} + \partial_{yy} + \partial_{zz})\, \psi \qquad \text{Laplacian of a scalar} \tag{7.15}$$

$$\int_{\mathcal{R}} \nabla \cdot \boldsymbol{F}\, \mathrm{d}V = \oint_{\partial \mathcal{R}} \boldsymbol{F} \cdot \hat{\boldsymbol{n}}\, \mathrm{d}\mathcal{S} \qquad \text{Gauss's divergence theorem} \tag{7.16}$$

$$\oint_{\partial \mathcal{S}} \boldsymbol{F} \cdot \mathrm{d}\boldsymbol{x} = \int_{\mathcal{S}} (\nabla \wedge \boldsymbol{F}) \cdot \hat{\boldsymbol{n}}\, \mathrm{d}\mathcal{S}. \qquad \text{Stokes' theorem.} \tag{7.17}$$

7.2 Spherical coordinates

We now consider spherical coordinates defined by Figure 7.1 and related to Cartesian coordinates through

$$x = r \cos\phi\, \cos\lambda \tag{7.18a}$$

$$y = r \cos\phi\, \sin\lambda \tag{7.18b}$$

$$z = r \sin\phi. \tag{7.18c}$$

The radial coordinate

$$r = |\boldsymbol{x}| = \sqrt{\boldsymbol{x} \cdot \boldsymbol{x}} = \sqrt{x^2 + y^2 + z^2} \tag{7.19}$$

measures the distance from the center of the sphere to position of the particle. The spherical angle coordinates

$$0 \le \lambda \le 2\pi \qquad \text{longitude} \tag{7.20}$$

$$-\pi/2 \le \phi \le \pi/2 \qquad \text{latitude} \tag{7.21}$$

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

$$(x, y, z) = (\xi^1, \xi^2, \xi^3) \equiv \xi^a \qquad \text{and} \qquad (\lambda, \phi, r) = (\xi^{\bar{1}}, \xi^{\bar{2}}, \xi^{\bar{3}}) \equiv \xi^{\bar{a}}. \tag{7.22}$$

Figure 7.1: *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, $(\hat{\boldsymbol{x}}, \hat{\boldsymbol{y}}, \hat{\boldsymbol{z}})$ points along the orthogonal axes and rotates with the sphere. The planetary spherical triad (also rotating with the sphere) of orthonormal basis vectors, $(\hat{\boldsymbol{\lambda}}, \hat{\boldsymbol{\phi}}, \hat{\boldsymbol{r}})$, makes use of the longitudinal unit vector $\hat{\boldsymbol{\lambda}}$, which points in the longitudinal direction (positive eastward), the latitudinal unit vector $\hat{\boldsymbol{\phi}}$, which points in the latitudinal direction (positive northward) and the radial unit vector $\hat{\boldsymbol{r}}$, which point in the radial direction (positive away from the center).*

7.2.1 Transforming between Cartesian and spherical coordinates

Following the general discussion in Section 6.1.4, we consider the infinitesimal distance along one of the Cartesian coordinate axes, $\mathrm{d}\xi^a$. The chain rule allows us to relate this distance to those along the axes of the spherical coordinate system

$$\mathrm{d}\xi^a = \frac{\partial \xi^a}{\partial \xi^{\overline{a}}}\,\mathrm{d}\xi^{\overline{a}} = \Lambda^a_{\ \overline{a}}\,\mathrm{d}\xi^{\overline{a}}. \tag{7.23}$$

The partial derivatives $\partial\xi^a/\partial\xi^{\overline{a}}$ form components to the transformation matrix that transforms between coordinate representations. For the coordinate relation (7.18a)-(7.18c), this transformation matrix is given by

$$\Lambda^a_{\ \overline{a}} = \begin{bmatrix} \partial\xi^1/\partial\xi^{\overline{1}} & \partial\xi^1/\partial\xi^{\overline{2}} & \partial\xi^1/\partial\xi^{\overline{3}} \\ \partial\xi^2/\partial\xi^{\overline{1}} & \partial\xi^2/\partial\xi^{\overline{2}} & \partial\xi^2/\partial\xi^{\overline{3}} \\ \partial\xi^3/\partial\xi^{\overline{1}} & \partial\xi^3/\partial\xi^{\overline{2}} & \partial\xi^3/\partial\xi^{\overline{3}} \end{bmatrix} = \begin{bmatrix} -r\cos\phi\sin\lambda & -r\sin\phi\,\cos\lambda & \cos\phi\cos\lambda \\ r\cos\phi\cos\lambda & -r\sin\phi\,\sin\lambda & \cos\phi\sin\lambda \\ 0 & r\cos\phi & \sin\phi \end{bmatrix}. \tag{7.24}$$

The determinant of the transformation (Jacobian) is given by

$$\det(\Lambda^a_{\ \overline{a}}) = r^2\cos\phi. \tag{7.25}$$

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

$$\Lambda^{\overline{a}}_{\ a} = \frac{1}{r^2\cos\phi}\begin{bmatrix} -r\sin\lambda & r\cos\lambda & 0 \\ -r\cos\phi\,\sin\phi\,\cos\lambda & -r\cos\phi\sin\phi\sin\lambda & r\cos^2\phi \\ r^2\cos^2\phi\cos\lambda & r^2\,\cos^2\phi\sin\lambda & r^2\cos\phi\sin\phi \end{bmatrix}. \tag{7.26}$$

7.2.2 Basis vectors

The spherical coordinate basis vectors, $\vec{e}_{\overline{a}}$, are related to the Cartesian coordinate basis vectors, $\vec{e}_a$, through the transformation

$$\vec{e}_{\overline{a}} = \Lambda^{a}_{\ \overline{a}}\,\vec{e}_a. \tag{7.27}$$

The transformation matrix (7.24) leads to

$$\vec{e}_\lambda = r\cos\phi\,(-\hat{\boldsymbol{x}}\sin\lambda + \hat{\boldsymbol{y}}\cos\lambda) \tag{7.28a}$$
$$\vec{e}_\phi = r(-\hat{\boldsymbol{x}}\sin\phi\,\cos\lambda - \hat{\boldsymbol{y}}\sin\phi\,\sin\lambda + \hat{\boldsymbol{z}}\cos\phi) \tag{7.28b}$$
$$\vec{e}_r = \hat{\boldsymbol{x}}\cos\phi\,\cos\lambda + \hat{\boldsymbol{y}}\cos\phi\,\sin\lambda + \hat{\boldsymbol{z}}\sin\phi. \tag{7.28c}$$

We can introduce the orthonormal unit vectors through

$$\vec{e}_\lambda = r\cos\phi\,\hat{\boldsymbol{\lambda}} \qquad \text{and} \qquad \vec{e}_\phi = r\,\hat{\boldsymbol{\phi}} \qquad \text{and} \qquad \vec{e}_r = \hat{\boldsymbol{r}}, \tag{7.29}$$

so that

$$\hat{\boldsymbol{\lambda}} = -\hat{\boldsymbol{x}}\sin\lambda + \hat{\boldsymbol{y}}\cos\lambda \tag{7.30a}$$
$$\hat{\boldsymbol{\phi}} = -\hat{\boldsymbol{x}}\cos\lambda\,\sin\phi - \hat{\boldsymbol{y}}\sin\lambda\,\sin\phi + \hat{\boldsymbol{z}}\cos\phi \tag{7.30b}$$
$$\hat{\boldsymbol{r}} = \hat{\boldsymbol{x}}\,\cos\lambda\,\cos\phi + \hat{\boldsymbol{y}}\sin\lambda\,\cos\phi + \hat{\boldsymbol{z}}\sin\phi \tag{7.30c}$$

along with the inverse relations

$$\hat{\boldsymbol{x}} = -\hat{\boldsymbol{\lambda}}\sin\lambda - \hat{\boldsymbol{\phi}}\cos\lambda\sin\phi + \hat{\boldsymbol{r}}\,\cos\lambda\,\cos\phi \tag{7.31a}$$
$$\hat{\boldsymbol{y}} = \hat{\boldsymbol{\lambda}}\cos\lambda - \hat{\boldsymbol{\phi}}\,\sin\lambda\,\sin\phi + \hat{\boldsymbol{r}}\sin\lambda\,\cos\phi \tag{7.31b}$$
$$\hat{\boldsymbol{z}} = \hat{\boldsymbol{\phi}}\,\cos\phi + \hat{\boldsymbol{r}}\,\sin\phi. \tag{7.31c}$$

7.2.3 Basis one-forms

Since spherical coordinates are orthogonal, we can readily derive the one-form basis through inverting the vector basis

$$\widetilde{e}^\lambda = (r\cos\phi)^{-1}\,\hat{\boldsymbol{\lambda}} \qquad \text{and} \qquad \widetilde{e}^\phi = r^{-1}\hat{\boldsymbol{\phi}} \qquad \text{and} \qquad \widetilde{e}^r = \hat{\boldsymbol{r}}, \tag{7.32}$$

which satisfy the bi-orthogonality relation with the basis vectors (Section 6.2.2)

$$\widetilde{e}^{\overline{b}}\cdot\vec{e}_{\overline{a}} = \delta^{\overline{b}}_{\ \overline{a}}. \tag{7.33}$$

7.2.4 Position and velocity

In spherical coordinates, the position of a point is fully specified by the radial position

$$\mathscr{P}(\tau) = r\,\vec{e}_r = r\,\hat{\boldsymbol{r}}. \tag{7.34}$$

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

$$\begin{aligned}
\vec{v}(\tau) &= \boldsymbol{v} && (7.35a)\\
&= \frac{\mathrm{d}\mathscr{P}}{\mathrm{d}\tau} && (7.35b)\\
&= \vec{e}_r\,\frac{\mathrm{d}r}{\mathrm{d}\tau} + r\,\frac{\mathrm{d}\vec{e}_r}{\mathrm{d}\tau} && (7.35c)\\
&= \vec{e}_r\,\frac{\mathrm{d}r}{\mathrm{d}\tau} + r\,\frac{\partial\vec{e}_r}{\partial\lambda}\,\frac{\mathrm{d}\lambda}{\mathrm{d}\tau} + r\,\frac{\partial\vec{e}_r}{\partial\phi}\,\frac{\mathrm{d}\phi}{\mathrm{d}\tau} && (7.35d)\\
&\equiv \vec{e}_r\,\frac{\mathrm{d}r}{\mathrm{d}\tau} + \vec{e}_\lambda\,\frac{\mathrm{d}\lambda}{\mathrm{d}\tau} + \vec{e}_\phi\,\frac{\mathrm{d}\phi}{\mathrm{d}\tau} && (7.35e)\\
&= \vec{e}_r\,v^r + \vec{e}_\lambda\,v^\lambda + \vec{e}_\phi\,v^\phi. && (7.35f)
\end{aligned}$$

To reach this result we made use of the identities satisfied by the spherical basis vectors

$$\vec{e}_\lambda = r\,\frac{\partial\vec{e}_r}{\partial\lambda} \qquad \text{and} \qquad \vec{e}_\phi = r\,\frac{\partial\vec{e}_r}{\partial\phi}. \tag{7.36}$$

7.2.5 Metric tensor

The metric tensor for spherical coordinates takes on the diagonal form

$$g_{\overline{a}\overline{b}} = \vec{e}_{\overline{a}}\cdot\vec{e}_{\overline{b}} = \begin{bmatrix} (r\cos\phi)^2 & 0 & 0\\ 0 & r^2 & 0\\ 0 & 0 & 1 \end{bmatrix}, \tag{7.37}$$

as does the spherical representation of the inverse metric tensor

$$g^{\overline{a}\overline{b}} = \widetilde{e}^{\overline{a}}\cdot\widetilde{e}^{\overline{b}} = \begin{bmatrix} (r\cos\phi)^{-2} & 0 & 0\\ 0 & r^{-2} & 0\\ 0 & 0 & 1 \end{bmatrix}. \tag{7.38}$$

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 (7.37)) is given by

$$\sqrt{\det(\overline{\mathcal{G}})} = r^2\cos\phi \tag{7.39}$$

so that the volume element is

$$\mathrm{d}V = r^2\cos\phi\,\mathrm{d}r\,\mathrm{d}\lambda\,\mathrm{d}\phi. \tag{7.40}$$

The covariant Levi-Civita tensor has the spherical representation

$$\varepsilon_{\overline{a}\overline{b}\overline{c}} = (r^2\cos\phi)\,\epsilon_{\overline{a}\overline{b}\overline{c}}, \tag{7.41}$$

where $\epsilon_{\overline{a}\overline{b}\overline{c}}$ 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 (6.67) for the vector cross product of two basis vectors using spherical coordinates

$$\vec{e}_{\overline{a}} \wedge \vec{e}_{\overline{b}} = \varepsilon_{\overline{a}\overline{b}\overline{c}}\, \widetilde{e}^{\overline{c}} \Longrightarrow \vec{e}_{\overline{a}} \wedge \vec{e}_{\overline{b}} = (r^2 \cos\phi)\epsilon_{\overline{a}\overline{b}\overline{c}}\, \widetilde{e}^{\overline{c}}. \tag{7.42}$$

Making use of the spherical coordinate basis vectors and one-forms renders

$$\vec{e}_r \wedge \vec{e}_\lambda = (r\cos\phi)\,(\hat{\boldsymbol{r}} \wedge \hat{\boldsymbol{\lambda}}) = (r\cos\phi)\,\hat{\boldsymbol{\phi}} = (r^2\cos\phi)\,\widetilde{e}^{\phi} = \varepsilon_{r\lambda\phi}\,\widetilde{e}^{\phi} \tag{7.43a}$$

$$\vec{e}_\lambda \wedge \vec{e}_\phi = (r^2\cos\phi)\,(\hat{\boldsymbol{\lambda}} \wedge \hat{\boldsymbol{\phi}}) = (r^2\cos\phi)\,\hat{\boldsymbol{r}} = (r^2\cos\phi)\,\widetilde{e}^{r} = \varepsilon_{\lambda\phi r}\,\widetilde{e}^{r} \tag{7.43b}$$

$$\vec{e}_\phi \wedge \vec{e}_r = r\,(\hat{\boldsymbol{\phi}} \wedge \hat{\boldsymbol{r}}) = r\,\hat{\boldsymbol{\lambda}} = (r^2\cos\phi)\,\widetilde{e}^{\lambda} = \varepsilon_{\phi r\lambda}\,\widetilde{e}^{\lambda}. \tag{7.43c}$$

To reach these results we made use of the cross products for the spherical coordinate unit vectors

$$\hat{\boldsymbol{r}} \wedge \hat{\boldsymbol{\lambda}} = \hat{\boldsymbol{\phi}} \qquad \text{and} \qquad \hat{\boldsymbol{\lambda}} \wedge \hat{\boldsymbol{\phi}} = \hat{\boldsymbol{r}} \qquad \text{and} \qquad \hat{\boldsymbol{\phi}} \wedge \hat{\boldsymbol{r}} = \hat{\boldsymbol{\lambda}}. \tag{7.44}$$

7.2.6 Components of a vector field

A vector field $\vec{F}$ has Cartesian components, F^a, related to spherical components, $F^{\overline{a}}$, via the transformation matrix, $F^{\overline{a}} = \Lambda^{\overline{a}}_{\ a}\, F^a$. This transformation leads to

$$F^{\overline{1}} = (r\cos\phi)^{-1}\,[-F^x \sin\lambda + F^y \cos\lambda] \tag{7.45a}$$

$$F^{\overline{2}} = r^{-1}\,[-F^x \sin\phi\cos\lambda - F^y \sin\phi\sin\lambda + F^z\cos\phi] \tag{7.45b}$$

$$F^{\overline{3}} = F^x \cos\phi\cos\lambda + F^y\cos\phi\sin\lambda + F^z\sin\phi. \tag{7.45c}$$

Making use of the spherical unit vector (7.30a)-(7.30c) leads to the identities

$$(r\cos\phi)\,F^{\overline{1}} = \hat{\boldsymbol{\lambda}}\cdot\boldsymbol{F} \tag{7.46a}$$

$$r\,F^{\overline{2}} = \hat{\boldsymbol{\phi}}\cdot\boldsymbol{F} \tag{7.46b}$$

$$F^{\overline{3}} = \hat{\boldsymbol{r}}\cdot\boldsymbol{F}. \tag{7.46c}$$

7.2.7 Differential operators

In spherical coordinates the gradient operator $\nabla = \widetilde{e}^a \partial_a$ takes on the form

$$\nabla = \hat{\boldsymbol{\lambda}}\,(r\,\cos\phi)^{-1}\,\partial_\lambda + \hat{\boldsymbol{\phi}}\,r^{-1}\,\partial_\phi + \hat{\boldsymbol{r}}\,\partial_r \tag{7.47}$$

and the covariant divergence of a vector field is given by

$$\nabla_{\overline{a}} F^{\overline{a}} = (r^2\cos\phi)^{-1}\partial_{\overline{a}}[r^2\cos\phi\, F^{\overline{a}}] \tag{7.48a}$$

$$= (r^2\cos\phi)^{-1}\left(\partial_\lambda[r^2\cos\phi\, F^{\overline{1}}] + \partial_\phi[r^2\cos\phi\, F^{\overline{2}}] + \partial_r[r^2\cos\phi\, F^{\overline{3}}]\right) \tag{7.48b}$$

$$= \frac{1}{r\cos\phi}\,\frac{\partial(\hat{\boldsymbol{\lambda}}\cdot\boldsymbol{F})}{\partial\lambda} + \frac{1}{r\cos\phi}\,\frac{\partial\,(\hat{\boldsymbol{\phi}}\cdot\boldsymbol{F}\,\cos\phi)}{\partial\phi} + \frac{1}{r^2}\,\frac{\partial\,(\hat{\boldsymbol{r}}\cdot\boldsymbol{F}\,r^2)}{\partial r}. \tag{7.48c}$$

The covariant Laplacian of a scalar, $\nabla^2\psi = \nabla\cdot\nabla\psi$, is given by

$$\nabla^2\psi = \nabla\cdot\left[\hat{\boldsymbol{\lambda}}\,(r\cos\phi)^{-1}\,\partial_\lambda\psi + \hat{\boldsymbol{\phi}}\,r^{-1}\,\partial_\phi\psi + \hat{\boldsymbol{r}}\,\partial_r\psi\right] \tag{7.49a}$$

$$= \frac{1}{(r\cos\phi)^2}\,\frac{\partial^2\psi}{\partial\lambda^2} + \frac{1}{r^2\cos\phi}\,\frac{\partial}{\partial\phi}\left[\cos\phi\,\frac{\partial\psi}{\partial\phi}\right] + \frac{1}{r^2}\,\frac{\partial}{\partial r}\left[r^2\,\frac{\partial\psi}{\partial r}\right]. \tag{7.49b}$$

The covariant curl (Section 6.16) takes the form

$$(\mathrm{curl}\vec{F})^{\overline{1}} = (r^2\cos\phi)^{-1}\left[\partial_\phi F^{\overline{3}} - \partial_r\left(r^2 F^{\overline{2}}\right)\right] \tag{7.50a}$$

$$(\mathrm{curl}\vec{F})^{\overline{2}} = (r^2\cos\phi)^{-1}\left[\partial_r\left(r^2\cos^2\phi\, F^{\overline{1}}\right) - \partial_\lambda F^{\overline{3}}\right] \tag{7.50b}$$

$$(\mathrm{curl}\vec{F})^{\overline{3}} = (r^2\cos\phi)^{-1}\left[\partial_\lambda\left(r^2 F^{\overline{2}}\right) - \partial_\phi\left(r^2\cos^2\phi\, F^{\overline{1}}\right)\right], \tag{7.50c}$$

which can be written in the more conventional form (e.g., equation (2.33) of *Vallis* (2017))

$$r\cos\phi\,(\mathrm{curl}\vec{F})^{\overline{1}} = \frac{1}{r}\left[\frac{\partial(\hat{\boldsymbol{r}}\cdot\boldsymbol{F})}{\partial\phi} - \frac{\partial(r\,\hat{\boldsymbol{\phi}}\cdot\boldsymbol{F})}{\partial r}\right] \tag{7.51a}$$

$$r\,(\mathrm{curl}\vec{F})^{\overline{2}} = \frac{1}{r}\left[\frac{\partial(r\,\hat{\boldsymbol{\lambda}}\cdot\boldsymbol{F})}{\partial r} - \frac{1}{\cos\phi}\frac{\partial(\hat{\boldsymbol{r}}\cdot\boldsymbol{F})}{\partial\lambda}\right] \tag{7.51b}$$

$$(\mathrm{curl}\vec{F})^{\overline{3}} = \frac{1}{r\cos\phi}\left[\frac{\partial(\hat{\boldsymbol{\phi}}\cdot\boldsymbol{F})}{\partial\lambda} - \frac{\partial(\cos\phi\,\hat{\boldsymbol{\lambda}}\cdot\boldsymbol{F})}{\partial\phi}\right]. \tag{7.51c}$$

7.2.8 Summary of spherical coordinate expressions

We here summarize the spherical coordinate version of some common mathematical operators.

$$(\lambda,\phi,r) = (x^{\overline{1}}, x^{\overline{2}}, x^{\overline{3}}) \qquad \text{spherical coordinates} \tag{7.52}$$

$$(r\cos\phi)\,F^{\overline{1}} = \hat{\boldsymbol{\lambda}}\cdot\boldsymbol{F} \qquad r\,F^{\overline{2}} = \hat{\boldsymbol{\phi}}\cdot\boldsymbol{F} \qquad F^{\overline{3}} = \hat{\boldsymbol{r}}\cdot\boldsymbol{F} \qquad \text{vector components} \tag{7.53}$$

$$\nabla = \hat{\boldsymbol{\lambda}}\,(r\,\cos\phi)^{-1}\,\partial_\lambda + \hat{\boldsymbol{\phi}}\,r^{-1}\,\partial_\phi + \hat{\boldsymbol{r}}\,\partial_r \qquad \text{gradient operator} \tag{7.54}$$

$$\nabla_{\overline{a}}F^{\overline{a}} = \frac{1}{r\cos\phi}\frac{\partial(\hat{\boldsymbol{\lambda}}\cdot\boldsymbol{F})}{\partial\lambda} + \frac{1}{r\cos\phi}\frac{\partial\left(\hat{\boldsymbol{\phi}}\cdot\boldsymbol{F}\,\cos\phi\right)}{\partial\phi} + \frac{1}{r^2}\frac{\partial\left(\hat{\boldsymbol{r}}\cdot\boldsymbol{F}\,r^2\right)}{\partial r} \qquad \text{divergence of vector} \tag{7.55}$$

$$\nabla^2\psi = \frac{1}{(r\cos\phi)^2}\frac{\partial^2\psi}{\partial\lambda^2} + \frac{1}{r^2\cos\phi}\frac{\partial}{\partial\phi}\left[\cos\phi\,\frac{\partial\psi}{\partial\phi}\right] + \frac{1}{r^2}\frac{\partial}{\partial r}\left[r^2\,\frac{\partial\psi}{\partial r}\right] \qquad \text{Laplacian of scalar} \tag{7.56}$$

$$(\nabla\wedge\boldsymbol{F})_{\overline{a}} = \varepsilon_{\overline{a}\overline{b}\overline{c}}\partial_{\overline{b}}F^{\overline{c}} \qquad \text{see equations (7.51a) – (7.51c)} \qquad \text{curl of vector.} \tag{7.57}$$

7.3 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 7.2, here focusing on cylindrical-polar coordinates as shown in Figure 7.2. 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 7.2), which is distinct from the radial distance, r, used to measure the distance from the origin in spherical coordinates (Figure 7.1).

The coordinate transformation between Cartesian coordinates and cylindrical-polar coordinates is given by

$$x = r\cos\vartheta \equiv \xi^{\overline{1}}\cos\xi^{\overline{2}} \tag{7.58a}$$

$$y = r\sin\vartheta \equiv \xi^{\overline{1}}\sin\xi^{\overline{2}} \tag{7.58b}$$

$$z = \xi^{\overline{3}}. \tag{7.58c}$$

Figure 7.2: *This schematic illustrates the geometry and notation for cylindrical-polar coordinates. The Cartesian triad of orthonormal basis vectors, $(\hat{\boldsymbol{x}}, \hat{\boldsymbol{y}}, \hat{\boldsymbol{z}})$ points along the orthogonal axes. The cylindrical-polar triad of orthonormal basis vectors, $(\hat{\boldsymbol{r}}, \hat{\boldsymbol{\vartheta}}, \hat{\boldsymbol{z}})$, makes use of the radial unit vector $\hat{\boldsymbol{r}}$, which points outward from the vertical axis, the angular unit vector $\hat{\boldsymbol{\vartheta}}$, which points in the counter-clockwise direction around the circle, and the vertical unit vector $\hat{\boldsymbol{z}}$. Note that the radial unit vector used for cylindrical-polar coordinates is distinct from that radial vector used in spherical coordinates shown in Figure 7.1.*

The radial coordinate for cylindrical-polar coordinates

$$r = \sqrt{x^2 + y^2} \tag{7.59}$$

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

$$(x, y, z) = (\xi^1, \xi^2, \xi^3) \equiv \xi^a \qquad \text{and} \qquad (r, \vartheta, z) = (\xi^{\bar{1}}, \xi^{\bar{2}}, \xi^{\bar{3}}) \equiv \xi^{\bar{a}}. \tag{7.60}$$

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 symbol ξ^3 and $\xi^{\bar{3}}$ to specify what other coordinates are held fixed when performing derivative operations.

7.3.1 Transforming between Cartesian and cylindrical-polar coordinates

The coordinate relation (7.58a)-(7.58c) leads to the transformation matrix

$$\Lambda^a{}_{\bar{a}} = \begin{bmatrix} \partial\xi^1/\partial\xi^{\bar{1}} & \partial\xi^1/\partial\xi^{\bar{2}} & \partial\xi^1/\partial\xi^{\bar{3}} \\ \partial\xi^2/\partial\xi^{\bar{1}} & \partial\xi^2/\partial\xi^{\bar{2}} & \partial\xi^2/\partial\xi^{\bar{3}} \\ \partial\xi^3/\partial\xi^{\bar{1}} & \partial\xi^3/\partial\xi^{\bar{2}} & \partial\xi^3/\partial\xi^{\bar{3}} \end{bmatrix} = \begin{bmatrix} \cos\vartheta & -r\sin\vartheta & 0 \\ \sin\vartheta & r\cos\vartheta & 0 \\ 0 & 0 & 1 \end{bmatrix} \tag{7.61}$$

and the inverse transformation is given by

$$\Lambda^{\bar{a}}{}_a = \frac{1}{r} \begin{bmatrix} r\cos\vartheta & r\sin\vartheta & 0 \\ -\sin\vartheta & \cos\vartheta & 0 \\ 0 & 0 & r \end{bmatrix}. \tag{7.62}$$

The determinant of the transformation (Jacobian) is given by

$$\det(\Lambda^{a}_{\overline{a}}) = r, \tag{7.63}$$

which vanishes along the vertical axis, which is where the transformation is singular.

7.3.2 Basis vectors

The cylindrical-polar coordinate basis vectors, $\vec{e}_{\overline{a}}$, are related to the Cartesian coordinate basis vectors, $\vec{e}_a$, through the transformation $\vec{e}_{\overline{a}} = \Lambda^{a}_{\overline{a}}\,\vec{e}_a$. The transformation matrix (7.61) leads to

$$\vec{e}_r = \hat{\boldsymbol{x}}\cos\vartheta + \hat{\boldsymbol{y}}\sin\vartheta \tag{7.64a}$$

$$\vec{e}_\vartheta = r\,(-\hat{\boldsymbol{x}}\sin\vartheta + \hat{\boldsymbol{y}}\cos\vartheta) \tag{7.64b}$$

$$\vec{e}_{\overline{z}} = \hat{\boldsymbol{z}}. \tag{7.64c}$$

We sometimes make use of the following orthonormal unit vectors $(\hat{\boldsymbol{r}}, \hat{\boldsymbol{\vartheta}}, \hat{\boldsymbol{z}})$

$$\vec{e}_r = \hat{\boldsymbol{r}} \qquad \text{and} \qquad \vec{e}_\vartheta = r\,\hat{\boldsymbol{\vartheta}} \qquad \text{and} \qquad \vec{e}_{\overline{z}} = \hat{\boldsymbol{z}} \tag{7.65}$$

along with the inverse relations

$$\hat{\boldsymbol{x}} = \hat{\boldsymbol{r}}\cos\vartheta - \hat{\boldsymbol{\vartheta}}\sin\vartheta \tag{7.66a}$$

$$\hat{\boldsymbol{y}} = \hat{\boldsymbol{r}}\sin\vartheta + \hat{\boldsymbol{\vartheta}}\cos\vartheta \tag{7.66b}$$

$$\hat{\boldsymbol{z}} = \hat{\boldsymbol{z}}. \tag{7.66c}$$

7.3.3 Basis one-forms

Since cylindrical-polar coordinates are orthogonal, we can readily derive the one-form basis through inverting the vector basis

$$\widetilde{e}^r = \hat{\boldsymbol{r}} \qquad \text{and} \qquad \widetilde{e}^\vartheta = r^{-1}\hat{\boldsymbol{\vartheta}} \qquad \text{and} \qquad \widetilde{e}^{\overline{z}} = \hat{\boldsymbol{z}}, \tag{7.67}$$

which satisfy the orthogonality relation (Section 6.2.2)

$$\widetilde{e}^{\overline{b}} \cdot \vec{e}_{\overline{a}} = \delta^{\overline{b}}_{\overline{a}}. \tag{7.68}$$

7.3.4 Position and velocity

In cylindrical-polar coordinates, the position of a point is specified by the radial position plus the vertical position

$$\mathscr{P}(\tau) = r\,\vec{e}_r + z\,\vec{e}_{\overline{z}}. \tag{7.69}$$

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

$$\vec{v}(\tau) = \frac{\mathrm{d}\mathscr{P}}{\mathrm{d}\tau} \tag{7.70a}$$

$$= \vec{e}_r\,\frac{\mathrm{d}r}{\mathrm{d}\tau} + r\,\frac{\mathrm{d}\vec{e}_r}{\mathrm{d}\tau} + \vec{e}_{\overline{z}}\,\frac{\mathrm{d}z}{\mathrm{d}\tau} \tag{7.70b}$$

$$= \vec{e}_r\,\frac{\mathrm{d}r}{\mathrm{d}\tau} + r\,\frac{\partial\vec{e}_r}{\partial\vartheta}\,\frac{\mathrm{d}\vartheta}{\mathrm{d}\tau} + +\vec{e}_{\overline{z}}\,\frac{\mathrm{d}z}{\mathrm{d}\tau} \tag{7.70c}$$

$$= \vec{e}_r\,\frac{\mathrm{d}r}{\mathrm{d}\tau} + \vec{e}_\vartheta\,\frac{\mathrm{d}\vartheta}{\mathrm{d}\tau} + \vec{e}_{\overline{z}}\,\frac{\mathrm{d}z}{\mathrm{d}\tau} \tag{7.70d}$$

$$= \vec{e}_r\,v^r + \vec{e}_\vartheta\,v^\vartheta + \vec{e}_{\overline{z}}\,v^{\overline{z}}. \tag{7.70e}$$

To reach this result we made use of the identity

$$\vec{e}_\vartheta = r\,\frac{\partial \vec{e}_r}{\partial \vartheta}. \tag{7.71}$$

7.3.5 Metric tensor

Cylindrical-polar coordinates are orthogonal so that the metric tensor and its inverse are represented by the diagonal matrices

$$g_{\overline{a}\overline{b}} = \vec{e}_{\overline{a}} \cdot \vec{e}_{\overline{b}} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & r^2 & 0 \\ 0 & 0 & 1 \end{bmatrix} \qquad \text{and} \qquad g^{\overline{a}\overline{b}} = \widetilde{e}^{\overline{a}} \cdot \widetilde{e}^{\overline{b}} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & r^{-2} & 0 \\ 0 & 0 & 1 \end{bmatrix}. \tag{7.72}$$

Volume element and Levi-Civita tensor

The square root of the determinant of the metric tensor written in cylindrical-polar coordinates (from equation (7.72)) is given by

$$\sqrt{\det(\overline{\mathcal{G}})} = r \tag{7.73}$$

so that the volume element is

$$\mathrm{d}V = r\,\mathrm{d}r\,\mathrm{d}\vartheta\,\mathrm{d}\overline{z}. \tag{7.74}$$

The covariant Levi-Civita tensor has the cylindrical-polar representation

$$\varepsilon_{\overline{a}\overline{b}\overline{c}} = r\,\epsilon_{\overline{a}\overline{b}\overline{c}}. \tag{7.75}$$

Vector cross product of basis vectors

As a check on the formalism for vector cross products, let us verify the relation (6.67) for the cross product of two basis vectors using cylindrical-polar coordinates

$$\vec{e}_{\overline{a}} \wedge \vec{e}_{\overline{b}} = \varepsilon_{\overline{a}\overline{b}\overline{c}}\,\widetilde{e}^{\overline{c}} \Longrightarrow \vec{e}_{\overline{a}} \wedge \vec{e}_{\overline{b}} = r\,\epsilon_{\overline{a}\overline{b}\overline{c}}\,\widetilde{e}^{\overline{c}}. \tag{7.76}$$

Making use of the cylindrical-polar coordinate basis vectors and one-forms renders

$$\vec{e}_r \wedge \vec{e}_\vartheta = r\,(\hat{\boldsymbol{r}} \wedge \hat{\boldsymbol{\vartheta}}) = r\,\widetilde{e}^{\overline{z}} = \varepsilon_{r\vartheta\overline{z}}\,\widetilde{e}^{\overline{z}} \tag{7.77a}$$

$$\vec{e}_\vartheta \wedge \vec{e}_{\overline{z}} = r\,(\hat{\boldsymbol{\vartheta}} \wedge \hat{\boldsymbol{z}}) = r\,\hat{\boldsymbol{r}} = r\,\widetilde{e}^{r} = \varepsilon_{\vartheta\overline{z}r}\,\widetilde{e}^{r} \tag{7.77b}$$

$$\vec{e}_{\overline{z}} \wedge \vec{e}_r = \hat{\boldsymbol{z}} \wedge \hat{\boldsymbol{r}} = \hat{\boldsymbol{\vartheta}} = r\,\widetilde{e}^{\vartheta} = \varepsilon_{\overline{z}r\vartheta}\,\widetilde{e}^{\vartheta}. \tag{7.77c}$$

To reach these results we made use of the cross products for the unit vectors

$$\hat{\boldsymbol{r}} \wedge \hat{\boldsymbol{\vartheta}} = \hat{\boldsymbol{z}} \qquad \text{and} \qquad \hat{\boldsymbol{\vartheta}} \wedge \hat{\boldsymbol{z}} = \hat{\boldsymbol{r}} \qquad \text{and} \qquad \hat{\boldsymbol{z}} \wedge \hat{\boldsymbol{r}} = \hat{\boldsymbol{\vartheta}}. \tag{7.78}$$

7.3.6 Components of a vector field

A vector field $\vec{F}$ has Cartesian components, F^a, related to cylindrical-polar components, $F^{\overline{a}}$, via the transformation matrix, $F^{\overline{a}} = \Lambda^{\overline{a}}{}_a\,F^a$. This transformation leads to

$$F^{\overline{1}} = F^x \cos\vartheta + F^y \sin\vartheta \tag{7.79a}$$

$$F^{\overline{2}} = r^{-1}\left[-F^x \sin\vartheta + F^y \cos\vartheta\right] \tag{7.79b}$$

$$F^{\overline{3}} = F^z. \tag{7.79c}$$

Introducing the cylindrical-polar unit vectors (7.65) leads to

$$F^{\bar{1}} = \hat{\boldsymbol{r}} \cdot \boldsymbol{F} \tag{7.80a}$$

$$r\, F^{\bar{2}} = \hat{\boldsymbol{\vartheta}} \cdot \boldsymbol{F} \tag{7.80b}$$

$$F^{\bar{3}} = \hat{\boldsymbol{z}} \cdot \boldsymbol{F}. \tag{7.80c}$$

7.3.7 Differential operators

In cylindrical-polar coordinates, the gradient operator $\nabla = \widetilde{e}^a \partial_a$ takes on the form

$$\nabla = \hat{\boldsymbol{r}}\, \frac{\partial}{\partial r} + \frac{\hat{\boldsymbol{\vartheta}}}{r}\, \frac{\partial}{\partial \vartheta} + \hat{\boldsymbol{z}}\, \frac{\partial}{\partial z} \tag{7.81}$$

and the covariant divergence of a vector field is given by

$$\nabla_{\bar{a}} F^{\bar{a}} = r^{-1}\, \partial_{\bar{a}}\, (r\, F^{\bar{a}}) \tag{7.82a}$$

$$= r^{-1} \left(\partial_r [r\, F^{\bar{1}}] + \partial_\vartheta [r\, F^{\bar{2}}] + \partial_{\bar{z}} [r\, F^{\bar{3}}] \right) \tag{7.82b}$$

$$= \frac{1}{r}\, \frac{\partial (r\, \hat{\boldsymbol{r}} \cdot \boldsymbol{F})}{\partial r} + \frac{1}{r}\, \frac{\partial\, (\hat{\boldsymbol{\vartheta}} \cdot \boldsymbol{F})}{\partial \vartheta} + \frac{\partial (\hat{\boldsymbol{z}} \cdot \boldsymbol{F})}{\partial z}. \tag{7.82c}$$

The covariant Laplacian of a scalar, $\nabla^2 \psi = \nabla \cdot \nabla \psi$, is given by

$$\nabla^2 \psi = \nabla \cdot \left[\hat{\boldsymbol{r}}\, \frac{\partial \psi}{\partial r} + \frac{\hat{\boldsymbol{\vartheta}}}{r}\, \frac{\partial \psi}{\partial \vartheta} + \hat{\boldsymbol{z}}\, \frac{\partial \psi}{\partial z} \right] \tag{7.83a}$$

$$= \frac{1}{r}\, \frac{\partial}{\partial r} \left[r\, \frac{\partial \psi}{\partial r} \right] + \frac{1}{r^2}\, \frac{\partial^2 \psi}{\partial \vartheta^2} + \frac{\partial^2 \psi}{\partial z^2}. \tag{7.83b}$$

The covariant curl (Section 6.16) takes the form

$$(\mathrm{curl} \vec{F})^{\bar{1}} = r^{-1}\, [\partial_\vartheta\, F^{\bar{3}} - \partial_{\bar{z}}\, (r^2\, F^{\bar{2}})] \tag{7.84a}$$

$$(\mathrm{curl} \vec{F})^{\bar{2}} = r^{-1}\, [\partial_{\bar{z}}\, F^{\bar{1}} - \partial_r\, F^{\bar{3}}] \tag{7.84b}$$

$$(\mathrm{curl} \vec{F})^{\bar{3}} = r^{-1}\, [\partial_r\, (r^2\, F^{\bar{2}}) - \partial_\vartheta\, F^{\bar{1}})], \tag{7.84c}$$

which can be written more conventionally as

$$(\mathrm{curl} \vec{F})^{\bar{1}} = \frac{1}{r}\, \frac{\partial (\hat{\boldsymbol{z}} \cdot \boldsymbol{F})}{\partial \vartheta} - \frac{\partial (\hat{\boldsymbol{\vartheta}} \cdot \boldsymbol{F})}{\partial z} \tag{7.85a}$$

$$r\, (\mathrm{curl} \vec{F})^{\bar{2}} = \frac{\partial (\hat{\boldsymbol{r}} \cdot \boldsymbol{F})}{\partial z} - \frac{\partial (\hat{\boldsymbol{z}} \cdot \boldsymbol{F})}{\partial r} \tag{7.85b}$$

$$(\mathrm{curl} \vec{F})^{\bar{3}} = \frac{1}{r}\, \frac{\partial (r\, \hat{\boldsymbol{\vartheta}} \cdot \boldsymbol{F})}{\partial r} - \frac{1}{r}\, \frac{\partial (\hat{\boldsymbol{r}} \cdot \boldsymbol{F})}{\partial \vartheta}. \tag{7.85c}$$

7.3.8 Summary of cylindrical coordinate expressions

We here summarize the cylindrical coordinate version of some common mathematical operators.

$$(r, \vartheta, z) = (x^{\overline{1}}, x^{\overline{2}}, x^{\overline{3}}) \qquad \text{cylindrical coordinates} \tag{7.86}$$

$$F^{\overline{1}} = \hat{\boldsymbol{r}} \cdot \boldsymbol{F} \qquad r\, F^{\overline{2}} = \hat{\boldsymbol{\vartheta}} \cdot \boldsymbol{F} \qquad F^{\overline{3}} = \hat{\boldsymbol{z}} \cdot \boldsymbol{F} \qquad \text{vector components} \tag{7.87}$$

$$\nabla = \hat{\boldsymbol{r}}\, \frac{\partial}{\partial r} + \frac{\hat{\boldsymbol{\vartheta}}}{r}\, \frac{\partial}{\partial \vartheta} + \hat{\boldsymbol{z}}\, \frac{\partial}{\partial z} \qquad \text{gradient} \tag{7.88}$$

$$\nabla_{\overline{a}} F^{\overline{a}} = \frac{1}{r}\, \frac{\partial (r\, \hat{\boldsymbol{r}} \cdot \boldsymbol{F})}{\partial r} + \frac{1}{r}\, \frac{\partial\, (\hat{\boldsymbol{\vartheta}} \cdot \boldsymbol{F})}{\partial \vartheta} + \frac{\partial (\hat{\boldsymbol{z}} \cdot \boldsymbol{F})}{\partial z}. \qquad \text{divergence} \tag{7.89}$$

$$\nabla^2 \psi = \frac{1}{r}\, \frac{\partial}{\partial r} \left[r\, \frac{\partial \psi}{\partial r} \right] + \frac{1}{r^2}\, \frac{\partial^2 \psi}{\partial \vartheta^2} + \frac{\partial^2 \psi}{\partial z^2} \qquad \text{Laplacian of scalar} \tag{7.90}$$

$$(\nabla \wedge \boldsymbol{F})_{\overline{a}} = \varepsilon_{\overline{a}\overline{b}\overline{c}} \partial_{\overline{b}} F^{\overline{c}} \qquad \text{see equations (7.84a)} - \text{(7.84c)}. \qquad \text{curl of a vector} \tag{7.91}$$

7.4 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

$$g_{\overline{a}\overline{b}} = \vec{e}_{\overline{a}} \cdot \vec{e}_{\overline{b}} = \begin{bmatrix} h_{\overline{1}} & 0 & 0 \\ 0 & h_{\overline{2}} & 0 \\ 0 & 0 & h_{\overline{3}} \end{bmatrix}, \tag{7.92}$$

where $h_{\overline{a}} > 0$ are "stretching" functions. The corresponding volume element is expressed as

$$\mathrm{d}V = h_{\overline{1}}\, h_{\overline{2}}\, h_{\overline{3}}\, \mathrm{d}\xi^{\overline{1}}\, \mathrm{d}\xi^{\overline{2}}\, \mathrm{d}\xi^{\overline{3}}. \tag{7.93}$$

These *generalized orthogonal curvilinear coordinates* have a corresponding orthogonal set of basis vectors

$$\vec{e}_{\overline{a}} = h_{\overline{a}}\, \hat{\mathbf{e}}_{(\overline{a})} \qquad \text{no implied sum.} \tag{7.94}$$

The objects $\hat{\mathbf{e}}_{(\overline{a})}$ are the dimensionless unit directions. The corresponding one-form basis is given by

$$\widetilde{e}^{\overline{a}} = (h_{\overline{a}})^{-1}\, \hat{\mathbf{e}}_{(\overline{a})}. \tag{7.95}$$

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 $h_{\overline{a}}$ carry the tensorial properties of the basis vectors $\vec{e}_{\overline{a}}$. Results for the trajectory and velocity are straightforward generalizations of the spherical results in Section 7.2 and cylindrical-polar results from Section 7.3. A thorough presentation of generalized orthogonal coordinates is found in Section 21.11 of *Griffies* (2004).

8

General vertical coordinates

General vertical coordinates (GVCs) offer a mathematical framework for a variety of topics in stratified fluid mechanics. Hence, they appear in many guises throughout this book. Notably, their non-orthogonality requires extra care beyond that needed with the orthogonal coordinates considered in Chapter 7, as does their time dependence. GVCs were introduced by *Starr* (1945) for atmospheric modeling and for ocean modeling by *Bleck* (1978). There is a growing use of GVC-based numerical ocean and atmospheric models, prompting the need to master their use for analysis, model formulation, and theory.

The key reason to favor the non-orthogonal GVCs, rather than a locally orthogonal set of coordinates, is that gravity plays a hugely dominant role in orienting geophysical fluid motion. Hence, it is prudent 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.

In contrast, 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 49.3.

READER'S GUIDE TO THIS CHAPTER

We make use of the general tensor analysis detailed in Chapter 6. We mostly consider just the spatial tensors in this chapter, consistent with the Newtonian perspective whereby time is universal. Particular chapters directly relying on the material from this chapter include the fluid kinematics discussed in Chapter 17, the general vertical coordinate dynamics discussed in Chapter 38, and the tracer equation diffusion and stirring operators discussed in Chapter 49,

8.1 Introducing GVCs **116**

8.1.1 Relating Cartesian and generalized vertical coordinates 117
8.1.2 A common confusion 118
8.2 Spatial basis vectors **118**
8.2.1 A few more points on the transformation matrix 118
8.2.2 Expressions for the basis vectors 119
8.3 Basis one-forms **119**
8.3.1 A few more points on the inverse transformation matrix 119
8.3.2 GVC basis one-forms 120
8.3.3 Verifying the orthogonality relation 120
8.4 Triple product identities **120**
8.5 Position vector **121**
8.6 Transforming components of a first order tensor **123**
8.7 Velocity **123**
8.7.1 Contravariant components to the velocity vector 124
8.7.2 Covariant components to the velocity vector 124
8.7.3 Introducing the material time derivative 124
8.7.4 Equivalence to the Cartesian velocity representation 124
8.8 Metric tensor **125**
8.8.1 Jacobian of transformation 125
8.8.2 Covariant and contravariant representations 125
8.9 Volume element and the Levi-Civita tensor **126**
8.10 Vector cross product of basis vectors **126**
8.11 Partial derivative operators **126**
8.11.1 Analytical derivation 126
8.11.2 Geometrical derivation 127
8.12 Material time derivative **128**
8.13 Divergence of a vector and the divergence theorem **128**
8.14 The diffusion operator **129**
8.14.1 Continuous expression 129
8.14.2 Layer thickness weighted diffusion operator 130
8.15 Vorticity **130**
8.15.1 The components 131
8.15.2 Transforming from Cartesian coordinates 131
8.16 Velocity circulation **131**

8.1 Introducing GVCs

In this chapter we present the mathematics of generalalized vertical coordinates (GVC), with Figure 8.1 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, pressure is a convenient choice since it naturally absorbs the appearance of density in many formula, such as mass continuity as discussed in Section 17.9.2 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 8.1: *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, $\hat{\boldsymbol{n}}$, always has a non-zero projection onto the vertical: $\hat{\boldsymbol{n}} \cdot \hat{\boldsymbol{z}} \neq 0$. That is, the surfaces never become vertical nor do they overturn. It also means that there is a one-to-one invertible relation between σ and geopotential, so that specifying $(x, y, \sigma(x, y, z, t))$ is sufficient to yield a unique z.*

8.1.1 Relating Cartesian and generalized vertical coordinates

We make use of the symbol σ for a generalalized 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

$$\xi^\alpha = (\xi^0, \xi^a) = (\xi^0, \xi^1, \xi^2, \xi^3) = (t, x, y, z) \text{ with } \alpha = 0, 1, 2, 3, \text{ and } a = 1, 2, 3. \tag{8.1}$$

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

$$\xi^{\overline{\alpha}} = (\xi^{\overline{0}}, \xi^{\overline{1}}, \xi^{\overline{2}}, \xi^{\overline{3}}) = (\overline{t}, \overline{x}, \overline{y}, \sigma). \tag{8.2}$$

The one-to-one coordinate transformation between Cartesian and GVC coordinates is written

$$\xi^{\overline{0}} = \xi^0 \Longleftrightarrow \overline{t} = t \tag{8.3a}$$

$$\xi^{\overline{1}} = \xi^1 \Longleftrightarrow \overline{x} = x \tag{8.3b}$$

$$\xi^{\overline{2}} = \xi^2 \Longleftrightarrow \overline{y} = y \tag{8.3c}$$

$$\xi^{\overline{3}} = \sigma(t, x, y, z). \tag{8.3d}$$

Since the coordinate transformation is invertible, we can define the inverse of equation (8.3d), thus providing an expression for the vertical position of a given GVC surface

$$\xi^3 = \xi^3(\xi^{\overline{a}}) = z(\overline{t}, \overline{x}, \overline{y}, \sigma). \tag{8.4}$$

This equation says that when we locate a position via the value of a chosen GVC surface (i.e., specify the value for σ), the geopotential position for that surface is a function of time, horizontal

position, and the chosen value for the generalized vertical coordinate. For example, the vertical position of a pressure surface of chosen value p is given by the functional relation $\xi^3 = z(\overline{t}, \overline{x}, \overline{y}, p)$. The full inverse coordinate transformation takes the form

$$\xi^0 = \xi^{\overline{0}} \tag{8.5a}$$

$$\xi^1 = \xi^{\overline{1}} \tag{8.5b}$$

$$\xi^2 = \xi^{\overline{2}} \tag{8.5c}$$

$$\xi^3 = z(\overline{t}, \overline{x}, \overline{y}, \sigma). \tag{8.5d}$$

8.1.2 A common confusion

The above discussion exposes a means for confusion. Namely, it is common in the literature to switch between using the symbol z to mean a particular geopotential; i.e., $z = -100$ m, versus the symbol z as the geopotential position of a particular σ surface, $z(\overline{t}, \overline{x}, \overline{y}, \sigma)$. One way to be careful is to always write $\xi^3(\overline{t}, \overline{x}, \overline{y}, \sigma)$ rather than $z(\overline{x}, \overline{y}, \sigma, \overline{t})$. We maintain care in places where it is crucial to keep the meaning clear. However, developing a GVC brain muscle allows us to be relaxed when the meaning is clear (or should be clear).

8.2 Spatial basis vectors

Making use of the tensor formalism from Chapter 6, consider the transformation of the Cartesian basis vectors into their corresponding GVC representation. This transformation takes is given by

$$\vec{e}_{\overline{a}} = \Lambda^a_{\overline{a}} \, \vec{e}_a, \tag{8.6}$$

where the transformation matrix is

$$\Lambda^a_{\overline{a}} = \begin{bmatrix} \partial x/\partial \overline{x} & \partial x/\partial \overline{y} & \partial x/\partial \sigma \\ \partial y/\partial \overline{x} & \partial y/\partial \overline{y} & \partial y/\partial \sigma \\ \partial z/\partial \overline{x} & \partial z/\partial \overline{y} & \partial z/\partial \sigma \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ \partial z/\partial \overline{x} & \partial z/\partial \overline{y} & \partial z/\partial \sigma \end{bmatrix}. \tag{8.7}$$

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 $\overline{x}$, $\overline{y}$, 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 $\partial z/\partial \overline{t}$ arises since we generally change the vertical position when following the temporal motion of a constant GVC surface. Likewise, $\partial z/\partial \overline{x}$ and $\partial z/\partial \overline{y}$ are nonzero since we generally change vertical position when moving horizontally along a sloped GVC surface. Finally, the element $\partial z/\partial \sigma$ is nonzero due to vertical stratification of the fluid when represented using GVCs.

8.2.1 A few more points on the transformation matrix

To further detail how to produce elements of the transformation matrix (8.7), 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

$$\Lambda^1_{\overline{a}} = \begin{bmatrix} [\partial x/\partial \overline{x}]_{\overline{y},\sigma} & [\partial x/\partial \overline{y}]_{\overline{x},\sigma} & [\partial x/\partial \sigma]_{\overline{x},\overline{y}} \end{bmatrix} \tag{8.8}$$

Since $x = \overline{x}$, all elements vanish except for the first. Namely, $[\partial x/\partial \overline{y}]_{\overline{x},\sigma} = 0$ since x cannot change when $\overline{x}$ is fixed. The same idea leads to the results for y.

8.2.2 Expressions for the basis vectors

Use of the transformation matrix (8.7) renders the spatial components of the GVC basis vectors

$$\vec{e}_{\overline{1}} = \hat{\boldsymbol{x}} + \hat{\boldsymbol{z}}\,(\partial z/\partial \overline{x}) \tag{8.9a}$$
$$\vec{e}_{\overline{2}} = \hat{\boldsymbol{y}} + \hat{\boldsymbol{z}}\,(\partial z/\partial \overline{y}) \tag{8.9b}$$
$$\vec{e}_{\overline{3}} = \hat{\boldsymbol{z}}\,(\partial z/\partial \sigma). \tag{8.9c}$$

The basis vectors $\vec{e}_{\overline{1}}$ and $\vec{e}_{\overline{2}}$ have a vertical component due to sloping GVC surfaces. These basis vectors lie within the tangent plane of the GVC surface. The basis vector $\vec{e}_{\overline{3}}$ is purely vertical and has a non-unit magnitude due to the inverse vertical stratification, $\partial z/\partial \sigma$. The left panel of Figure 8.2 illustrates the basis vectors.

Figure 8.2: *Illustrating the basis vectors (left panel) and basis one-forms (right panel) for generalized vertical coordinates. The $\vec{e}_{\overline{3}}$ basis vector is vertical whereas $\vec{e}_{\overline{1}}$ and $\vec{e}_{\overline{2}}$ lie within the tangent plane to the σ surface. As a complement, the basis one-form $\widetilde{e}^{\overline{3}}$ is normal to the σ surface whereas the basis one-forms $\widetilde{e}^{\overline{1}}$ and $\widetilde{e}^{\overline{2}}$ are horizontal.*

8.3 Basis one-forms

The basis one-forms are obtained by transforming from Cartesian into GVCs through use of the inverse transformation

$$\widetilde{e}^{\overline{a}} = \Lambda^{\overline{a}}{}_{a}\,\widetilde{e}^{a}, \tag{8.10}$$

where the inverse transformation matrix takes the form

$$\Lambda^{\overline{a}}{}_{a} = \begin{bmatrix} \partial\overline{x}/\partial x & \partial\overline{x}/\partial y & \partial\overline{x}/\partial z \\ \partial\overline{y}/\partial x & \partial\overline{y}/\partial y & \partial\overline{y}/\partial z \\ \partial\sigma/\partial x & \partial\sigma/\partial y & \partial\sigma/\partial z \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ \partial\sigma/\partial x & \partial\sigma/\partial y & \partial\sigma/\partial z \end{bmatrix}. \tag{8.11}$$

As for the transformation matrix (8.7), the unit diagonal values arise since a horizontal position in Cartesian and GVCs is the same and the horizontal directions are orthogonal. Likewise, $\partial\overline{x}/\partial z = \partial\overline{y}/\partial z = 0$ since the horizontal position on a GVC surface remains unchanged when moving across a depth surface. Finally, 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.

8.3.1 A few more points on the inverse transformation matrix

When computing elements of the inverse transformation matrix (8.11), it is crucial to ensure that the proper variables are held fixed. For example, consider the top row where we compute

$$\Lambda^{\overline{1}}{}_{a} = \begin{bmatrix} [\partial\overline{x}/\partial x]_{y,z} & [\partial\overline{x}/\partial y]_{x,z} & [\partial\overline{x}/\partial z]_{x,y} \end{bmatrix}. \tag{8.12}$$

Just as for the transformation matrix (8.8), since $x = \overline{x}$, all but the first element vanish in equation (8.12). Namely, $[\partial\overline{x}/\partial y]_{x,z} = 0$ since the $\overline{x}$ cannot change when x is fixed. The same idea holds for the $\overline{y}$ row.

8.3.2 GVC basis one-forms

Use of the inverse transformation matrix (8.11) renders the spatial components of the GVC basis one-forms

$$\widetilde{e}^{\overline{1}} = \hat{\boldsymbol{x}} \tag{8.13a}$$

$$\widetilde{e}^{\overline{2}} = \hat{\boldsymbol{y}} \tag{8.13b}$$

$$\widetilde{e}^{\overline{3}} = \widetilde{e}^{a}\,\partial_a\sigma = \hat{\boldsymbol{x}}\,(\partial\sigma/\partial x) + \hat{\boldsymbol{y}}\,(\partial\sigma/\partial y) + \hat{\boldsymbol{z}}\,(\partial\sigma/\partial z) = \nabla\sigma. \tag{8.13c}$$

The left panel of Figure 8.2 illustrates the basis one-forms.

8.3.3 Verifying the orthogonality relation

The basis one-forms satisfy the orthogonality relation (6.25) with the basis vectors

$$\widetilde{e}^{\overline{a}} \cdot \vec{e}_{\overline{b}} = \delta^{\overline{a}}_{\overline{b}}. \tag{8.14}$$

This identity is trivial to verify for all $\overline{a} = 1, 2, 3$.

8.4 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 6.8.3

$$\sigma = \sigma(t, x, y, z) = \sigma[t, x, y, z(\overline{t}, \overline{x}, \overline{y}, \sigma)], \tag{8.15}$$

so that the chain rule leads to the differential increment

$$\mathrm{d}\sigma = \mathrm{d}t\left[\frac{\partial\sigma}{\partial t}\right]_{x,y,z} + \mathrm{d}x\left[\frac{\partial\sigma}{\partial x}\right]_{t,y,z} + \mathrm{d}y\left[\frac{\partial\sigma}{\partial y}\right]_{t,x,z} + \mathrm{d}z\left[\frac{\partial\sigma}{\partial z}\right]_{t,x,y}. \tag{8.16}$$

Likewise, writing $z = z[\overline{t}, \overline{x}, \overline{y}, \sigma]$ leads to the differential increment $\mathrm{d}z$

$$\mathrm{d}z = \mathrm{d}\overline{t}\left[\frac{\partial z}{\partial \overline{t}}\right]_{\overline{x},\overline{y},\sigma} + \mathrm{d}\overline{x}\left[\frac{\partial z}{\partial \overline{x}}\right]_{\overline{t},\overline{y},\sigma} + \mathrm{d}\overline{y}\left[\frac{\partial z}{\partial \overline{y}}\right]_{\overline{t},\overline{x},\sigma} + \mathrm{d}\sigma\left[\frac{\partial z}{\partial \sigma}\right]_{\overline{t},\overline{x},\overline{y}}. \tag{8.17}$$

We note the identities

$$\left[\frac{\partial\sigma}{\partial z}\right]_{t,x,y}\left[\frac{\partial z}{\partial\sigma}\right]_{\overline{t},\overline{x},\overline{y}} = 1 \qquad \mathrm{d}\overline{t} = \mathrm{d}t \qquad \mathrm{d}\overline{x} = \mathrm{d}x \qquad \mathrm{d}\overline{y} = \mathrm{d}y, \tag{8.18}$$

which follow since $t = \overline{t}$, $x = \overline{x}$, and $y = \overline{y}$. Substituting equation (8.17) into equation (8.16) and making use of the identities (8.18) yields

$$\begin{aligned} 0 = \mathrm{d}t &\left[\left[\frac{\partial\sigma}{\partial t}\right]_{x,y,z} + \left[\frac{\partial\sigma}{\partial z}\right]_{t,x,y}\left[\frac{\partial z}{\partial\overline{t}}\right]_{\overline{x},\overline{y},\sigma}\right] \\ &+ \mathrm{d}x\left[\left[\frac{\partial\sigma}{\partial x}\right]_{t,y,z} + \left[\frac{\partial\sigma}{\partial z}\right]_{t,x,y}\left[\frac{\partial z}{\partial\overline{x}}\right]_{\overline{t},\overline{y},\sigma}\right] + \mathrm{d}y\left[\left[\frac{\partial\sigma}{\partial y}\right]_{t,x,z} + \left[\frac{\partial\sigma}{\partial z}\right]_{t,x,y}\left[\frac{\partial z}{\partial\overline{y}}\right]_{\overline{t},\overline{x},\sigma}\right]. \end{aligned} \tag{8.19}$$

For this equation to hold with general increments $\mathrm{d}t$, $\mathrm{d}x$, and $\mathrm{d}y$ requires that each bracketed term vanish, which in turn leads to the following set of triple product identities[1]

$$\left[\frac{\partial\sigma}{\partial z}\right]_{t,x,y}\left[\frac{\partial z}{\partial\overline{t}}\right]_{\overline{x},\overline{y},\sigma} = -\left[\frac{\partial\sigma}{\partial t}\right]_{x,y,z} \tag{8.20a}$$

$$\left[\frac{\partial\sigma}{\partial z}\right]_{t,x,y}\left[\frac{\partial z}{\partial\overline{x}}\right]_{\overline{t},\overline{y},\sigma} = -\left[\frac{\partial\sigma}{\partial x}\right]_{t,y,z} \tag{8.20b}$$

$$\left[\frac{\partial\sigma}{\partial z}\right]_{t,x,y}\left[\frac{\partial z}{\partial\overline{y}}\right]_{\overline{t},\overline{x},\sigma} = -\left[\frac{\partial\sigma}{\partial y}\right]_{t,x,z}. \tag{8.20c}$$

If the vertical stratification, $\partial\sigma/\partial z$, is non-zero, the triple product identities are equivalent to

$$\left[\frac{\partial z}{\partial\overline{t}}\right]_{\overline{x},\overline{y},\sigma} = -\frac{[\partial\sigma/\partial t]_{x,y,z}}{[\partial\sigma/\partial z]_{t,x,y}} = -\left[\frac{\partial\sigma}{\partial t}\right]_{x,y,z}\left[\frac{\partial z}{\partial\sigma}\right]_{t,x,y} \tag{8.21a}$$

$$\left[\frac{\partial z}{\partial\overline{x}}\right]_{\overline{t},\overline{y},\sigma} = -\frac{[\partial\sigma/\partial x]_{t,y,z}}{[\partial\sigma/\partial z]_{t,x,y}} = -\left[\frac{\partial\sigma}{\partial x}\right]_{t,y,z}\left[\frac{\partial z}{\partial\sigma}\right]_{t,x,y} \tag{8.21b}$$

$$\left[\frac{\partial z}{\partial\overline{y}}\right]_{\overline{t},\overline{x},\sigma} = -\frac{[\partial\sigma/\partial y]_{t,x,z}}{[\partial\sigma/\partial z]_{t,x,y}} = -\left[\frac{\partial\sigma}{\partial y}\right]_{t,x,z}\left[\frac{\partial z}{\partial\sigma}\right]_{t,x,y}. \tag{8.21c}$$

Since $t = \overline{t}$, $x = \overline{x}$, and $y = \overline{y}$ we can write these identities in the more succinct form

$$\left[\frac{\partial z}{\partial\overline{t}}\right]_{\sigma} = -\frac{[\partial\sigma/\partial t]_z}{[\partial\sigma/\partial z]} \tag{8.22a}$$

$$\left[\frac{\partial z}{\partial\overline{x}}\right]_{\sigma} = -\frac{[\partial\sigma/\partial x]_z}{[\partial\sigma/\partial z]} \tag{8.22b}$$

$$\left[\frac{\partial z}{\partial\overline{y}}\right]_{\sigma} = -\frac{[\partial\sigma/\partial y]_z}{[\partial\sigma/\partial z]}. \tag{8.22c}$$

These identities are quite useful for manipulating equations involving GVCs. In particular, equations (8.22b) and (8.22c) provide alternate expressions for the slope of σ isosurfaces relative to the horizontal plane (see Section 8.11).

8.5 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 (8.9a)-(8.9c) so that the position of an arbitrary point in space is given by

$$\mathcal{P} = \xi^{\overline{a}}\,\vec{e}_{\overline{a}} \tag{8.23a}$$

$$= \overline{x}\,[\hat{\boldsymbol{x}} + (\partial z/\partial\overline{x})\,\hat{\boldsymbol{z}}] + \overline{y}\,[\hat{\boldsymbol{y}} + (\partial z/\partial\overline{y})\,\hat{\boldsymbol{z}}] + \sigma\,(\partial z/\partial\sigma)\,\hat{\boldsymbol{z}} \tag{8.23b}$$

$$= \hat{\boldsymbol{x}}\,\overline{x} + \hat{\boldsymbol{y}}\,\overline{y} + \hat{\boldsymbol{z}}\,[\overline{x}\,(\partial z/\partial\overline{x}) + \overline{y}\,(\partial z/\partial\overline{y}) + \sigma\,(\partial z/\partial\sigma)] \tag{8.23c}$$

$$= \hat{\boldsymbol{x}}\,\overline{x} + \hat{\boldsymbol{y}}\,\overline{y} + \hat{\boldsymbol{z}}\,\xi^{\overline{a}}\,\partial_{\overline{a}}z. \tag{8.23d}$$

We identify the following properties as a means to help understand these expressions, with Figure 8.3 offering a schematic.

[1] These identities are directly analogous to the Maxwell relations from thermodynamics, with an introduction in Section 22.6 and full details in the book by *Callen* (1985).

Figure 8.3: *The position of a point in space as represented using GVCs following equation (8.24a). For this example, $\overline{y} = 0$ so that the horizontal position is determined by the coordinate $\overline{x} = x$, whereas the vertical postion is determined by $\overline{x}\,(\partial z/\partial \overline{x}) + \sigma\,(\partial z/\partial \sigma)$.*

- The form (8.23b) has horizontal positions $\overline{x}$ and $\overline{y}$ multiplying the basis vectors $\vec{e}_{\overline{x}}$ and $\vec{e}_{\overline{y}}$, with these vectors oriented parallel to a surface of constant GVC as in Figure 8.3. Likewise, the third term, $\sigma\,(\partial z/\partial \sigma)\,\hat{\boldsymbol{z}}$, positions the point vertically according to the value of the GVC and its inverse stratification.

- Consider the case of $\overline{y} = 0$ so that

$$\mathscr{P} = \overline{x}\,\hat{\boldsymbol{x}} + \hat{\boldsymbol{z}}\,[\overline{x}\,(\partial z/\partial \overline{x}) + \sigma\,(\partial z/\partial \sigma)] \tag{8.24a}$$
$$= \overline{x}\,\hat{\boldsymbol{x}} + \hat{\boldsymbol{z}}\,(\partial z/\partial \sigma)\,[\overline{x}\,(\partial \sigma/\partial z)_x (\partial z/\partial \overline{x})_\sigma + \sigma] \tag{8.24b}$$
$$= \overline{x}\,\hat{\boldsymbol{x}} + \hat{\boldsymbol{z}}\,(\partial z/\partial \sigma)\,[-\overline{x}\,(\partial \sigma/\partial x)_z + \sigma], \tag{8.24c}$$

  where we used the triple product identity (8.22b) for the final equality. Hence, 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

$$\mathscr{P} \cdot \tilde{e}^{\overline{b}} = \xi^{\overline{a}}\,\vec{e}_{\overline{a}} \cdot \tilde{e}^{\overline{b}} = \xi^{\overline{b}}. \tag{8.25}$$

  This result follows from the orthogonality relation (6.25). So the projection of the position vector onto a basis one-form picks out the corresponding coordinate value.

- Equation (8.4) provides the spatial dependence for the vertical position of the surface of constant GVC

$$z = z(\xi^{\overline{a}}). \tag{8.26}$$

  At any particular time instance we can perform a Taylor series about a reference depth z_0, so that

$$z(\xi^{\overline{a}}) \approx z_0 + \xi^{\overline{a}}\,\partial_{\overline{a}} z. \tag{8.27}$$

  We can thus write the position (8.23d) in the form

$$\mathscr{P} = \hat{\boldsymbol{x}}\,\overline{x} + \hat{\boldsymbol{y}}\,\overline{y} + \hat{\boldsymbol{z}}\,[z - z_0]. \tag{8.28}$$

Taking the reference depth to be $z_0 = 0$ 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 coordinate representation.

8.6 Transforming components of a first order tensor

Consider a vector field $\vec{F}$ with Cartesian representation

$$\vec{F} = \boldsymbol{F} = F^a \, \vec{e}_a = F^x \, \hat{\boldsymbol{x}} + F^y \, \hat{\boldsymbol{y}} + F^z \, \hat{\boldsymbol{z}}. \tag{8.29}$$

The corresponding GVC components are related through the transformation matrix

$$F^{\overline{a}} = \Lambda^{\overline{a}}_{\ a} \, F^a. \tag{8.30}$$

Making use of the transformation matrix (8.11) yields the relations between GVC components and Cartesian components

$$F^{\overline{1}} = F^1 \qquad F^{\overline{2}} = F^2 \qquad F^{\overline{3}} = \nabla\sigma \cdot \boldsymbol{F}, \tag{8.31}$$

where we wrote

$$\nabla\sigma \cdot \boldsymbol{F} = (\partial\sigma/\partial x) \, F^1 + (\partial\sigma/\partial y) \, F^2 + (\partial\sigma/\partial z) \, F^3. \tag{8.32}$$

The vector field thus can be represented in GVC coordinates as

$$\vec{F} = F^{\overline{a}} \, \vec{e}_{\overline{a}} = F^1 \, \vec{e}_{\overline{1}} + F^2 \, \vec{e}_{\overline{2}} + (\nabla\sigma \cdot \boldsymbol{F}) \, \vec{e}_{\overline{3}}. \tag{8.33}$$

Similarly, the covariant components transform as $F_{\overline{a}} = \Lambda^{a}_{\ \overline{a}} \, F_a$, where use of the inverse transformation matrix (8.11) renders

$$F_{\overline{1}} = F_1 + \frac{\partial z}{\partial \overline{x}} \, F_3 \qquad F_{\overline{2}} = F_2 + \frac{\partial z}{\partial \overline{y}} \, F_3 \qquad F_{\overline{3}} = \frac{\partial z}{\partial \sigma} \, F_3, \tag{8.34}$$

and the expression for the vector field

$$\vec{F} = F_{\overline{a}} \, \widetilde{e}^{\overline{a}} = [F_1 + (\partial z/\partial \overline{x}) \, F_3] \, \widetilde{e}^{\overline{1}} + [F_2 + (\partial z/\partial \overline{y}) \, F_3] \, \widetilde{e}^{\overline{2}} + (\partial z/\partial \sigma) \, F_3 \, \widetilde{e}^{\overline{3}}. \tag{8.35}$$

Recall also that for Cartesian coordinates the contravarient and covariant components to a vector are identical: $F^a = F_a$.

8.7 Velocity

We now make use of the results from Section 8.6 to represent the velocity vector, considering both covariant and contravariant representations. As for the position vector detailed in Section 8.5, 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 8.7.4 we verify that the transformation formalism indeed respects this equivalance, with the GVC representations equivalent to the Cartesian representation

$$\vec{v} = u \, \hat{\boldsymbol{x}} + v \, \hat{\boldsymbol{y}} + w \, \hat{\boldsymbol{z}}. \tag{8.36}$$

8.7.1 Contravariant components to the velocity vector

Following Section 8.6, we have the contravariant velocity components

$$v^{\overline{1}} = u \qquad \text{and} \qquad v^{\overline{2}} = v \qquad \text{and} \qquad v^{\overline{3}} = \boldsymbol{v} \cdot \nabla\sigma. \tag{8.37}$$

Use of the basis vectors (8.9a)-(8.9c) then leads to

$$\vec{v} = v^{\overline{a}}\, \vec{e}_{\overline{a}} \tag{8.38a}$$

$$= u\, \vec{e}_{\overline{x}} + v\, \vec{e}_{\overline{y}} + (\boldsymbol{v} \cdot \nabla\sigma)\, \vec{e}_{\sigma} \tag{8.38b}$$

$$= u\, [\hat{\boldsymbol{x}} + (\partial z/\partial\overline{x})\, \hat{\boldsymbol{z}}] + v\, [\hat{\boldsymbol{y}} + (\partial z/\partial\overline{y})\, \hat{\boldsymbol{z}}] + (\boldsymbol{v} \cdot \nabla\sigma)\, (\partial z/\partial\sigma)\, \hat{\boldsymbol{z}}. \tag{8.38c}$$

8.7.2 Covariant components to the velocity vector

The covariant velocity components are given by

$$v_{\overline{1}} = u + \frac{\partial z}{\partial\overline{x}}\, w \qquad \text{and} \qquad v_{\overline{2}} = v + \frac{\partial z}{\partial\overline{y}}\, w \qquad \text{and} \qquad v_{\overline{3}} = \frac{\partial z}{\partial\sigma}\, w. \tag{8.39}$$

The one-form basis (8.13a)–(8.13c) thus leads to the velocity vector

$$\vec{v} = v_{\overline{a}}\, \tilde{e}^{\overline{a}} = [u + (\partial z/\partial\overline{x})\, w]\, \hat{\boldsymbol{x}} + [v + (\partial z/\partial\overline{y})\, w]\, \hat{\boldsymbol{y}} + w\, (\partial z/\partial\sigma)\, \nabla\sigma. \tag{8.40}$$

8.7.3 Introducing the material time derivative

The material evolution for the generalized vertical coordinate can be written

$$\frac{\mathrm{D}\sigma}{\mathrm{D}t} = \frac{\partial\sigma}{\partial t} + \boldsymbol{v} \cdot \nabla\sigma = \dot{\sigma}, \tag{8.41}$$

with $\dot{\sigma}$ symbolizing any process contributing to motion across σ isosurfaces (as fully explained in Section 17.3). Using the expression (8.41) in the velocity vector expression (8.38c) leads to

$$\vec{v} = u\, [\hat{\boldsymbol{x}} + (\partial z/\partial\overline{x})\, \hat{\boldsymbol{z}}] + v\, [\hat{\boldsymbol{y}} + (\partial z/\partial\overline{y})\, \hat{\boldsymbol{z}}] + (\boldsymbol{v} \cdot \nabla\sigma)\, (\partial z/\partial\sigma)\, \hat{\boldsymbol{z}}. \tag{8.42a}$$

$$= u\, [\hat{\boldsymbol{x}} + (\partial z/\partial\overline{x})\, \hat{\boldsymbol{z}}] + v\, [\hat{\boldsymbol{y}} + (\partial z/\partial\overline{y})\, \hat{\boldsymbol{z}}] + (\dot{\sigma} - \partial\sigma/\partial t)\, (\partial z/\partial\sigma)\, \hat{\boldsymbol{z}} \tag{8.42b}$$

$$= u\, \hat{\boldsymbol{x}} + v\, \hat{\boldsymbol{y}} + [\partial z/\partial\overline{t} + \boldsymbol{u} \cdot \nabla_{\sigma} z + (\partial z/\partial\sigma)\dot{\sigma}]\, \hat{\boldsymbol{z}}, \tag{8.42c}$$

where the final equality made use of the triple product (8.21a): $(\partial\sigma/\partial t)\, (\partial z/\partial\sigma) = -\partial z/\partial\overline{t}$. 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 $\boldsymbol{v} \cdot \nabla\sigma = 0$ and

$$\vec{v} = u\, [\hat{\boldsymbol{x}} + (\partial z/\partial\overline{x})\, \hat{\boldsymbol{z}}] + v\, [\hat{\boldsymbol{y}} + (\partial z/\partial\overline{y})\, \hat{\boldsymbol{z}}] \qquad \text{if } \partial_t\sigma = 0 \text{ and } \dot{\sigma} = 0. \tag{8.43}$$

However, in general there are transient fluctuations and material changes to σ so that $\boldsymbol{v} \cdot \nabla\sigma \neq 0$.

8.7.4 Equivalence to the Cartesian velocity representation

Use of the triple product identities (8.22b)-(8.22c) allows us to manipulate both expressions (8.38c) and (8.40) to recover the Cartesian expression

$$\vec{v} = u\, \hat{\boldsymbol{x}} + v\, \hat{\boldsymbol{y}} + w\, \hat{\boldsymbol{z}}. \tag{8.44}$$

Another way to see this identity is to note that in equation (8.42c), the vertical component is an expression for the material time derivative of the vertical position

$$w = \frac{\mathrm{D}z}{\mathrm{D}t} = \frac{\partial z}{\partial\overline{t}} + \boldsymbol{u} \cdot \nabla_{\sigma} z + \frac{\partial z}{\partial\sigma}\, \dot{\sigma}. \tag{8.45}$$

We derive this identity in Section 17.4 where we discuss further kinematic results using GVCs.

8.8 Metric tensor

Recall from Section 6.1 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

$$g_{\overline{a}\overline{b}} = \vec{e}_{\overline{a}} \cdot \vec{e}_{\overline{b}} = \begin{bmatrix} 1 + (\partial z/\partial \overline{x})^2 & (\partial z/\partial \overline{x})\,(\partial z/\partial \overline{y}) & (\partial z/\partial \overline{x})\,(\partial z/\partial \sigma) \\ (\partial z/\partial \overline{x})\,(\partial z/\partial \overline{y}) & 1 + (\partial z/\partial \overline{y})^2 & (\partial z/\partial \overline{y})\,(\partial z/\partial \sigma) \\ (\partial z/\partial \overline{x})\,(\partial z/\partial \sigma) & (\partial z/\partial \overline{y})\,(\partial z/\partial \sigma) & (\partial z/\partial \sigma)^2 \end{bmatrix}, \tag{8.46}$$

and the GVC representation of the inverse metric tensor is given by the somewhat simpler form

$$g^{\overline{a}\overline{b}} = \widetilde{e}^{\overline{a}} \cdot \widetilde{e}^{\overline{b}} = \begin{bmatrix} 1 & 0 & \partial \sigma/\partial x \\ 0 & 1 & \partial \sigma/\partial y \\ \partial \sigma/\partial x & \partial \sigma/\partial y & |\nabla \sigma|^2 \end{bmatrix}. \tag{8.47}$$

Proof that $g^{\overline{a}\overline{b}}\, g_{\overline{b}\overline{c}} = \delta^{\overline{a}}_{\overline{c}}$ requires use of the triple product identities (8.22b) and (8.22c).

8.8.1 Jacobian of transformation

The determinant of the GVC representation of the metric tensor (8.46) is

$$\det(g_{\overline{a}\overline{b}}) = (\partial z/\partial \sigma)^2 \tag{8.48}$$

so that the Jacobian of transformation (Section 6.4) is the specific thickness

$$\frac{\partial(x, y, z)}{\partial(\overline{x}, \overline{y}, \sigma)} = \frac{\partial z}{\partial \sigma}. \tag{8.49}$$

The coordinate transformation from Cartesian to generalized vertical is invertible only so long as the Jabobian 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 by[2]

$$\frac{\partial z}{\partial \sigma} = \frac{\partial z}{\partial p} = -\frac{1}{\rho g}, \tag{8.50}$$

which is indeed single-signed since the mass density is always positive.

8.8.2 Covariant and contravariant representations

The metric tensor allows us to convert between the covariant and contravariant representations of a vector via the identity (Section 6.2.3)

$$F_{\overline{a}} = g_{\overline{a}\overline{b}}\, F^{\overline{b}}. \tag{8.51}$$

We use triple product identities (8.22b)-(8.22c) to verify that this relation agrees with the transformation matrix approach detailed in Section 8.6. For example,

$$\begin{aligned} F_{\overline{1}} &= g_{\overline{1}\overline{b}}\, F^{\overline{b}} && (8.52a) \\ &= [1 + (\partial z/\partial \overline{x})^2]\, F^{\overline{1}} + (\partial z/\partial \overline{x})\,(\partial z/\partial \overline{y})\, F^{\overline{2}} + (\partial z/\partial \overline{x})\,(\partial z/\partial \sigma)\, F^{\sigma} && (8.52b) \\ &= [1 + (\partial z/\partial \overline{x})^2]\, F^{1} + (\partial z/\partial \overline{x})\,(\partial z/\partial \overline{y})\, F^{2} + (\partial z/\partial \overline{x})\,(\partial z/\partial \sigma)\, \nabla \sigma \cdot \boldsymbol{F} && (8.52c) \\ &= F^1 + (\partial z/\partial \overline{x})\, F^3 && (8.52d) \\ &= F_1 + (\partial z/\partial \overline{x})\, F_3, && (8.52e) \end{aligned}$$

[2] We derive the hydrostatic balance in Section 24.3.

where the final equality holds since $F^1 = F_1$ and $F^3 = F_3$ for Cartesian tensor components.

8.9 Volume element and the Levi-Civita tensor

The square root of the determinant of the metric tensor (8.46) is

$$\sqrt{\det(g_{\overline{a}\overline{b}})} = \partial z/\partial\sigma \tag{8.53}$$

so that the volume element (Section 6.4) is

$$\mathrm{d}V = (\partial z/\partial\sigma)\,\mathrm{d}x\,\mathrm{d}y\,\mathrm{d}\sigma. \tag{8.54}$$

The covariant Levi-Civita tensor (Section 6.6) has the GVC representations

$$\varepsilon_{\overline{a}\overline{b}\overline{c}} = (\partial z/\partial\sigma)\,\epsilon_{\overline{a}\overline{b}\overline{c}} \qquad \varepsilon^{\overline{a}\overline{b}\overline{c}} = (\partial z/\partial\sigma)^{-1}\,\epsilon^{\overline{a}\overline{b}\overline{c}} \tag{8.55}$$

where ϵ is the permutation symbol introduced in Section 1.4.1 with its components independent of coordinate representation.

8.10 Vector cross product of basis vectors

We now verify the relation (6.67) for the cross product of two basis vectors using GVCs

$$\vec{e}_{\overline{a}} \wedge \vec{e}_{\overline{b}} = \varepsilon_{\overline{a}\overline{b}\overline{c}}\,\widetilde{e}^{\overline{c}} \Longrightarrow \vec{e}_{\overline{a}} \wedge \vec{e}_{\overline{b}} = (\partial z/\partial\sigma)\epsilon_{\overline{a}\overline{b}\overline{c}}\,\widetilde{e}^{\overline{c}}. \tag{8.56}$$

Making use of the basis vectors from Section 8.2 and the basis one-forms from Section 8.3 renders

$$\vec{e}_{\overline{x}} \wedge \vec{e}_{\overline{y}} = \hat{\boldsymbol{z}} - \hat{\boldsymbol{x}}\,(\partial z/\partial\overline{x}) - \hat{\boldsymbol{y}}\,(\partial z/\partial\overline{y}) = (\partial z/\partial\sigma)\,\nabla\sigma = \varepsilon_{\overline{x}\overline{y}\sigma}\,\widetilde{e}^{\sigma} \tag{8.57a}$$

$$\vec{e}_{\overline{y}} \wedge \vec{e}_{\sigma} = \hat{\boldsymbol{x}}\,(\partial z/\partial\sigma) = \varepsilon_{\overline{y}\sigma\overline{x}}\,\widetilde{e}^{\overline{x}} \tag{8.57b}$$

$$\vec{e}_{\sigma} \wedge \vec{e}_{\overline{x}} = \hat{\boldsymbol{y}}\,(\partial z/\partial\sigma) = \varepsilon_{\sigma\overline{x}\overline{y}}\,\widetilde{e}^{\overline{y}}. \tag{8.57c}$$

8.11 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 8.11.2 requires minimal use of the previous tensor formalism.

8.11.1 Analytical derivation

The partial derivative operators in GVCs are computed via $\partial_{\overline{a}} = \Lambda^{a}_{\overline{a}}\,\partial_a$. Including also the time component leads to the relations

$$\partial_{\overline{t}} = \partial_t + (\partial z/\partial\overline{t})\,\partial_z \qquad \partial_{\overline{x}} = \partial_x + (\partial z/\partial\overline{x})\,\partial_z \qquad \partial_{\overline{y}} = \partial_y + (\partial z/\partial\overline{y})\,\partial_z \qquad \partial_{\sigma} = (\partial z/\partial\sigma)\,\partial_z. \tag{8.58}$$

We can make use of the triple product identities (8.22b) and (8.22c) to express the slope of a constant GVC surface in two equivalent manners

$$\nabla_{\sigma} z = -\frac{\nabla_z \sigma}{\partial\sigma/\partial z} \tag{8.59}$$

where we introduced the standard shorthand notation

$$\nabla_\sigma = \hat{\boldsymbol{x}}\,\frac{\partial}{\partial \overline{x}} + \hat{\boldsymbol{y}}\,\frac{\partial}{\partial \overline{y}} \qquad \text{and} \qquad \nabla_z = \hat{\boldsymbol{x}}\,\frac{\partial}{\partial x} + \hat{\boldsymbol{y}}\,\frac{\partial}{\partial y}. \tag{8.60}$$

It is common to transform between the horizontal gradient operators, in which case we write

$$\nabla_\sigma = \nabla_z + (\nabla_\sigma z)\,\partial_z \equiv \nabla_z + \boldsymbol{S}\,\partial_z, \tag{8.61}$$

where we introduced the notation for the slope of the σ surface relative to the horizontal plane

$$\boldsymbol{S} = \nabla_\sigma z = -\frac{\nabla_z \sigma}{\partial\sigma/\partial z}. \tag{8.62}$$

It is important to note that ∇_σ is merely a shorthand for the two partial derivative operators, and that it furthermore only has components in the horizontal directions.

8.11.2 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 8.4. This feature of the horizontal derivative operator is a key aspect of the GVCs' non-orthogonality.

$$\left[\frac{\partial \psi}{\partial x}\right]_\sigma \approx \frac{\psi(B) - \psi(A)}{x(C) - x(A)}$$

Figure 8.4: *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 (8.64a)-(8.64d), 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 8.4, which shows a generalized vertical coordinate surface (constant σ surface) along with a sample tangent plane with a slope

$$S^{(x)} = \frac{\text{rise}}{\text{run}} = \tan\vartheta = \frac{z(B) - z(C)}{x(C) - x(A)} \approx \left[\frac{\partial z}{\partial x}\right]_\sigma = -\frac{(\partial\sigma/\partial x)_z}{(\partial\sigma/\partial z)} \tag{8.63}$$

relative to the horizontal. We readily verify the following identities based on finite difference operations for an arbitrary function

$$\left[\frac{\partial \psi}{\partial x}\right]_\sigma \approx \frac{\psi(B) - \psi(A)}{x(C) - x(A)} \tag{8.64a}$$

$$= \frac{\psi(C) - \psi(A)}{x(C) - x(A)} + \frac{\psi(B) - \psi(C)}{x(C) - x(A)} \tag{8.64b}$$

$$= \frac{\psi(C) - \psi(A)}{x(C) - x(A)} + \left[\frac{z(B) - z(C)}{x(C) - x(A)}\right]\frac{\psi(B) - \psi(C)}{z(B) - z(C)} \tag{8.64c}$$

$$= \left[\frac{\partial \psi}{\partial x}\right]_z + S^{(x)}\left[\frac{\partial \psi}{\partial z}\right]_x. \tag{8.64d}$$

Taking the continuum limit then leads to the relations between horizontal derivatives computed on constant σ surfaces to those computed on constant z surfaces

$$\left[\frac{\partial}{\partial x}\right]_\sigma = \left[\frac{\partial}{\partial x}\right]_z + \left[\frac{\partial z}{\partial x}\right]_\sigma \frac{\partial}{\partial z} \tag{8.65a}$$

$$\left[\frac{\partial}{\partial y}\right]_\sigma = \left[\frac{\partial}{\partial y}\right]_z + \left[\frac{\partial z}{\partial y}\right]_\sigma \frac{\partial}{\partial z}, \tag{8.65b}$$

which can be written in the shorthand vector notation

$$\nabla_\sigma = \hat{\boldsymbol{x}} \left[\frac{\partial}{\partial x}\right]_\sigma + \hat{\boldsymbol{y}} \left[\frac{\partial}{\partial y}\right]_\sigma = \nabla_z + (\nabla_\sigma z)\, \partial_z. \tag{8.66}$$

8.12 Material time derivative

Making use of the relations for the partial derivative operators in Section 8.11 allows us to write the material time derivative in the following equivalent forms

$$\frac{\mathrm{D}}{\mathrm{D}t} = \left[\frac{\partial}{\partial t}\right]_z + \boldsymbol{u}\cdot\nabla_z + w\,\frac{\partial}{\partial z} \tag{8.67a}$$

$$= \left[\frac{\partial}{\partial t}\right]_\sigma - (\partial z/\partial \bar{t})\,\partial_z + \boldsymbol{u}\cdot[\nabla_\sigma - (\nabla_\sigma z)\,\partial_z] + w\,\partial/\partial z \tag{8.67b}$$

$$= \left[\frac{\partial}{\partial t}\right]_\sigma + \boldsymbol{u}\cdot\nabla_\sigma + \left[w - \boldsymbol{u}\cdot\nabla_\sigma z - \partial z/\partial \bar{t}\right](\partial\sigma/\partial z)\,\partial/\partial\sigma \tag{8.67c}$$

$$= \left[\frac{\partial}{\partial t}\right]_\sigma + \boldsymbol{u}\cdot\nabla_\sigma + \frac{\mathrm{D}\sigma}{\mathrm{D}t}\,\frac{\partial}{\partial\sigma} \tag{8.67d}$$

$$= \left[\frac{\partial}{\partial t}\right]_\sigma + \boldsymbol{u}\cdot\nabla_\sigma + \frac{\partial z}{\partial\sigma}\,\frac{\mathrm{D}\sigma}{\mathrm{D}t}\,\frac{\partial}{\partial z}. \tag{8.67e}$$

The equality (8.67d) made use of the identity (8.45), which is itself derived in Section 17.4 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

$$(u, v) = \frac{\mathrm{D}(x, y)}{\mathrm{D}t}. \tag{8.68}$$

8.13 Divergence of a vector and the divergence theorem

Making use of the general expression (6.14) for the covariant divergence of a vector renders the GVC expression

$$\nabla_{\bar{a}}\, F^{\bar{a}} = [\det(g_{\bar{a}\bar{b}})]^{-1/2}\,\partial_{\bar{a}}\left[[\det(g_{\bar{a}\bar{b}})]^{1/2}\; F^{\bar{a}}\right] = (\partial z/\partial\sigma)^{-1}\partial_{\bar{a}}\,[(\partial z/\partial\sigma)\,F^{\bar{a}}]. \tag{8.69}$$

Recall that the GVC vector components, $F^{\bar{a}}$, are related to the Cartesian components in equation (8.31), and the GVC components of the partial derivative operator, $\partial_{\bar{a}}$, are related to the Cartesian operator in equation (8.60).

When making use of the divergence theorem (Section 6.17), we require the product of the volume element and the covariant divergence. For GVCs this product takes on the form

$$(\nabla_{\overline{a}}\, F^{\overline{a}})\, \mathrm{d}V = \partial_{\overline{a}}\, [(\partial z/\partial\sigma)\, F^{\overline{a}}]\, \mathrm{d}\overline{x}\, \mathrm{d}\overline{y}\, \mathrm{d}\sigma, \tag{8.70}$$

which reduces to a boundary integral when integrating over a volume.

8.14 The diffusion operator

As an explicit example of the covariant divergence operator (8.69), we here consider the diffusion operator discussed in Chapter 47. The derivation here recovers much of what we just discussed in Section 8.13, yet we make use of a bit less tensor formalism though at the cost of more algebra.

8.14.1 Continuous expression

The diffusion operator is the convergence of the diffusive flux

$$\mathcal{R} = -\nabla \cdot \boldsymbol{J}, \tag{8.71}$$

where $\boldsymbol{J}$ 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 8.11

$$\begin{aligned}
-\mathcal{R} &= \nabla \cdot \boldsymbol{J} && \text{(8.72a)}\\
&= \nabla_z \cdot \boldsymbol{J}^h + \partial_z J^z && \text{(8.72b)}\\
&= (\nabla_\sigma - \nabla_\sigma z\, \partial_z) \cdot \boldsymbol{J}^h + (\sigma_z)\, \partial_\sigma J^z && \text{(8.72c)}\\
&= \sigma_z \left[z_\sigma\, \nabla_\sigma \cdot \boldsymbol{J}^h + (\hat{\boldsymbol{z}}\, \partial_\sigma - \nabla_\sigma z\, \partial_\sigma) \cdot \boldsymbol{J} \right] && \text{(8.72d)}\\
&= \sigma_z \left[\nabla_\sigma \cdot (z_\sigma\, \boldsymbol{J}^h) - \boldsymbol{J}^h \cdot \nabla_\sigma\, (z_\sigma) + \partial_\sigma J^z - \partial_\sigma\, (\nabla_\sigma z \cdot \boldsymbol{J}) + \boldsymbol{J} \cdot \partial_\sigma\, (\nabla_\sigma z) \right] && \text{(8.72e)}\\
&= \sigma_z \left[\nabla_\sigma \cdot (z_\sigma\, \boldsymbol{J}^h) + \partial_\sigma J^z - \partial_\sigma\, (\nabla_\sigma z \cdot \boldsymbol{J}^h) \right] && \text{(8.72f)}\\
&= \sigma_z \left(\nabla_\sigma \cdot (\partial_\sigma z\, \boldsymbol{J}^h) + \partial_\sigma\, [(\hat{\boldsymbol{z}} - \nabla_\sigma z) \cdot \boldsymbol{J}] \right) && \text{(8.72g)}\\
&= \sigma_z \left[\nabla_\sigma \cdot (z_\sigma\, \boldsymbol{J}^h) + \partial_\sigma\, (z_\sigma\, \nabla\sigma \cdot \boldsymbol{J}) \right], && \text{(8.72h)}
\end{aligned}$$

where we used

$$z_\sigma \nabla\sigma = \hat{\boldsymbol{z}} - \nabla_\sigma z \tag{8.73}$$

to reach the final equality, and made use of the shorthand

$$z_\sigma = \frac{\partial z}{\partial \sigma} \qquad \text{and} \qquad \sigma_z = \frac{\partial \sigma}{\partial z}. \tag{8.74}$$

The coordinate transformations in Section 8.6 for vector components reveal that the expression (8.72h) is identical to equation (8.69) derived using formal tensor methods. Likewise, when multiplying by the volume element

$$\mathrm{d}V = \mathrm{d}x\, \mathrm{d}y\, \mathrm{d}z = \mathrm{d}x\, \mathrm{d}y\, z_\sigma\, \mathrm{d}\sigma, \tag{8.75}$$

leads to

$$-\,\mathcal{R}\, \mathrm{d}V = \left[\nabla_\sigma \cdot (z_\sigma\, \boldsymbol{J}^h) + \partial_\sigma\, (z_\sigma\, \nabla\sigma \cdot \boldsymbol{J}) \right] \mathrm{d}x\, \mathrm{d}y\, \mathrm{d}\sigma, \tag{8.76}$$

which is identical to the expression (8.70).

8.14.2 Layer thickness weighted diffusion operator

Consider a presecribed 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

$$-\mathcal{R}\,\delta V = \left[\nabla_\sigma \cdot (\delta\sigma\, z_\sigma\, \boldsymbol{J}^h) + \delta\sigma\, \partial_\sigma\, (z_\sigma\, \nabla\sigma \cdot \boldsymbol{J})\right] \delta x\, \delta y \tag{8.77a}$$

$$= \frac{1}{\delta z}\left[\nabla_\sigma \cdot (\delta\sigma\, z_\sigma\, \boldsymbol{J}^h) + \delta\sigma\, \partial_\sigma\, (z_\sigma\, \nabla\sigma \cdot \boldsymbol{J})\right] \delta x\, \delta y\, \delta z \tag{8.77b}$$

$$= \frac{1}{h}\left[\nabla_\sigma \cdot (h\, \boldsymbol{J}^h) + \Delta_\sigma(z_\sigma\, \nabla\sigma \cdot \boldsymbol{J})\right] \delta x\, \delta y\, h, \tag{8.77c}$$

where we introduced the infinitesimal layer thickness

$$h = z_\sigma\, \delta\sigma \tag{8.78}$$

and the non-dimensional differential operator

$$\Delta_\sigma \equiv \delta\sigma\, \frac{\partial}{\partial\sigma}. \tag{8.79}$$

Cancelling the volume element on both sides leads to the diffusion operator

$$\mathcal{R} = -\frac{1}{h}\left[\nabla_\sigma \cdot (h\, \boldsymbol{J}^h) + \Delta_\sigma(z_\sigma\, \nabla\sigma \cdot \boldsymbol{J})\right]. \tag{8.80}$$

This form is commonly found in the ocean modeling literature when considering layered models of ocean circulation.

We make the following comments concerning the diffusion operator in equation (8.80).

- Our introduction of the layer thickness $h = z_\sigma\, \delta\sigma$ is treated a more formally in Sections 17.9 and 17.10 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, $h\, \boldsymbol{J}^h$, is oriented within the horizontal plane. However, its contribution to the diffusion operator is computed by taking its convergence 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 $z_\sigma \nabla\sigma \cdot \boldsymbol{J}$ 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

$$\mathcal{R} = -\frac{1}{h}\left[\nabla_\sigma \cdot (h\, \boldsymbol{J}^h)\right] \qquad \text{if } \nabla\sigma \cdot \boldsymbol{J} = 0. \tag{8.81}$$

  The neutral diffusion operator of Section 49.3.3 is an example of such an operator, with σ in that case given by the locally referenced potential density.

8.15 Vorticity

As detailed in Chapter 34, vorticity is the curl of the velocity

$$\vec{\omega} = \text{curl}(\vec{v}), \tag{8.82}$$

where the curl has components (Section 6.16)

$$\text{curl}(\vec{v}) = \widetilde{e}_a\, \varepsilon^{abc} \partial_b v_c = \widetilde{e}_{\overline{a}}\, \varepsilon^{\overline{a}\overline{b}\overline{c}} \partial_{\overline{b}} v_{\overline{c}}. \tag{8.83}$$

8.15.1 The components

We identify the contravariant components of the vorticity via

$$\omega^{\overline{a}} = \varepsilon^{\overline{a}\overline{b}\overline{c}} \partial_{\overline{b}} v_{\overline{c}} = (\partial z/\partial \sigma)^{-1} \, \epsilon^{\overline{a}\overline{b}\overline{c}} \, \partial_{\overline{b}} v_{\overline{c}} \tag{8.84}$$

where we made use of equation (8.55) to introduce the permutation symbol. Expanding the components leads to

$$\omega^{\overline{1}} = (\partial \sigma/\partial z) \, (\partial_{\overline{2}} \, v_{\overline{3}} - \partial_{\overline{3}} \, v_{\overline{2}}) \tag{8.85a}$$

$$\omega^{\overline{2}} = (\partial \sigma/\partial z) \, (\partial_{\overline{3}} \, v_{\overline{1}} - \partial_{\overline{1}} \, v_{\overline{3}}) \tag{8.85b}$$

$$\omega^{\overline{3}} = \omega^{\sigma} = (\partial \sigma/\partial z) \, (\partial_{\overline{1}} \, v_{\overline{2}} - \partial_{\overline{2}} \, v_{\overline{1}}). \tag{8.85c}$$

8.15.2 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

$$\omega^{\overline{a}} = \Lambda^{\overline{a}}_{\ a} \, \omega^{a}, \tag{8.86}$$

where ω^a are the Cartesian components

$$\boldsymbol{\omega} = \hat{\boldsymbol{x}} \left(\frac{\partial w}{\partial y} - \frac{\partial v}{\partial z} \right) + \hat{\boldsymbol{y}} \left(\frac{\partial u}{\partial z} - \frac{\partial w}{\partial x} \right) + \hat{\boldsymbol{z}} \left(\frac{\partial v}{\partial x} - \frac{\partial u}{\partial y} \right). \tag{8.87}$$

Making use of the transformation matrix $\Lambda^{\overline{a}}_{\ a}$ from equation (8.11) yields (as in Section 8.6)

$$\omega^{\overline{x}} = \omega^{x} = \frac{\partial w}{\partial y} - \frac{\partial v}{\partial z} \qquad \text{and} \qquad \omega^{\overline{y}} = \omega^{y} = \frac{\partial u}{\partial z} - \frac{\partial w}{\partial x} \qquad \text{and} \qquad \omega^{\sigma} = \boldsymbol{\omega} \cdot \nabla \sigma. \tag{8.88}$$

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.

8.16 Velocity circulation

The velocity circulation (Section 32.3) is given by the closed oriented path integral of the velocity projected into the direction of the path

$$\mathcal{C} \equiv \oint_{\partial \mathcal{S}} \boldsymbol{v} \cdot \mathrm{d}\boldsymbol{x} \tag{8.89}$$

where $\mathrm{d}\boldsymbol{x}$ 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

$$\mathcal{C} = \oint_{\partial \mathcal{S}} \boldsymbol{v} \cdot \mathrm{d}\boldsymbol{x} = \int_{\mathcal{S}} (\nabla \wedge \boldsymbol{v}) \cdot \hat{\boldsymbol{n}} \, \mathrm{d}\mathcal{S} = \int_{\mathcal{S}} \boldsymbol{\omega} \cdot \hat{\boldsymbol{n}} \, \mathrm{d}\mathcal{S}, \tag{8.90}$$

where $\hat{\boldsymbol{n}}$ is the outward normal vector orienting the area element $\mathrm{d}\mathcal{S}$ according to the right-hand rule applied to the bounding circuit. These results are all written in a generally covariant manner (Section 5.2) 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

$$\hat{\boldsymbol{n}} = \frac{\nabla\sigma}{|\nabla\sigma|} \tag{8.91}$$

is the outward normal and

$$\boldsymbol{\omega}\cdot\hat{\boldsymbol{n}} = \frac{\omega^\sigma}{|\nabla\sigma|} \tag{8.92}$$

where $\omega^\sigma = \boldsymbol{\omega}\cdot\nabla\sigma$ (equation (8.88)). So long as the vertical stratification remains non-zero ($\partial\sigma/\partial z \neq 0$) we can write the area factor in the form

$$\frac{\mathrm{d}\mathcal{S}}{|\nabla\sigma|} = \frac{\mathrm{d}\mathcal{S}}{\sqrt{(\partial\sigma/\partial x)^2 + (\partial\sigma/\partial y)^2 + (\partial\sigma/\partial z)^2}} \tag{8.93a}$$

$$= \frac{\mathrm{d}\mathcal{S}}{|\partial\sigma/\partial z|\,\sqrt{[(\partial\sigma/\partial x)/(\partial\sigma/\partial z)]^2 + [(\partial\sigma/\partial y)/(\partial\sigma/\partial z)]^2 + 1}} \tag{8.93b}$$

$$= \frac{\mathrm{d}\mathcal{S}}{|\partial\sigma/\partial z|\sqrt{1+\tan^2\vartheta}} \tag{8.93c}$$

$$= \left|\frac{\partial z}{\partial\sigma}\right| |\cos\vartheta|\,\mathrm{d}\mathcal{S} \tag{8.93d}$$

$$= \left|\frac{\partial z}{\partial\sigma}\right| \mathrm{d}A. \tag{8.93e}$$

The equality (8.93c) introduces the angle, ϑ, between the boundary surface and the horizontal plane as in Figure 8.4. The squared slope of this surface given by

$$\tan^2\vartheta = \frac{\nabla_z\sigma\cdot\nabla_z\sigma}{(\partial\sigma/\partial z)^2} = \nabla_\sigma z\cdot\nabla_\sigma z. \tag{8.94}$$

The equality (8.93d) made use of a trigonometric identity, and the equality (8.93e) introduced the horizontal projection of the area,

$$\mathrm{d}A = |\cos\vartheta|\,\mathrm{d}\mathcal{S}. \tag{8.95}$$

Bringing these results together leads to the expression for circulation around a closed loop on a constant σ surface

$$\mathcal{C}_{\sigma-\mathrm{surface}} = \int_{\mathcal{S}} (\boldsymbol{\omega}\cdot\nabla\sigma)\,|\partial z/\partial\sigma|\,\mathrm{d}A. \tag{8.96}$$

Part II

Geophysical particle mechanics

Rotation and gravitation of the planet 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 convenient, if not essential, to make use of a rotating (non-inertial) terrestrial reference frame for our description.

In this part of the book, we study the Newtonian mechanics of a point particle moving around a rotating and gravitating sphere as viewed in the terrestrial reference frame. This analysis exposes concepts that later appear 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. We make use of spherical coordinates to develop expressions for position, velocity, and acceleration, with the mathematical treatment for particles sufficient for geophysical fluids.

Geophysical fluid mechanics involves concepts from both fluid mechanics and classical point particle mechanics. Hence, in our development of geophysical fluid mechanics, we find it useful to first examine rotational and gravitational particle mechanics; i.e., to study *geophysical particle mechanics*. Doing so also allows those less versed in classical mechanics to learn, or perhaps relearn, basic notions from particle mechanics also of use for understanding their more complex realizations in fluids. In this manner, geophysical particle mechanics provides a useful foundation for geophysical fluid mechanics.

9

Particle kinematics and dynamics

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, 5, 6, and Section 7.2. 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.

9.1 A survey of Newtonian particle mechanics **136**
9.1.1 Mechanics of a point particle 137
9.1.2 Galilean invariance of inertial reference frames 137
9.1.3 Inertial and non-inertial reference frames 138
9.1.4 Angular momentum 138
9.1.5 Mechanical work and kinetic energy 138
9.1.6 Conservative forces 139

9.1.7 Friction as a non-conservative force . . . 139
9.1.8 Comments . . . 140
9.2 The rotating earth . . . **140**
9.3 Reference frames and non-inertial accelerations . . . **140**
9.4 Rotation of a vector . . . **142**
9.4.1 Change in direction: brief derivation . . . 142
9.4.2 Change in direction: detailed derivation for planar motion . . . 142
9.4.3 Comments . . . 143
9.5 Some tensor algebra . . . **143**
9.5.1 Why we need general tensors . . . 143
9.5.2 The coordinate representation of a vector . . . 144
9.6 The velocity vector . . . **145**
9.6.1 Coordinate velocity . . . 145
9.6.2 Changes to the basis vectors . . . 145
9.7 Inertial acceleration and its decomposition . . . **145**
9.8 Representing the position vector . . . **146**
9.9 Representing the velocity vector . . . **146**
9.9.1 Planetary Cartesian coordinate representation . . . 147
9.9.2 Planetary spherical coordinate representation . . . 147
9.9.3 Axial angular momentum . . . 148
9.10 Planetary Cartesian representation of acceleration . . . **148**
9.10.1 Planetary Cartesian representation . . . 149
9.10.2 Summary of acceleration representation . . . 150
9.10.3 Further study . . . 150
9.11 Spherical representation of acceleration . . . **151**
9.11.1 Decomposing the acceleration . . . 152
9.11.2 Spherical coordinate acceleration . . . 152
9.11.3 Metric acceleration . . . 152
9.11.4 Centrifugal acceleration . . . 153
9.11.5 Coriolis acceleration . . . 153
9.11.6 Coriolis acceleration for large-scale motions . . . 153
9.12 Gravitational force and potential energy . . . **154**
9.12.1 Newton's Gravitational Law . . . 155
9.12.2 Effective gravitational force from the geopotential . . . 156
9.12.3 Further study . . . 156
9.13 Newton's law of motion . . . **156**
9.13.1 Cartesian coordinate representation . . . 156
9.13.2 Spherical coordinate representation . . . 157
9.13.3 Geopotential coordinates for slightly oblate spheriods . . . 158
9.13.4 Further study . . . 158
9.14 Exercises . . . **158**

9.1 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 rather than precise.

9.1.1 Mechanics of a point particle

Let $\boldsymbol{X}(t)$ 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 $\boldsymbol{X}$ 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

$$\boldsymbol{V} = \frac{\mathrm{d}\boldsymbol{X}}{\mathrm{d}t} = \dot{\boldsymbol{X}}, \tag{9.1}$$

where the second equality introduced a common "dot" notation for time derivatives that will sometimes be used in this book. Geometrically, the velocity is the tangent to the trajectory at each time instance.

The *linear momentum* of the particle equals to the mass of the particle, m, times its velocity

$$\boldsymbol{P} = m\,\boldsymbol{X}. \tag{9.2}$$

The momentum of the particle changes when it interacts with other objects and/or experiences forces. The vector sum of these forces is written $\boldsymbol{F}$, 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

$$\frac{\mathrm{d}\boldsymbol{P}}{\mathrm{d}t} = \boldsymbol{F}. \tag{9.3}$$

If the particle mass is fixed, then this equation becomes a differential equation for the particle position or trajectory

$$\boldsymbol{A} = \frac{\mathrm{d}\boldsymbol{V}}{\mathrm{d}t} = \frac{\mathrm{d}^2\boldsymbol{X}}{\mathrm{d}t^2} = \boldsymbol{F}/m. \tag{9.4}$$

We refer to $\boldsymbol{A} = \dot{\boldsymbol{V}}$ as the *acceleration*.

Many conclusions in mechanics are expressed in terms of conservation laws (Chapter 10), 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: $\mathrm{d}\boldsymbol{P}/\mathrm{d}t = 0$. Depending on the nature of the forces, this conservation law might hold for only one or two of the vector components to the linear momentum.

9.1.2 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 (9.4) is unchanged if shifting the velocity by a constant. This arbitrariness in the velocity represents a *symmetry* respected by the equation of motion. 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 (9.4). The reason we cannot make a distinction is that the mathematical expression for the equation of motion in the two inertial frames is indistinguishable. 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.

We find this discussion to offer a simple example of the mathematical transformation theory introduced in Part I of this book. Mathematically, the Galilean transformation is written

$$\bar{t} = t \qquad \text{and} \qquad \overline{\boldsymbol{X}} = \boldsymbol{X} + \boldsymbol{U}t, \tag{9.5}$$

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 $\boldsymbol{U}$. Some refer to the barred reference frame as a *Galilean boosted* frame. The particle velocity in the moving (boosted) reference frame is given by

$$\overline{\boldsymbol{V}} = \frac{\mathrm{d}\overline{\boldsymbol{X}}}{\mathrm{d}\bar{t}} = \frac{\mathrm{d}\boldsymbol{X}}{\mathrm{d}t} + \frac{\mathrm{d}\left(\boldsymbol{U}t\right)}{\mathrm{d}t} = \boldsymbol{V} + \boldsymbol{U}, \tag{9.6}$$

where we set $\mathrm{d}\boldsymbol{U}/\mathrm{d}t = 0$ since $\boldsymbol{U}$ 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 $\boldsymbol{U}$. Furthermore, as advertised, the acceleration in the two reference frames is identical

$$\overline{\boldsymbol{A}} = \frac{\mathrm{d}^2\overline{\boldsymbol{X}}}{\mathrm{d}\bar{t}^2} = \frac{\mathrm{d}\boldsymbol{V}}{\mathrm{d}t} = \boldsymbol{A}. \tag{9.7}$$

9.1.3 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 (9.3) or (9.4). 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 the system experiences the earth's rotation. In such cases the rotating earth reference frame is no longer a good approximation to an inertial frame.

9.1.4 Angular momentum

The angular momentum of a particle defined relative to a chosen origin is given by the vector cross product

$$\boldsymbol{L} = \boldsymbol{X} \wedge \boldsymbol{P}. \tag{9.8}$$

Making use of Newton's second law (9.3) leads to the equation of motion for the angular momentum

$$\frac{\mathrm{d}\boldsymbol{L}}{\mathrm{d}t} = \boldsymbol{X} \wedge \boldsymbol{F}, \tag{9.9}$$

where we set $\dot{\boldsymbol{X}} \wedge \boldsymbol{P} = \dot{\boldsymbol{X}} \wedge m\,\dot{\boldsymbol{X}} = 0$. The cross product, $\boldsymbol{X} \wedge \boldsymbol{F}$, is the *torque* acting on the system relative to the origin. Equation (9.9) leads to our second conservation law: a particle has a constant angular momentum when experiencing zero torques. We emphasize that these statements are dependent on the choice of origin for the angular momentum and the corresponding torques.

9.1.5 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

$$W = \int_{\boldsymbol{x}(t_1)}^{\boldsymbol{x}(t_2)} \boldsymbol{F} \cdot \mathrm{d}\boldsymbol{x}, \tag{9.10}$$

where $\boldsymbol{x}(t_1)$ and $\boldsymbol{x}(t_2)$ are the spatial coordinates of the endpoints for the trajectory at times t_1 and t_2, and $\mathrm{d}\boldsymbol{x}$ is the differential vector increment along the trajectory. Since the particle is moving along its dynamical trajectory, we can write $\mathrm{d}\boldsymbol{x} = \boldsymbol{V}\,\mathrm{d}t$ and make use of Newton's equation of motion (9.4) to reach

$$W = \int_{\boldsymbol{x}(t_1)}^{\boldsymbol{x}(t_2)} \boldsymbol{F} \cdot \mathrm{d}\boldsymbol{x} = \frac{m}{2} \int_{\boldsymbol{x}(t_1)}^{\boldsymbol{x}(t_2)} \frac{\mathrm{d}(\boldsymbol{V} \cdot \boldsymbol{V})}{\mathrm{d}t}\,\mathrm{d}t = K(t_2) - K(t_1), \tag{9.11}$$

where we identified the kinetic energy of the particle

$$K = \frac{m}{2}\,\boldsymbol{V} \cdot \boldsymbol{V}. \tag{9.12}$$

We conclude that the work done on the particle over a time interval is equal to its change in kinetic energy.

9.1.6 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*

$$\boldsymbol{F} = -\nabla V, \tag{9.13}$$

where V is called the *potential* or the *potential energy*. Inserting the potential into the work equation (9.11) leads to an expression of mechanical energy conservation

$$W = \int_{\boldsymbol{x}(t_1)}^{\boldsymbol{x}(t_2)} \boldsymbol{F} \cdot \mathrm{d}\boldsymbol{x} = \Delta K = -\Delta V \Longrightarrow K(t_2) + V(t_2) = K(t_1) + V(t_1). \tag{9.14}$$

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.

9.1.7 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*

$$\boldsymbol{F}_{\text{Rayleigh}} = -\lambda\, m\, \boldsymbol{V}, \tag{9.15}$$

where $\lambda > 0$ is a constant with dimensions of inverse time. Newton's equation of motion with Rayleigh drag (and no other forces) takes the form

$$\frac{\mathrm{d}\boldsymbol{V}}{\mathrm{d}t} = -\lambda\, \boldsymbol{V}, \tag{9.16}$$

which acts to drive the velocity to zero. We see that Rayleigh drag is not Galilean invariant since it is dependent on velocity. In the presence of Rayleigh drag (and no other forces), the kinetic energy is dissipated in time according to to the exponential decay

$$K(t) = K(0)\, e^{-2\,\lambda\, t}. \tag{9.17}$$

9.1.8 Comments

At this point we move on 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 here as a reminder of the fundamental physical notions. Such reminders are particularly useful later in the chapter when encountering mathematical manipulations with spherical coordinates in a rotating reference frame!

Every book on classical mechanics has some form of the material presented in this section, such as that from *Goldstein* (1980) targeted at the entering physical graduate student or *Marion and Thornton* (1988) targeted at second year undergraduates.

9.2 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 9.1). 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 velocity of the earth is given by

$$\Omega = \frac{2\pi + 2\pi/365.24}{86400\text{s}} = \left[\frac{\pi}{43082}\right] \text{s}^{-1} = 7.2921 \times 10^{-5}\ \text{s}^{-1}. \tag{9.18}$$

The earth's angular velocity, both its direction and its magnitude, is assumed constant in time for purposes of geophysical fluid mechanics

$$\frac{\mathrm{d}\vec{\Omega}}{\mathrm{d}t} = 0. \tag{9.19}$$

The angular velocity (9.18) seems quite small. However, a terrestrial reference frame on the surface of the earth, undergoing *solid body motion*, moves with linear velocity

$$U_{\text{solid-body}} = \Omega\, R_{\text{e}} \approx 465\ \text{m s}^{-1} = 1672\ \text{km hr}^{-1}, \tag{9.20}$$

where we set the earth's radius to

$$R_{\text{e}} = 6.367 \times 10^6\,\text{m}. \tag{9.21}$$

This speed is quite large relative to motion of the planet. Yet that is the point: this speed is not relative to the planet but instead it is the speed of the rotating planet. Motion of the atmosphere and ocean fluids are generally quite close to the solid-body, so that it makes sense to describe geophysical fluid motion from the non-inertial rotating terrestrial reference frame.

9.3 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 9.2 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

Figure 9.1: *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.*

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 Sections 10.7 and 10.8, the Coriolis acceleration appearing 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 9.2: *The position vector, $\boldsymbol{X}(t)$, 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 $\boldsymbol{U}_{solid} = \boldsymbol{\Omega} \wedge \boldsymbol{X}$ (equation (9.40)). For a particle on the earth's surface at the equator, the solid body speed is $\Omega\, R_e = 7.2921 \times 10^{-5}\ s^{-1} \times 6.371 \times 10^6\ m \approx 465\ m\ s^{-1} = 1672\ km\ hr^{-1}$. This speed is much greater than that of a fluid element relative to the moving earth. Hence, geophysical fluids are moving in near solid-body motion with the planet. It is motion relative to the solid-body that is of interest, thus motivating us to work in a rotating terrestrial reference frame.*

9.4 Rotation of a vector

How does a vector change under a solid-body rotation such as that shown in Figure 9.3? 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? Answering this question is fundamental to the kinematics of rotational motion. To develop an answer, observe a pure rotation does not change the magnitude of a vector, so that

$$|\boldsymbol{X}(t)| = |\boldsymbol{X}(t+\delta t)|. \tag{9.22}$$

The condition (9.22) can be written as

$$\frac{\mathrm{d}(\boldsymbol{X}\cdot\boldsymbol{X})}{\mathrm{d}t} = 0, \tag{9.23}$$

which leads to the constraint

$$\boldsymbol{X}\cdot\frac{\mathrm{d}\boldsymbol{X}}{\mathrm{d}t} = 0. \tag{9.24}$$

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.

9.4.1 Change in direction: brief derivation

Referring to Figure 9.3, we see that the infinitesimal difference $\boldsymbol{X}(t+\delta t) - \boldsymbol{X}(t)$ equals to the vector cross product of the angular velocity with the position vector

$$\boldsymbol{X}(t+\delta t) - \boldsymbol{X}(t) = \delta t\,\boldsymbol{\Omega}\wedge\boldsymbol{X}(t). \tag{9.25}$$

Dividing by δt leads to

$$\frac{\mathrm{d}\boldsymbol{X}}{\mathrm{d}t} = \boldsymbol{\Omega}\wedge\boldsymbol{X} \qquad \text{pure rotation.} \tag{9.26}$$

Note that this evolution equation satisfies the constraint (9.24) since $\boldsymbol{X}\cdot(\boldsymbol{\Omega}\wedge\boldsymbol{X}) = 0$, meaning that the magnitude of the vector indeed remains fixed.

9.4.2 Change in direction: detailed derivation for planar motion

To determine the change in direction generated by a pure rotation, we first consider the simplified case of planar rotation. Let $\boldsymbol{\Omega}$ be entirely in the vertical, and let $\boldsymbol{X}$ be confined to the horizontal plane. In a time increment δt, the vector $\boldsymbol{X}(t)$ is rotated by an angle

$$\delta\vartheta = |\boldsymbol{\Omega}|\,\delta t \tag{9.27}$$

to $\boldsymbol{X}(t+\delta t)$. In the limit of small $\delta\vartheta$, the difference vector, $\boldsymbol{X}(t+\delta t) - \boldsymbol{X}(t)$, is perpendicular to $\boldsymbol{X}(t)$ and is of magnitude equal to the arc length

$$\delta s = |\boldsymbol{X}(t)|\,\delta\vartheta = |\boldsymbol{X}(t)|\,|\boldsymbol{\Omega}|\,\delta t. \tag{9.28}$$

We observe that the vector $\boldsymbol{\Omega}\wedge\boldsymbol{X}(t)$ points in the same direction as $\boldsymbol{X}(t+\delta t) - \boldsymbol{X}(t)$ and is of length $|\boldsymbol{X}(t)|\,|\boldsymbol{\Omega}|$. We conclude that

$$\boldsymbol{X}(t+\delta t) - \boldsymbol{X}(t) = \boldsymbol{\Omega}\wedge\boldsymbol{X}(t)\,\delta t. \tag{9.29}$$

Figure 9.3: *The change in a vector under a pure rotation leaves its origin fixed and the vector magnitude unchanged, $|\boldsymbol{X}(t)| = |\boldsymbol{X}(t+\delta t)|$. Only the vector direction changes, here shown to be $\delta\vartheta = \Omega\,\delta t$. Infinitesimal changes generated by the angular velocity $\boldsymbol{\Omega}$ lead to the vector differences $\boldsymbol{X}(t+\delta t) - \boldsymbol{X}(t) = \delta t\,\boldsymbol{\Omega}\wedge\boldsymbol{X}(t)$.*

Dividing through by δt and taking the limit $\delta t \to 0$ gives

$$\frac{\mathrm{d}\boldsymbol{X}}{\mathrm{d}t} = \boldsymbol{\Omega}\wedge\boldsymbol{X} \qquad \text{pure rotation.} \tag{9.30}$$

The proof for the case in which $\boldsymbol{\Omega}$ has a component along $\boldsymbol{X}$ 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.

9.4.3 Comments

We make use of the result (9.26) in many places when working with rotating physics. One place of particular note concerns the changes to unit vectors under rotation. We emphasize that Cartesian unit vectors exhibit changes solely due to pure rotations, so that, for example,

$$\frac{\mathrm{d}\hat{\boldsymbol{x}}}{\mathrm{d}t} = \boldsymbol{\Omega}\wedge\hat{\boldsymbol{x}}. \tag{9.31}$$

Other unit vectors, such as those for spherical coordinates and polar coordinates, also change due to the solid-body rotation. Additionally, these non-Cartesian unit vectors change when their orientation is modified relative to the Cartesian coordinate axes at a rate distinct from the solid-body. We further discuss this point in Section 9.6.2 and detail the spherical coordinate case in Sections 9.9.2 and 9.11.

9.5 Some tensor algebra

In Part I of this book, we detailed the use of tensor analysis for geophysical motions. We here summarize the salient points for the reader who skimmed that earlier discussion.

9.5.1 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

24.2). However, we make routine use of spherical coordinates when describing geophysical motion, and cylindrical-polar coordinates for studies of rotating tank experiments (see Exercise 24.4 and Section 29.3). Finally, we use generalized vertical coordinates in the description of stratified flows (Chapters 8, 17, and 38). 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.

9.5.2 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

$$\vec{X} = \sum_{a=1}^{3} \xi^a \, \vec{e}_a = \xi^a \, \vec{e}_a, \tag{9.32}$$

where the Einstein summation convention is defined by the final equality. In this equation,

$$\vec{e}_a = (\vec{e}_1, \vec{e}_2, \vec{e}_3) \tag{9.33}$$

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 7.2.2). Note that we commonly make use of the boldface notation for a vector rather than the arrowed symbol (Section 5.6)

$$\boldsymbol{X} = \vec{X} = \xi^a \, \vec{e}_a. \tag{9.34}$$

The basis vectors in equation (9.32) 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 our work we can keep this mathematical framework at a modest distance, with exposure only in selected places. The key crutch we are relying 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 even the marriage of space and time afforded by special relativity. These assumptions simplify much of our work with 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*, $\widetilde{e}^a$. Duality here is defined using the familiar (Euclidean) inner product

$$\vec{e}_a \cdot \widetilde{e}^b = \delta^b_{\ a}, \tag{9.35}$$

with $\delta^b_{\ a}$ the Kronecker delta tensor

$$\delta^b_{\ a} = \begin{cases} 1 & \text{if } b = a \\ 0 & \text{if } b \neq a. \end{cases} \tag{9.36}$$

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 7 and 8.

9.6 The velocity vector

The velocity is the time derivative of the position vector

$$\boldsymbol{V} = \frac{\mathrm{d}\boldsymbol{X}}{\mathrm{d}t}. \tag{9.37}$$

In general, both the coordinate representation and the basis vectors are time dependent, so that the velocity has two contributions, one from the time dependence of the coordinates and another from the basis vectors

$$\boldsymbol{V} = \frac{\mathrm{d}\boldsymbol{X}}{\mathrm{d}t} = \frac{\mathrm{d}\left(\xi^a\,\vec{e}_a\right)}{\mathrm{d}t} = \frac{\mathrm{d}\xi^a}{\mathrm{d}t}\,\vec{e}_a + \xi^a\,\frac{\mathrm{d}\vec{e}_a}{\mathrm{d}t}. \tag{9.38}$$

9.6.1 Coordinate velocity

The first term on the right hand side of equation (9.38) is the velocity as measured within the chosen reference frame using the chosen coordinates

$$\boldsymbol{V}_{\text{coord}} \equiv \frac{\mathrm{d}\xi^a}{\mathrm{d}t}\,\vec{e}_a. \tag{9.39}$$

That is, this is the contribution 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.

9.6.2 Changes to the basis vectors

The second term on the right hand side of equation (9.38) 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 9.4)

$$\boldsymbol{U}_{\text{solid}} = \boldsymbol{\Omega} \wedge \boldsymbol{X}. \tag{9.40}$$

- ROTATION RELATIVE TO SOLID-BODY: Another manner to change a vector through rotation occurs when the basis vector maintains a constant magnitude but changes its direction at a different rate than 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).

9.7 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

$$\boldsymbol{A} = \frac{\mathrm{d}\boldsymbol{V}}{\mathrm{d}t} = \frac{\mathrm{d}^2\boldsymbol{X}}{\mathrm{d}t^2}. \tag{9.41}$$

This equation is independent of any coordinate representation. Correspondingly, the physical and geometrical content are manifest. When introducing a coordinate representation, the expression becomes subject to details of the chosen coordinates that can obscure the underlying geometric basis.

Consequently, it is important to keep the geometric form in mind when offering an interpretation for coordinate dependent terms.

Introducing a coordinate representation $\boldsymbol{X} = \xi^a \, \vec{e}_a$ into the acceleration (9.41), and making use of the chain rule, leads to

$$\boldsymbol{A} = \frac{\mathrm{d}}{\mathrm{d}t} \frac{\mathrm{d}\boldsymbol{X}}{\mathrm{d}t} \tag{9.42a}$$

$$= \frac{\mathrm{d}}{\mathrm{d}t} \frac{\mathrm{d}\,(\xi^a \, \vec{e}_a)}{\mathrm{d}t} \tag{9.42b}$$

$$= \frac{\mathrm{d}}{\mathrm{d}t} \left[\frac{\mathrm{d}\,\xi^a}{\mathrm{d}t} \, \vec{e}_a + \xi^a \, \frac{\mathrm{d}\,\vec{e}_a}{\mathrm{d}t} \right] \tag{9.42c}$$

$$= \frac{\mathrm{d}^2 \xi^a}{\mathrm{d}t^2} \, \vec{e}_a + 2 \, \frac{\mathrm{d}\xi^a}{\mathrm{d}t} \, \frac{\mathrm{d}\vec{e}_a}{\mathrm{d}t} + \xi^a \, \frac{\mathrm{d}^2 \vec{e}_a}{\mathrm{d}t^2}. \tag{9.42d}$$

The first term on the right hand side is the acceleration of the coordinate representation as measured in the rotating reference frame

$$\boldsymbol{A}_{\mathrm{coord}} \equiv \frac{\mathrm{d}^2 \xi^a}{\mathrm{d}t^2} \, \vec{e}_a. \tag{9.43}$$

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 9.6.2).

Some presentations of the kinematic result (9.42d) 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 results from the need to take two time derivatives of the basis vectors as part of a representation of acceleration. It appears throughout rotational physics as part of the Coriolis acceleration.

9.8 Representing the position vector

We make use of some results from Section 7.2 relating Cartesian and spherical coordinates and as furthermore defined by Figure 7.1. Starting with the position vector, we introduce the *planetary* Cartesian basis vectors, $(\hat{\boldsymbol{x}}, \hat{\boldsymbol{y}}, \hat{\boldsymbol{z}})$, and corresponding spherical basis vectors, $(\hat{\boldsymbol{\lambda}}, \hat{\boldsymbol{\phi}}, \hat{\boldsymbol{r}})$. We thus have the suite of equivalent expressions for the position of a particle moving around the sphere

$$\boldsymbol{X} = x \, \hat{\boldsymbol{x}} + y \, \hat{\boldsymbol{y}} + z \, \hat{\boldsymbol{z}} \tag{9.44a}$$

$$= (r \cos\phi \, \cos\lambda) \, \hat{\boldsymbol{x}} + (r \cos\phi \, \sin\lambda) \, \hat{\boldsymbol{y}} + (r \sin\phi) \, \hat{\boldsymbol{z}} \tag{9.44b}$$

$$= r \, \hat{\boldsymbol{r}} \tag{9.44c}$$

$$= |\boldsymbol{X}| \, \hat{\boldsymbol{r}}. \tag{9.44d}$$

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.

9.9 Representing the velocity vector

As seen in Section 9.6, the inertial velocity vector has a coordinate representation written as

$$\boldsymbol{V} = \frac{\mathrm{d}\boldsymbol{X}}{\mathrm{d}t} = \frac{\mathrm{d}(\xi^a \, \vec{e}_a)}{\mathrm{d}t} = \frac{\mathrm{d}\xi^a}{\mathrm{d}t} \, \vec{e}_a + \xi^a \, \frac{\mathrm{d}\vec{e}_a}{\mathrm{d}t}. \tag{9.45}$$

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.

9.9.1 Planetary Cartesian coordinate representation

The basis vectors for the Cartesian coordinates, $(\vec{e}_1, \vec{e}_2, \vec{e}_3) = (\hat{\boldsymbol{x}}, \hat{\boldsymbol{y}}, \hat{\boldsymbol{z}})$, 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 24.2 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.

The angular velocity is oriented around the polar axis

$$\boldsymbol{\Omega} = \Omega\, \hat{\boldsymbol{z}}, \tag{9.46}$$

so that the solid-body velocity only has components in the $\hat{\boldsymbol{x}}$ and $\hat{\boldsymbol{y}}$ directions

$$\boldsymbol{U}_{\text{solid}} = \boldsymbol{\Omega} \wedge \boldsymbol{X} = \Omega\,(-\hat{\boldsymbol{x}}\, y + \hat{\boldsymbol{y}}\, x). \tag{9.47}$$

The inertial velocity thus has the following representation in terms of planetary Cartesian coordinates in the rotating reference frame

$$\boldsymbol{V} = \left[-y\,\Omega + \frac{\mathrm{d}x}{\mathrm{d}t}\right]\hat{\boldsymbol{x}} + \left[x\,\Omega + \frac{\mathrm{d}y}{\mathrm{d}t}\right]\hat{\boldsymbol{y}} + \frac{\mathrm{d}z}{\mathrm{d}t}\,\hat{\boldsymbol{z}} = \boldsymbol{V}_{\text{Cartesian}} + \boldsymbol{\Omega} \wedge \boldsymbol{X} = \boldsymbol{V}_{\text{Cartesian}} + \boldsymbol{U}_{\text{solid}}, \tag{9.48}$$

where we defined the Cartesian velocity vector

$$\boldsymbol{V}_{\text{Cartesian}} \equiv \frac{\mathrm{d}x}{\mathrm{d}t}\,\hat{\boldsymbol{x}} + \frac{\mathrm{d}y}{\mathrm{d}t}\,\hat{\boldsymbol{y}} + \frac{\mathrm{d}z}{\mathrm{d}t}\,\hat{\boldsymbol{z}}, \tag{9.49}$$

which is the velocity as measured in the rotating reference frame when using planetary Cartesian coordinates.

9.9.2 Planetary spherical coordinate representation

The position vector in the planetary spherical coordinate representation is given by

$$\boldsymbol{X} = r\,\hat{\boldsymbol{r}}. \tag{9.50}$$

The basis vector $\hat{\boldsymbol{r}}$ is normalized, so that its evolution arises just from rotational motion. 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 change by taking the time derivative of $\hat{\boldsymbol{r}}$ as given by equation (7.30c)

$$\frac{\mathrm{d}\hat{\boldsymbol{r}}}{\mathrm{d}t} = \frac{\mathrm{d}}{\mathrm{d}t}\left[\hat{\boldsymbol{x}}\,\cos\lambda\,\cos\phi + \hat{\boldsymbol{y}}\sin\lambda\,\cos\phi + \hat{\boldsymbol{z}}\sin\phi\right] \tag{9.51a}$$

$$= \cos\phi\left[\frac{\mathrm{d}\lambda}{\mathrm{d}t} + \Omega\right]\hat{\boldsymbol{\lambda}} + \frac{\mathrm{d}\phi}{\mathrm{d}t}\,\hat{\boldsymbol{\phi}}. \tag{9.51b}$$

Consequently, the inertial velocity has the following spherical coordinate representation

$$\boldsymbol{V} = \frac{\mathrm{d}\boldsymbol{X}}{\mathrm{d}t} \tag{9.52a}$$

$$= \frac{\mathrm{d}(r\,\hat{\boldsymbol{r}})}{\mathrm{d}t} \tag{9.52b}$$

$$= \frac{\mathrm{d}r}{\mathrm{d}t}\,\hat{\boldsymbol{r}} + r\,\frac{\mathrm{d}\hat{\boldsymbol{r}}}{\mathrm{d}t} \tag{9.52c}$$

$$= r_\perp \left[\frac{\mathrm{d}\lambda}{\mathrm{d}t} + \Omega\right]\hat{\boldsymbol{\lambda}} + r\,\frac{\mathrm{d}\phi}{\mathrm{d}t}\,\hat{\boldsymbol{\phi}} + \frac{\mathrm{d}r}{\mathrm{d}t}\,\hat{\boldsymbol{r}} \tag{9.52d}$$

$$= (u + r_\perp\,\Omega)\,\hat{\boldsymbol{\lambda}} + v\,\hat{\boldsymbol{\phi}} + w\,\hat{\boldsymbol{r}} \tag{9.52e}$$

$$= \boldsymbol{V}_{\text{spherical}} + \boldsymbol{U}_{\text{solid}}. \tag{9.52f}$$

In this equation we introduced the spherical coordinate velocity vector

$$\boldsymbol{V}_{\text{spherical}} = r_\perp\,\frac{\mathrm{d}\lambda}{\mathrm{d}t}\,\hat{\boldsymbol{\lambda}} + r\,\frac{\mathrm{d}\phi}{\mathrm{d}t}\,\hat{\boldsymbol{\phi}} + \frac{\mathrm{d}r}{\mathrm{d}t}\,\hat{\boldsymbol{r}} = u\,\hat{\boldsymbol{\lambda}} + v\,\hat{\boldsymbol{\phi}} + w\,\hat{\boldsymbol{r}}, \tag{9.53}$$

where

$$u = r_\perp\,\frac{\mathrm{d}\lambda}{\mathrm{d}t} \qquad \text{and} \qquad v = r\,\frac{\mathrm{d}\phi}{\mathrm{d}t} \qquad \text{and} \qquad w = \frac{\mathrm{d}r}{\mathrm{d}t}, \tag{9.54}$$

are components to the spherical velocity vector, and with

$$r_\perp = r\cos\phi \tag{9.55}$$

the distance to the polar axis. The spherical velocity, $\boldsymbol{V}_{\text{spherical}}$, 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

$$\boldsymbol{U}_{\text{solid}} = \boldsymbol{\Omega} \wedge \boldsymbol{X} = r_\perp\,\Omega\,\hat{\boldsymbol{\lambda}}. \tag{9.56}$$

That is, the solid-body velocity is purely zonal.

9.9.3 Axial angular momentum

As seen in Section 10.6, the zonal component of the inertial velocity equals to the axial angular momentum per unit mass

$$L^z = m\,\hat{\boldsymbol{\lambda}} \cdot \boldsymbol{V} = m\,(u + r_\perp\,\Omega). \tag{9.57}$$

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 10.6, this conservation law offers a very important constraint on the particle trajectory. It also plays a role in the motion of geophysical fluids (Section 18.4).

9.10 Planetary Cartesian representation of acceleration

The inertial acceleration vector is given by the second time derivative of the position vector

$$\boldsymbol{A} = \frac{\mathrm{d}\boldsymbol{V}}{\mathrm{d}t} = \frac{\mathrm{d}^2\boldsymbol{X}}{\mathrm{d}t^2}. \tag{9.58}$$

We here consider its representation using planetary Cartesian coordinates (x, y, z) and the Cartesian basis $(\vec{e}_1, \vec{e}_2, \vec{e}_3) = (\hat{\boldsymbol{x}}, \hat{\boldsymbol{y}}, \hat{\boldsymbol{z}})$.

9.10.1 Planetary Cartesian representation

For our study of geophysical fluid motion, we assume the planetary angular velocity, $\boldsymbol{\Omega}$, is a constant in time

$$\frac{\mathrm{d}\boldsymbol{\Omega}}{\mathrm{d}t} = 0. \tag{9.59}$$

Making use of the results from Section 9.4 leads to

$$\frac{\mathrm{d}\vec{e}_a}{\mathrm{d}t} = \boldsymbol{\Omega} \wedge \vec{e}_a \tag{9.60}$$

and

$$\frac{\mathrm{d}^2\vec{e}_a}{\mathrm{d}t^2} = \frac{\mathrm{d}}{\mathrm{d}t}\left(\boldsymbol{\Omega} \wedge \vec{e}_a\right) = \boldsymbol{\Omega} \wedge \frac{\mathrm{d}\vec{e}_a}{\mathrm{d}t} = \boldsymbol{\Omega} \wedge \left(\boldsymbol{\Omega} \wedge \vec{e}_a\right), \tag{9.61}$$

which yields the inertial acceleration

$$\begin{aligned}
\boldsymbol{A} &= \frac{\mathrm{d}}{\mathrm{d}t}\frac{\mathrm{d}\boldsymbol{X}}{\mathrm{d}t} && (9.62\text{a})\\
&= \frac{\mathrm{d}^2\xi^a}{\mathrm{d}t^2}\,\vec{e}_a + 2\,\frac{\mathrm{d}\xi^a}{\mathrm{d}t}\left(\boldsymbol{\Omega} \wedge \vec{e}_a\right) + \xi^a\,\boldsymbol{\Omega} \wedge \left(\boldsymbol{\Omega} \wedge \vec{e}_a\right) && (9.62\text{b})\\
&= \boldsymbol{A}_{\text{Cartesian}} + 2\,\boldsymbol{\Omega} \wedge \boldsymbol{V}_{\text{Cartesian}} + \boldsymbol{\Omega} \wedge \left(\boldsymbol{\Omega} \wedge \boldsymbol{X}\right) && (9.62\text{c})\\
&= \boldsymbol{A}_{\text{Cartesian}} + 2\,\boldsymbol{\Omega} \wedge \boldsymbol{V}_{\text{Cartesian}} - \Omega^2\left(x\,\hat{\boldsymbol{x}} + y\,\hat{\boldsymbol{y}}\right) && (9.62\text{d})\\
&= \boldsymbol{A}_{\text{Cartesian}} - \boldsymbol{A}_{\text{Coriolis}} - \boldsymbol{A}_{\text{centrifugal}}. && (9.62\text{e})
\end{aligned}$$

The inertial acceleration is thus decomposed into three terms. The first contribution is the Cartesian expression

$$\boldsymbol{A}_{\text{Cartesian}} = \frac{\mathrm{d}^2x}{\mathrm{d}t^2}\,\hat{\boldsymbol{x}} + \frac{\mathrm{d}^2y}{\mathrm{d}t^2}\,\hat{\boldsymbol{y}} + \frac{\mathrm{d}^2z}{\mathrm{d}t^2}\,\hat{\boldsymbol{z}}, \tag{9.63}$$

which is the coordinate acceleration measured in the rotating frame using planetary Cartesian coordinates. The second contribution on the right hand side of equation (9.62e) is minus the Coriolis acceleration

$$\boldsymbol{A}_{\text{Coriolis}} = -2\,\boldsymbol{\Omega} \wedge \boldsymbol{V}_{\text{Cartesian}}. \tag{9.64}$$

The Coriolis acceleration plays a fundamental role in geophysical fluid mechanics and will be central to our development in this book. The third contribution is the centripetal acceleration, which is also minus the centrifugal acceleration

$$\boldsymbol{A}_{\text{centripetal}} = -\boldsymbol{A}_{\text{centrifugal}} = \boldsymbol{\Omega} \wedge \left(\boldsymbol{\Omega} \wedge \boldsymbol{X}\right) = -\Omega^2\,\boldsymbol{X}. \tag{9.65}$$

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

$$\boldsymbol{A}_{\text{centrifugal}} = -\nabla\Phi_{\text{centrifugal}} \qquad \Longleftrightarrow \qquad \Phi_{\text{centrifugal}} \equiv -\frac{\Omega^2\,r_\perp^2}{2} = -\frac{\Omega^2\left(x^2+y^2\right)}{2}, \tag{9.66}$$

which is a form to be used in the following. 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. Its 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, or causes you to lean when moving around a corner in a fast car.

9.10.2 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

$$\boldsymbol{A}_{\text{Cartesian}} = \boldsymbol{A} + \boldsymbol{A}_{\text{Coriolis}} + \boldsymbol{A}_{\text{centrifugal}} \tag{9.67a}$$

$$= \boldsymbol{A} - 2\,\boldsymbol{\Omega} \wedge \boldsymbol{V}_{\text{Cartesian}} - \nabla\Phi_{\text{centrifugal}} \tag{9.67b}$$

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, $\boldsymbol{A}$, that is directly affected by forces such as gravitation.

- CORIOLIS: The Coriolis acceleration

$$\boldsymbol{A}_{\text{Coriolis}} = -2\,\boldsymbol{\Omega} \wedge \boldsymbol{V}_{\text{Cartesian}} = -2\,\Omega\,\hat{\boldsymbol{z}} \wedge \boldsymbol{V}_{\text{Cartesian}} = -2\,\Omega\left[-\frac{\mathrm{d}y}{\mathrm{d}t}\,\hat{\boldsymbol{x}} + \frac{\mathrm{d}x}{\mathrm{d}t}\,\hat{\boldsymbol{y}}\right], \tag{9.68}$$

  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 $(\hat{\boldsymbol{x}}, \hat{\boldsymbol{y}})$. This geometric results is to be expected since the Coriolis acceleration arises from rotation about the polar $\hat{\boldsymbol{z}}$ axis.

- CENTRIFUGAL: The centrifugal acceleration

$$\boldsymbol{A}_{\text{centrifugal}} = -\nabla\Phi_{\text{centrifugal}} = \Omega^2\,\boldsymbol{r}_\perp = \Omega^2\,(x\,\hat{\boldsymbol{x}} + y\,\hat{\boldsymbol{y}}) \tag{9.69}$$

  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 $(\hat{\boldsymbol{x}}, \hat{\boldsymbol{y}})$. The centrifugal acceleration is directed outward from (perpendicular to) to the polar axis of rotation. We see this orientation in Figure 9.5 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, $\boldsymbol{A}_{\text{centrifugal}} = -\nabla\Phi_{\text{centrifugal}}$ where $\Phi_{\text{centrifugal}} = -\Omega^2\,(x^2 + y^2)/2$ (equation (9.66)). Hence, the centrifugal acceleration can be combined with the gravitational acceleration in the equation of motion (see Section 9.12). 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 9.12.2.

9.10.3 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.

9.11 Spherical representation of acceleration

The spherical representation of the inertial velocity is given by equation (9.52f)

$$\boldsymbol{V} = \frac{\mathrm{d}\boldsymbol{X}}{\mathrm{d}t} = (u + r_\perp\,\Omega)\,\hat{\boldsymbol{\lambda}} + v\,\hat{\boldsymbol{\phi}} + w\,\hat{\boldsymbol{r}} = \boldsymbol{V}_{\text{sphere}} + r_\perp\,\Omega\,\hat{\boldsymbol{\lambda}}, \tag{9.70}$$

where we introduced the spherical velocity from equation (9.54)

$$\boldsymbol{V}_{\text{sphere}} \equiv u\,\hat{\boldsymbol{\lambda}} + v\,\hat{\boldsymbol{\phi}} + w\,\hat{\boldsymbol{r}}. \tag{9.71}$$

We will also make use of the notation for the zonal component of the inertial velocity,

$$u_{\mathrm{I}} = u + r_\perp\,\Omega. \tag{9.72}$$

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

$$\boldsymbol{A} = \frac{\mathrm{d}}{\mathrm{d}t}\left(u_{\mathrm{I}}\,\hat{\boldsymbol{\lambda}} + v\,\hat{\boldsymbol{\phi}} + w\,\hat{\boldsymbol{r}}\right) \tag{9.73a}$$

$$= \frac{\mathrm{d}u_{\mathrm{I}}}{\mathrm{d}t}\,\hat{\boldsymbol{\lambda}} + \frac{\mathrm{d}v}{\mathrm{d}t}\,\hat{\boldsymbol{\phi}} + \frac{\mathrm{d}w}{\mathrm{d}t}\,\hat{\boldsymbol{r}} + u_{\mathrm{I}}\,\frac{\mathrm{d}\hat{\boldsymbol{\lambda}}}{\mathrm{d}t} + v\,\frac{\mathrm{d}\hat{\boldsymbol{\phi}}}{\mathrm{d}t} + w\,\frac{\mathrm{d}\hat{\boldsymbol{r}}}{\mathrm{d}t}. \tag{9.73b}$$

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 (7.2.2), a bit of algebra yields the time derivatives

$$\frac{\mathrm{d}\hat{\boldsymbol{\lambda}}}{\mathrm{d}t} = \left[\Omega + \frac{\mathrm{d}\lambda}{\mathrm{d}t}\right]\,(\hat{\boldsymbol{\phi}}\,\sin\phi - \hat{\boldsymbol{r}}\cos\phi) \tag{9.74}$$

$$\frac{\mathrm{d}\hat{\boldsymbol{\phi}}}{\mathrm{d}t} = -\hat{\boldsymbol{\lambda}}\,\left[\Omega + \frac{\mathrm{d}\lambda}{\mathrm{d}t}\right]\sin\phi - \hat{\boldsymbol{r}}\,\dot{\phi} \tag{9.75}$$

$$\frac{\mathrm{d}\hat{\boldsymbol{r}}}{\mathrm{d}t} = \hat{\boldsymbol{\lambda}}\,\left[\frac{\mathrm{d}\lambda}{\mathrm{d}t} + \Omega\right]\cos\phi + \frac{\mathrm{d}\phi}{\mathrm{d}t}\,\hat{\boldsymbol{\phi}}. \tag{9.76}$$

We are thus led to the inertial acceleration components

$$\boldsymbol{A}\cdot\hat{\boldsymbol{\lambda}} = \frac{\mathrm{d}u_{\mathrm{I}}}{\mathrm{d}t} + \left[\frac{\mathrm{d}\lambda}{\mathrm{d}t} + \Omega\right]\,(w\cos\phi - v\sin\phi) \tag{9.77a}$$

$$\boldsymbol{A}\cdot\hat{\boldsymbol{\phi}} = \frac{\mathrm{d}v}{\mathrm{d}t} + \left[\frac{\mathrm{d}\lambda}{\mathrm{d}t} + \Omega\right]\,u_{\mathrm{I}}\,\sin\phi + w\,\frac{\mathrm{d}\phi}{\mathrm{d}t} \tag{9.77b}$$

$$\boldsymbol{A}\cdot\hat{\boldsymbol{r}} = \frac{\mathrm{d}w}{\mathrm{d}t} - \left[\frac{\mathrm{d}\lambda}{\mathrm{d}t} + \Omega\right]\,u_{\mathrm{I}}\,\cos\phi - v\,\frac{\mathrm{d}\phi}{\mathrm{d}t}. \tag{9.77c}$$

Use of the identities

$$u = r_\perp\,\frac{\mathrm{d}\lambda}{\mathrm{d}t} \qquad \text{and} \qquad u_{\mathrm{I}} = u + r_\perp\,\Omega \qquad \text{and} \qquad \frac{\mathrm{d}u_{\mathrm{I}}}{\mathrm{d}t} = \frac{\mathrm{d}u}{\mathrm{d}t} + \Omega\,(w\cos\phi - v\sin\phi) \tag{9.78}$$

and some reorganization leads to the spherical coordinate representation of the inertial acceleration

$$\begin{aligned}\boldsymbol{A} = \hat{\boldsymbol{\lambda}}\,&\left[\frac{\mathrm{d}u}{\mathrm{d}t} + \frac{u\,(w - v\tan\phi)}{r} + 2\,\Omega\,(w\cos\phi - v\sin\phi)\right]\\ &+ \hat{\boldsymbol{\phi}}\left[\frac{\mathrm{d}v}{\mathrm{d}t} + \frac{v\,w + u^2\tan\phi}{r} + 2\,\Omega\,u\sin\phi + r_\perp\,\Omega^2\sin\phi\right]\\ &+ \hat{\boldsymbol{r}}\left[\frac{\mathrm{d}w}{\mathrm{d}t} - \frac{u^2 + v^2}{r} - 2\,\Omega\,u\cos\phi - r_\perp\,\Omega^2\cos\phi\right].\end{aligned} \tag{9.79}$$

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.

9.11.1 Decomposing the acceleration

We decompose the inertial acceleration (9.79) into the following terms

$$\boldsymbol{A} = \boldsymbol{A}_{\text{sphere}} + \boldsymbol{A}_{\text{metric}} - \boldsymbol{A}_{\text{Coriolis}} - \boldsymbol{A}_{\text{centrifugal}}, \tag{9.80}$$

with signs chosen so that in the rotating frame the acceleration is written

$$\underbrace{\boldsymbol{A}_{\text{sphere}} + \boldsymbol{A}_{\text{metric}}}_{\text{net spherical acceleration}} = \boldsymbol{A} + \boldsymbol{A}_{\text{Coriolis}} + \boldsymbol{A}_{\text{centrifugal}}. \tag{9.81}$$

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.

9.11.2 Spherical coordinate acceleration

The spherical coordinate acceleration is given by the time change in the spherical velocity components

$$\boldsymbol{A}_{\text{sphere}} = \frac{\mathrm{d}u}{\mathrm{d}t}\,\hat{\boldsymbol{\lambda}} + \frac{\mathrm{d}v}{\mathrm{d}t}\,\hat{\boldsymbol{\phi}} + \frac{\mathrm{d}w}{\mathrm{d}t}\,\hat{\boldsymbol{r}}. \tag{9.82}$$

This term has no contribution from changes to the spherical unit vectors.

9.11.3 Metric acceleration

The metric acceleration arises from changes to the spherical unit vectors and it is given by

$$\boldsymbol{A}_{\text{metric}} = \hat{\boldsymbol{\lambda}}\left[\frac{u\,(w - v\tan\phi)}{r}\right] + \hat{\boldsymbol{\phi}}\left[\frac{v\,w + u^2\tan\phi}{r}\right] - \hat{\boldsymbol{r}}\left[\frac{u^2+v^2}{r}\right] \tag{9.83a}$$

$$= \hat{\boldsymbol{\lambda}}\left[\frac{u\,(w\cos\phi - v\sin\phi)}{r\cos\phi}\right] + \hat{\boldsymbol{\phi}}\left[\frac{v\,w\cos\phi + u^2\sin\phi}{r\cos\phi}\right] - \hat{\boldsymbol{r}}\left[\frac{u^2+v^2}{r}\right] \tag{9.83b}$$

$$= \frac{1}{r}\left[u\tan\phi\,(\hat{\boldsymbol{r}}\wedge\boldsymbol{V}_{\text{sphere}}) + w\,\boldsymbol{U}_{\text{sphere}} - \hat{\boldsymbol{r}}\,\boldsymbol{U}_{\text{sphere}}\cdot\boldsymbol{U}_{\text{sphere}}\right], \tag{9.83c}$$

where we wrote the horizontal (angular) and vertical (radial) components of the spherical velocity according to

$$\boldsymbol{V}_{\text{sphere}} = \boldsymbol{U}_{\text{sphere}} + \hat{\boldsymbol{r}}\,w = \hat{\boldsymbol{\lambda}}\,u + \hat{\boldsymbol{\phi}}\,v + \hat{\boldsymbol{r}}\,w. \tag{9.84}$$

For purposes of developing a kinetic energy budget, note that

$$\boldsymbol{V}_{\text{sphere}}\cdot\boldsymbol{A}_{\text{metric}} = 0, \tag{9.85}$$

so that

$$\boldsymbol{V}\cdot\boldsymbol{A}_{\text{metric}} = \Omega\,u\,r_{\perp}\,(w - v\tan\phi) = \Omega\,u\,r\,(w\cos\phi - v\sin\phi). \tag{9.86}$$

9.11.4 Centrifugal acceleration

The spherical coordinate representation of the centrifugal acceleration is given by

$$\boldsymbol{A}_{\text{centrifugal}} = -\nabla\Phi_{\text{centrifugal}} = \Omega^2\,(x\,\hat{\boldsymbol{x}} + y\,\hat{\boldsymbol{y}}) = r_\perp\,\Omega^2\,(\hat{\boldsymbol{\phi}}\sin\phi - \hat{\boldsymbol{r}}\cos\phi). \tag{9.87}$$

The centrifugal acceleration points outward from the axis of rotation (see Figure 9.5 to be discussed later), so that it has no component in the longitudinal direction. For purposes of developing a kinetic energy budget (Section 10.5.1), note that

$$\boldsymbol{V}\cdot\boldsymbol{A}_{\text{centrifugal}} = \boldsymbol{V}_{\text{sphere}}\cdot\boldsymbol{A}_{\text{centrifugal}} = -\Omega^2\,r\cos\phi\,(w\cos\phi - v\sin\phi). \tag{9.88}$$

9.11.5 Coriolis acceleration

The spherical coordinate representation of the Coriolis acceleration makes use of the spherical representation of the earth's angular velocity

$$\boldsymbol{\Omega} = \Omega\,\hat{\boldsymbol{z}} = \Omega\,(\hat{\boldsymbol{\phi}}\,\cos\phi + \hat{\boldsymbol{r}}\,\sin\phi) \tag{9.89}$$

to reach the form

$$\begin{aligned}
\boldsymbol{A}_{\text{Coriolis}} &= -2\,\boldsymbol{\Omega}\wedge\boldsymbol{V}_{\text{sphere}} && (9.90\text{a})\\
&= -2\,\Omega\,(\hat{\boldsymbol{\phi}}\,\cos\phi + \hat{\boldsymbol{r}}\,\sin\phi)\wedge\boldsymbol{V}_{\text{sphere}} && (9.90\text{b})\\
&= -2\,\Omega\,(\hat{\boldsymbol{\phi}}\,\cos\phi + \hat{\boldsymbol{r}}\,\sin\phi)\wedge(u\,\hat{\boldsymbol{\lambda}} + v\,\hat{\boldsymbol{\phi}} + w\,\hat{\boldsymbol{r}}) && (9.90\text{c})\\
&= -2\,\Omega\left[\hat{\boldsymbol{\lambda}}\,(w\cos\phi - v\sin\phi) + \hat{\boldsymbol{\phi}}\,u\sin\phi - \hat{\boldsymbol{r}}\,u\cos\phi\right]. && (9.90\text{d})
\end{aligned}$$

Note that

$$\boldsymbol{V}_{\text{sphere}}\cdot\boldsymbol{A}_{\text{Coriolis}} = 0, \tag{9.91}$$

so that

$$\boldsymbol{V}\cdot\boldsymbol{A}_{\text{Coriolis}} = -2\,\Omega^2\,r\cos\phi\,(w\cos\phi - v\sin\phi). \tag{9.92}$$

We find it convenient to introduce a shorthand notation

$$\boldsymbol{f} = 2\,\Omega\,\sin\phi\,\hat{\boldsymbol{r}} \tag{9.93}$$

$$\boldsymbol{f}^* = 2\,\Omega\,\cos\phi\,\hat{\boldsymbol{\phi}}, \tag{9.94}$$

so that the Coriolis acceleration takes the form

$$\boldsymbol{A}_{\text{Coriolis}} = -(\boldsymbol{f} + \boldsymbol{f}^*)\wedge\boldsymbol{V}_{\text{sphere}}. \tag{9.95}$$

There are two contributions to the Coriolis acceleration: one from the radial and one from the meridional component of the earth's rotation vector.

9.11.6 Coriolis acceleration for large-scale motions

Let us again write the Coriolis acceleration in equation (9.90d), only now underlining two terms

$$\boldsymbol{A}_{\text{Coriolis}} = -2\,\Omega\left[\hat{\boldsymbol{\lambda}}\,(\underline{w\cos\phi} - v\sin\phi) + \hat{\boldsymbol{\phi}}\,u\sin\phi - \hat{\boldsymbol{r}}\,\underline{u\cos\phi}\right]. \tag{9.96}$$

For many applications in geophysical fluid dynamics, the term $\hat{\boldsymbol{r}}\,(2\,\Omega\,u\cos\phi)$ is much smaller than the competing gravitational acceleration that also contributes to the radial acceleration, thus

prompting $\hat{\boldsymbol{r}}\,(2\,\Omega\,u\cos\phi)$ to be dropped from the $\hat{\boldsymbol{r}}$ equation of motion.[1] Furthermore, the vertical velocity term is generally much smaller than the horizontal velocity term appearing in the $\hat{\boldsymbol{\lambda}}$ 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 24.1)

$$\boldsymbol{A}_{\text{Coriolis}}^{\text{large-scale}} \equiv -2\,\Omega\sin\phi\,(-\hat{\boldsymbol{\lambda}}\,v + \hat{\boldsymbol{\phi}}\,u) \equiv -f\,\hat{\boldsymbol{r}}\wedge\boldsymbol{V}_{\text{sphere}}. \tag{9.97}$$

For the last equality we introduced the Coriolis parameter

$$f \equiv 2\,\Omega\sin\phi. \tag{9.98}$$

As illustrated in Figure 9.4, we see that it is the local vertical component of the earth's angular rotation that plays the most important role in large-scale fluid mechanics

$$\boldsymbol{\Omega} = \Omega\,\hat{\boldsymbol{z}} = \Omega\,(\hat{\boldsymbol{\phi}}\,\cos\phi + \hat{\boldsymbol{r}}\,\sin\phi) \approx \Omega\sin\phi\,\hat{\boldsymbol{r}} = \boldsymbol{f}/2. \tag{9.99}$$

Figure 9.4: *This figure illustrates the two components of the earth's rotational velocity,* $\boldsymbol{\Omega} = \Omega\,\hat{\boldsymbol{z}} = \Omega\,(\hat{\boldsymbol{\phi}}\cos\phi + \hat{\boldsymbol{r}}\sin\phi)$. *The local vertical component,* $\boldsymbol{\Omega} \approx \Omega\sin\phi\,\hat{\boldsymbol{r}}$, *is the most important component for large scale geophysical fluid dynamics.*

9.12 Gravitational force and potential energy

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 10). We here discuss the gravitational potential energy and the associated gravitational force.

[1] The term $\hat{\boldsymbol{r}}\,(2\,\Omega\,u\cos\phi)$ is called the Eötvös correction in the study of marine gravity.

9.12.1 Newton's Gravitational Law

For a point particle of mass m moving around the sphere, potential energy (SI units kg m^2 s^{-2}) is associated with motion through the gravitational field. We write this potential energy as

$$P = m\,\Phi_{\rm e}, \tag{9.100}$$

where $\Phi_{\rm e}$ is the gravitational potential (SI units m^2 s^{-2}) determined from Newton's Law of Gravity. For a spherical mass, the gravitational potential is given by

$$\Phi_{\rm e} = -\frac{G\,M}{r} \tag{9.101}$$

where M is the mass of the sphere and

$$G = 6.674\times 10^{-11}\ {\rm N\ m^2\ kg^{-2}} = 6.674\times 10^{-11}\ {\rm m^3\ kg^{-1}\ s^{-2}} \tag{9.102}$$

is Newton's gravitational constant. The gradient of the gravitational potential gives the gravitational acceleration

$$\boldsymbol{g}_{\rm e} = -\nabla\Phi_{\rm e} = -\frac{G\,M}{r^2}\,\hat{\boldsymbol{r}} \tag{9.103}$$

so that the gravitational force acting on the particle is given by

$$\boldsymbol{F}_{\rm gravity} = m\,\boldsymbol{g}_{\rm e} = -m\,\nabla\Phi_{\rm e}. \tag{9.104}$$

For atmospheric and oceanic fluid dynamics, it is often sufficient to assume the gravitational acceleration is constant and equal to its value at the surface of the sphere.[2] In this case

$$\boldsymbol{g}_{\rm e} = -g_{\rm e}\,\hat{\boldsymbol{r}}, \tag{9.105}$$

where

$$g_{\rm e} = \frac{G\,M_{\rm e}}{R_{\rm e}^2} \approx 9.8\ {\rm m\ s^{-2}}. \tag{9.106}$$

To reach this value, we assumed a sphere of mass equal to the earth mass

$$M_{\rm e} = 5.977\times 10^{24}\ {\rm kg}, \tag{9.107}$$

and radius

$$R_{\rm e} = 6.371\times 10^{6}\ {\rm m} \tag{9.108}$$

determined so that the sphere has the same volume as the earth. The corresponding gravitational potential for the particle is given by

$$\Phi_{\rm e} = g_{\rm e}\,r, \tag{9.109}$$

and the gravitational potential energy is

$$m\,\Phi_{\rm e} = m\,g_{\rm e}\,r. \tag{9.110}$$

[2]The 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 43.

9.12.2 Effective gravitational force from the geopotential

We can combine the potential for the centrifugal acceleration as given by equation (9.69) with the gravitational potential (9.101), thus resulting in the *geopotential*

$$\Phi = r\,[g_{\mathrm{e}} - \boldsymbol{U}^2_{\mathrm{solid}}/(2r)]. \tag{9.111}$$

The contribution from the centrifugal term can be estimated by making use of terrestrial values, in which $R = R_{\mathrm{e}} = 6.371 \times 10^6$ m (equation (9.108)), and $\Omega_{\mathrm{e}} = 7.292 \times 10^{-5}$ s^{-1} (Section 9.2). The centrifugal term is its largest at the equator, $\phi = 0$, where

$$\frac{\boldsymbol{U}^2_{\mathrm{solid}}}{2\,R_{\mathrm{e}}} \approx 0.017 \text{ m s}^{-2}, \tag{9.112}$$

so that the ratio of the gravitational to centrifugal accelerations is (at most)

$$\frac{g_{\mathrm{e}}}{\boldsymbol{U}^2_{\mathrm{solid}}/(2R_{\mathrm{e}})} = \frac{M_{\mathrm{e}}\,G/R_{\mathrm{e}}^2}{\Omega_{\mathrm{e}}^2\,R_{\mathrm{e}}/2} \approx 576. \tag{9.113}$$

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 $\hat{\boldsymbol{r}}$ as pointing parallel to $\nabla\Phi$ rather than parallel to $\nabla\Phi_{\mathrm{e}}$. We have more to say on this topic in Section 9.13.3.

9.12.3 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).

9.13 Newton's law of motion

As seen in Section 9.1, 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

$$m\,\boldsymbol{A} = -m\,\nabla\Phi_{\mathrm{e}}. \tag{9.114}$$

This is a relatively simple equation of motion. However, it becomes more complex when moving to the rotating reference frame of terrestrial observers, and when represented using spherical coordinates.

9.13.1 Cartesian coordinate representation

The inertial acceleration using Cartesian coordinates is given by equation (9.67b)

$$\boldsymbol{A} = \boldsymbol{A}_{\mathrm{Cartesian}} - \boldsymbol{A}_{\mathrm{Coriolis}} - \boldsymbol{A}_{\mathrm{centrifugal}} \tag{9.115a}$$

$$= \boldsymbol{A}_{\mathrm{Cartesian}} + 2\,\boldsymbol{\Omega} \wedge \boldsymbol{V}_{\mathrm{Cartesian}} + \nabla\Phi_{\mathrm{centrifugal}}, \tag{9.115b}$$

so that the rotating frame Cartesian equation of motion is given by

$$\boldsymbol{A}_{\mathrm{Cartesian}} = -\nabla\Phi_{\mathrm{e}} - 2\,\boldsymbol{\Omega} \wedge \boldsymbol{V}_{\mathrm{Cartesian}} - \nabla\Phi_{\mathrm{centrifugal}} \tag{9.116a}$$

$$= -2\,\boldsymbol{\Omega} \wedge \boldsymbol{V}_{\mathrm{Cartesian}} - \nabla\Phi, \tag{9.116b}$$

where the geopotential is the sum of the gravitational and centrifugal potentials (equation (9.111))

$$\Phi = \Phi_{\mathrm{e}} + \Phi_{\mathrm{centrifugal}}. \tag{9.117}$$

We can write the equation of motion in component form by exposing the indices (Cartesian tensors) and using a dot for time derivative

$$\ddot{X}_a + 2\,\epsilon_{abc}\,\boldsymbol{\Omega}_b\,\dot{X}_c = -\partial_a\Phi. \tag{9.118}$$

We can write this equation in the standard vector form within the rotating reference frame

$$\frac{\mathrm{d}^2\boldsymbol{X}}{\mathrm{d}t^2} + 2\,\boldsymbol{\Omega}\wedge\dot{\boldsymbol{X}} = -\nabla\Phi. \tag{9.119}$$

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 (9.119) can be written

$$\frac{\mathrm{d}\boldsymbol{M}}{\mathrm{d}t} = \frac{\mathrm{d}}{\mathrm{d}t}(\boldsymbol{V}_{\mathrm{Cartesian}} + 2\,\boldsymbol{\Omega}\wedge\boldsymbol{X}) = -\nabla\Phi, \tag{9.120}$$

where we introduced the *potential momentum* per mass

$$\boldsymbol{M} = \boldsymbol{V}_{\mathrm{Cartesian}} + 2\,\boldsymbol{\Omega}\wedge\boldsymbol{X}. \tag{9.121}$$

We motivate the name in Section 10.2.

9.13.2 Spherical coordinate representation

We now make use of the acceleration written in spherical coordinates as given in Section 9.11

$$\boldsymbol{A}_{\mathrm{sphere}} + \boldsymbol{A}_{\mathrm{metric}} = \boldsymbol{A}_{\mathrm{Coriolis}} + \boldsymbol{A} + \boldsymbol{A}_{\mathrm{centrifugal}} \tag{9.122a}$$

$$= -2\,\boldsymbol{\Omega}\wedge\boldsymbol{V}_{\mathrm{sphere}} - \nabla\Phi. \tag{9.122b}$$

The effective gravitational force is not a central force due to the contribution from the centrifugal acceleration. We see this more explicitly by using the equations in Section 9.11 to write the spherical equations

$$\dot{u} + \frac{u\,(w - v\tan\phi)}{r} + 2\,\Omega\,(w\cos\phi - v\sin\phi) = 0 \tag{9.123}$$

$$\dot{v} + \frac{v\,w + u^2\tan\phi}{r} + 2\,\Omega\,u\sin\phi = -r_\perp\,\Omega^2\sin\phi \tag{9.124}$$

$$\dot{w} - \frac{u^2 + v^2}{r} - 2\,\Omega\,u\cos\phi = r_\perp\,\Omega^2\cos\phi - g_{\mathrm{e}}. \tag{9.125}$$

The presence of Ω^2 terms in both the meridional and radial equations signals the non-central nature of the effective gravitational acceleration.

9.13.3 Geopotential coordinates for slightly oblate spheriods

As we saw in Section 9.12.1, the radius of a sphere that best fits the volume of the earth is given by

$$R_{\mathrm{e}} = 6.367 \times 10^6 \,\mathrm{m}. \tag{9.126}$$

However, the non-central nature of the effective gravitational force leads to an oblate spheroidal shape for planets such as the earth. The result is a distinction between the equatorial and polar radii

$$R_{\mathrm{equator}} = 6.378 \times 10^6 \,\mathrm{m} \qquad \text{and} \qquad R_{\mathrm{pole}} = 6.357 \times 10^6 \,\mathrm{m}, \tag{9.127}$$

with a corresponding ratio

$$1 - \frac{R_{\mathrm{pole}}}{R_{\mathrm{equator}}} \approx 3 \times 10^{-3}. \tag{9.128}$$

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 planetary scale mechanics. In this case, the radial coordinate is constant on the oblate spheroid shaped geopotential.

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 non-radial component to the effective gravitational force. We illustrate the situation in Figure 9.5. Absorbing the centrifugal term into an effective gravitational potential then leads to the effective gravitational acceleration vector

$$-\nabla\Phi = -g\,\hat{\boldsymbol{r}}, \tag{9.129}$$

with g the effective gravitational acceleration. Using this convention, the particle equations of motion take the following form

$$\dot{u} + \frac{u\,(w - v\tan\phi)}{r} + 2\,\Omega\,(w\cos\phi - v\sin\phi) = 0 \tag{9.130}$$

$$\dot{v} + \frac{v\,w + u^2\tan\phi}{r} + 2\,\Omega\,u\sin\phi = 0 \tag{9.131}$$

$$\dot{w} - \frac{u^2 + v^2}{r} - 2\,\Omega\,u\cos\phi = -g. \tag{9.132}$$

Notably, the effective gravitational acceleration only impacts the radial equation of motion.

9.13.4 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.

9.14 Exercises

EXERCISE 9.1: WORKING THROUGH THE SPHERICAL ACCELERATION

Convince yourself that the spherical form of the acceleration given by equation (9.79) is indeed correct.

Figure 9.5: *This figure illustrates the geopotential vertical coordinate system used to study geophysical fluids. The left panel shows the non-central nature of the effective gravitational force $\boldsymbol{g}$, which is given by the sum of the gravitational acceleration $\boldsymbol{g}_r$ plus centrifugal acceleration $\boldsymbol{A}_{\text{centrifugal}}$. The gravitational acceleration points radially to the center of the sphere, whereas the centrifugal acceleration points outward from the axis of rotation. The right panel shows a reinterpreted vertical coordinate $r = R + z$ as measuring distance perpendicular to the geopotential surface (dotted surface). Using this geopotential vertical coordinate, the effective gravitational force is aligned with the vertical coordinate and so has no "horizontal" component ($g_\phi = 0$). This figure is not drawn to scale, with the oblate nature highly exaggerated compared to the real earth system (see equation (9.128)).*

EXERCISE 9.2: VELOCITY AND ACCELERATION IN CYLINDRICAL-POLAR COORDINATES

In Section 7.3 we worked through the transformation from Cartesian coordinates to cylindrical-polar coordinates for describing motion in a rotating reference frame. Here we develop an expression for the position, velocity, and acceleration vectors in a frame rotating about the vertical axis using 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 30.5) or when studying cyclostrophically balanced flow.

(A) Determine the representation of the inertial velocity vector, $\boldsymbol{V} = \mathrm{d}\boldsymbol{X}/\mathrm{d}t$, in terms of cylindrical-polar coordinates.

(B) Determine the representation of the inertial acceleration vector, $\boldsymbol{A} = \mathrm{d}\boldsymbol{V}/\mathrm{d}t$, in terms of cylindrical-polar coordinates.

(C) Writing the inertial acceleration in the form

$$\boldsymbol{A} = \boldsymbol{A}_{\text{cylindrical-polar}} - \boldsymbol{A}_{\text{centrifugal}} - \boldsymbol{A}_{\text{Coriolis}}, \tag{9.133}$$

give the mathematical expressions for the relative acceleration written in cylindrical-polar coordinates, $\boldsymbol{A}_{\text{cylindrical-polar}}$, the centrifugal acceleration, $\boldsymbol{A}_{\text{centrifugal}}$, and the Coriolis acceleration, $\boldsymbol{A}_{\text{Coriolis}}$.

EXERCISE 9.3: VELOCITY PROJECTED ONTO ACCELERATION

The kinetic energy per mass of a particle is given by

$$\mathcal{K} = \boldsymbol{V} \cdot \boldsymbol{V}/2, \tag{9.134}$$

where $\boldsymbol{V}$ is the inertial velocity of a particle. In an inertial reference frame it is trivial to show that

$$\frac{1}{2}\frac{\mathrm{d}\mathcal{K}}{\mathrm{d}t} = \boldsymbol{V} \cdot \boldsymbol{A} \tag{9.135}$$

through use of the chain rule, where $\boldsymbol{A} = \mathrm{d}\boldsymbol{V}/\mathrm{d}t$ is the inertial acceleration. Verify that this identity also holds in the rotating reference frame. For simplicity make use of planetary Cartesian coordinates.

EXERCISE 9.4: GEOMETRY OF CONSTANT GEOPOTENTIAL SURFACES

Here we examine some properties of the geopotential given by equation (9.111), where the squared solid-body speed is $\boldsymbol{U}^2_{\text{solid}} = (\Omega\, r \cos\phi)^2$.

(A) Sketch surfaces of constant geopotential according to the expression (9.111).

(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 (9.127).

EXERCISE 9.5: 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).

10

Symmetries and conservation laws

This chapter explores the constraints on particle motion due to conservation laws resulting from space-time symmetries. These ideas are fundamental to how we garner a qualitative understanding of motion, whereby it is generally more useful to know the dynamically conserved properties shared by all trajectories rather than details of any particular trajectory. Conservation laws provide useful predictive capabilities and help in designing analytical and numerical solution methods. These ideas extend to our study of geophysical fluid mechanics.

READER'S GUIDE TO THIS CHAPTER

This chapter is an extension of the particle mechanics studied in Chapter 9. We make use of ideas from this chapter throughout the book, and in particular when studying properties of the momentum equations for a geophysical fluid in Chapter 18. Furthermore, we here explore many facets of rotational physics, with this material central to geophysical fluid motions.

10.1 Trajectories and dynamical constraints **162**
10.1.1 Connecting symmetries to conservation laws 163
10.1.2 Further study 163
10.2 Potential momentum **163**
10.2.1 Basics 163
10.2.2 Comment about terminology 164
10.3 Inertial oscillations **165**
10.3.1 Oscillator equation 165
10.3.2 Particle trajectory and velocity 165
10.3.3 Period of inertial oscillations 166
10.3.4 Comments and further study 166
10.4 Dynamical constraints arising from geometric symmetries **167**
10.4.1 Linear momentum conservation 167
10.4.2 Potential momentum conservation 167
10.4.3 Angular momentum conservation 167
10.4.4 Kinetic energy and Galilean transformation 168
10.4.5 Cartesian expression in a rotating reference frame 168
10.4.6 Kinetic energy using spherical coordinates: Part I 169
10.4.7 Kinetic energy using spherical coordinates: Part II 169
10.5 Mechanical energy conservation **170**
10.5.1 General expression for kinetic energy evolution 170

10.5.2 Conservation of mechanical energy 170
10.5.3 Comments and further study 171
10.6 Axial angular momentum conservation **171**
10.6.1 Angular momentum 171
10.6.2 Conservation of axial angular momentum 172
10.7 Constraints from axial angular momentum conservation **172**
10.7.1 Deriving the constraints 173
10.7.2 Axial angular momentum conservation yields zonal Coriolis acceleration 173
10.7.3 Zonal acceleration induced by meridional motion 174
10.7.4 Zonal acceleration induced by radial motion 174
10.7.5 Zonal acceleration derived from axial angular momentum conservation 175
10.7.6 Nearly horizontal motions 175
10.8 More facets of motion on the rotating sphere **176**
10.8.1 Constant zonal motion is not steady motion 176
10.8.2 Coriolis acceleration in the meridional equation 177
10.8.3 Comments 177
10.9 Exercise **177**

10.1 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 dynamical insights often hidden when just having information about the trajectory. Constraints also provide predictive statements of great 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 angular momentum around the rotational axis.

Knowledge of the trajectory for a single particle does not readily reveal that this particle is constrained in its motion to conserve angular momentum. In contrast, knowing that all particles, no matter what trajectory, conserve angular momentum gives us the ability to understand basic properties of the motion and furthermore to predict the response under perturbations. We offer specifics to this predictive capability later in this chapter and provide more examples throughout this book. Such dynamical constraints are especially relevant for fluids since it is rare to determine the analytical expression for fluid motion, making the knowledge of constraints incredibly valuable.

10.1.1 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 E. Noether in 1918. Noether's Theorem is fundamental to all areas of physics. We will not delve into the mathematical details of Noether's Theorem. Instead, it is sufficient for our study to make use of it as a conceptual framework for understanding conservation laws. Namely, if there is a symmetry, then there is a corresponding conservation law.

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 9.1.7, 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 as well. 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 36).

10.1.2 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 amazing mathematician whose work, conducted under some very unfortunate circumstances, united symmetry and conservation laws.

10.2 Potential momentum

In Section 9.13.1 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.

10.2.1 Basics

Again as in Section 9.13.1, recall that if the planetary rotation is a constant in time, then the Cartesian equation of motion

$$\frac{\mathrm{d}}{\mathrm{d}t}\left[\dot{\boldsymbol{X}} + 2\,\boldsymbol{\Omega}\wedge\boldsymbol{X}\right] = -\nabla\Phi, \tag{10.1}$$

suggests we introduce the *potential momentum* per mass

$$\boldsymbol{M} \equiv \frac{\mathrm{d}\boldsymbol{X}}{\mathrm{d}t} + 2\,\boldsymbol{\Omega} \wedge \boldsymbol{X} = \hat{\boldsymbol{x}}\,(u - 2\,\Omega\,y) + \hat{\boldsymbol{y}}\,(v + 2\,\Omega\,x) + \hat{\boldsymbol{z}}\,w, \tag{10.2}$$

in which case

$$\frac{\mathrm{d}\boldsymbol{M}}{\mathrm{d}t} = -\nabla\Phi. \tag{10.3}$$

Now let $\hat{\boldsymbol{s}}$ be a unit vector tangent to the geopotential surface so that $\hat{\boldsymbol{s}} \cdot \nabla\Phi = 0$. Assuming the geopotential surface is time independent, the equation of motion (10.3) leads to

$$\frac{\mathrm{d}(\hat{\boldsymbol{s}} \cdot \boldsymbol{M})}{\mathrm{d}t} = 0. \tag{10.4}$$

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 10.1.1) 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 10.1 with a horizontal geopotential surface.

Figure 10.1: *The projection of the potential momentum onto the geopotential surface is a constant of the motion,* $\mathrm{d}(\hat{\boldsymbol{s}} \cdot \boldsymbol{M})/\mathrm{d}t = 0$. *Here the geopotential surface is the x-y-plane so that* $\hat{\boldsymbol{x}} \cdot \boldsymbol{M} = u - 2\,\Omega\,y$ *and* $\hat{\boldsymbol{y}} \cdot \boldsymbol{M} = v + 2\,\Omega\,x$ *are the two conserved components of potential momentum.*

Consider a particle with potential momentum $\boldsymbol{M}$ and move it from an arbitrary point to a reference position with $\boldsymbol{X} = 0$. Upon reaching the reference position, the horizontal velocity of the particle must equal to $\boldsymbol{M}$ in order to maintain the same potential momentum. This example motivates the name "potential momentum", since $\boldsymbol{M}$ measures the potential for relative motion contained in the particle as it moves along a geopotential.

10.2.2 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 9.9.1)

$$\boldsymbol{V} = \boldsymbol{V}_{\text{Cartesian}} + \boldsymbol{U}_{\text{solid-body}} = \boldsymbol{V}_{\text{Cartesian}} + \boldsymbol{O}mega \wedge \boldsymbol{X}. \tag{10.5}$$

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.

10.3 Inertial oscillations

In Section 24.2 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. The $f-$plane approximation furthermore sets the Coriolis parameter

$$f = 2\,\Omega \sin\phi_0 \tag{10.6}$$

to a constant, where ϕ_0 is a chosen latitude. Consequently, a particle moving on the f-plane maintains a constant horizontal potential momentum

$$\frac{\mathrm{d}M_x}{\mathrm{d}t} = \frac{\mathrm{d}(u - f\,y)}{\mathrm{d}t} = 0 \tag{10.7a}$$

$$\frac{\mathrm{d}M_y}{\mathrm{d}t} = \frac{\mathrm{d}(v + f\,x)}{\mathrm{d}t} = 0, \tag{10.7b}$$

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.

10.3.1 Oscillator equation

Taking the time derivative of the zonal equation (10.7a) and using the meridional equation (10.7b) leads to

$$\ddot{u} - f\,\dot{v} = \ddot{u} + f^2\,u = 0. \tag{10.8}$$

Similar manipulations for the meridional velocity equation render the free oscillator equation for each component of the horizontal velocity

$$\frac{\mathrm{d}^2\boldsymbol{U}}{\mathrm{d}t^2} + f^2\,\boldsymbol{U} = 0. \tag{10.9}$$

Motions that satisfy this equation are termed *inertial oscillations*.

10.3.2 Particle trajectory and velocity

Time integrating the equation of motion (10.9) renders the particle trajectory and its velocity

$$\boldsymbol{X}(t) = \frac{U}{f}\Big[\hat{\boldsymbol{x}}\sin(ft) + \hat{\boldsymbol{y}}\cos(ft)\Big] \tag{10.10a}$$

$$\boldsymbol{U}(t) = U\Big[\hat{\boldsymbol{x}}\cos(ft) - \hat{\boldsymbol{y}}\sin(ft)\Big], \tag{10.10b}$$

where U is the particle speed, which is a constant, and we assumed the initial conditions

$$\boldsymbol{X}(0) = \frac{U}{f}\,\hat{\boldsymbol{y}} \qquad \text{and} \qquad \boldsymbol{U}(0) = U\,\hat{\boldsymbol{x}}. \tag{10.11}$$

From the particle trajectory equation (10.10a), we see that motion of a particle exhibiting inertial oscillations is circular and with radius

$$R = |U|\,f^{-1}. \tag{10.12}$$

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 discussed in Section 27.3.2, inertial oscillations arise from a balance between the Coriolis acceleration of the rotating frame and the centrifugal 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

$$\boldsymbol{M}(t) = \boldsymbol{U}(t) + f\,\hat{\boldsymbol{z}} \wedge \boldsymbol{X}(t) = 0. \tag{10.13}$$

Adding an arbitrary constant to the initial position makes the potential momentum equal to a nonzero constant.

10.3.3 Period of inertial oscillations

Inertial oscillations possess a constant speed and move around the inertial circle with a period

$$T_{\text{inertial}} = \frac{2\,\pi}{f} = \frac{11.97}{|\sin\phi_0|}\,\text{hour}, \tag{10.14}$$

where we set $\Omega = 7.292 \times 10^{-5}\text{s}^{-1}$ (equation (9.18)). 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 $T_{\text{smallest}} \approx 12\,\text{hour}$. 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.

10.3.4 Comments and further study

Inertial oscillations of fluid parcels 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 also 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, thus allowing us to drop the pressure gradient from the momentum equation.

We encounter inertial motions again in Section 27.3 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 to refer to motion in an inertial reference frame (Section 9.3). 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 offers a useful controlled setting to observe inertial ocillations, such as shown near the 18 minute mark in this video from Prof. Fultz of the University of Chicago.

[1] This centrifugal acceleration is distinct from the planetary centrifugal acceleration absorbed into the geopotential (Section 9.12.2). The planetary centrifugal acceleration plays no role in inertial oscillations.

10.4 Dynamical constraints arising from geometric 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.

10.4.1 Linear momentum conservation

As seen in Section 9.1.1, linear momentum remains constant for a particle moving without any forces acting on it; i.e., a *free particle*. The conservation of linear momentum is most readily viewed within the particle's inertial reference frame, where a vanishing inertial acceleration leads to a constant inertial velocity

$$\boldsymbol{A} = 0. \tag{10.15}$$

When viewed from a rotating frame using Cartesian coordinates, a vanishing inertial acceleration means that the Cartesian acceleration balances Coriolis and centrifugal accelerations

$$\ddot{\boldsymbol{X}} = -2\,\boldsymbol{\Omega} \wedge \dot{\boldsymbol{X}} - \nabla\Phi_{\text{centrifugal}}. \tag{10.16}$$

This equation can be written in terms of the potential momentum

$$\frac{\mathrm{d}\boldsymbol{M}}{\mathrm{d}t} = -\nabla\Phi_{\text{centrifugal}}. \tag{10.17}$$

Hence, if $\Phi_{\text{centrifugal}}$ is static, then motion maintaining constant inertial momentum also maintains $\hat{\boldsymbol{s}} \cdot \boldsymbol{M} = 0$ constant, where $\hat{\boldsymbol{s}}$ is a vector tangent to constant $\Phi_{\text{centrifugal}}$ surfaces.

10.4.2 Potential momentum conservation

Again, the 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 (10.4))

$$\frac{\mathrm{d}(\hat{\boldsymbol{s}} \cdot \boldsymbol{M})}{\mathrm{d}t} = 0. \tag{10.18}$$

A geopotential is a two-dimensional surface so that this conservation law corresponds to two dynamical constraints such as shown in Figure 10.1.

10.4.3 Angular momentum conservation

As detailed in Section 10.6, the angular momentum computed with respect to the axis of rotation is a constant of the motion (Figure 10.2). This conservation law arises from rotational symmetry of the system about the rotational axis. Axial angular momentum conservation takes the form

$$\frac{\mathrm{d}L^z}{\mathrm{d}t} = 0, \tag{10.19}$$

where the axial angular momentum is

$$L^z = m\,r_\perp^2\,(\dot{\lambda} + \Omega) \qquad \text{with} \qquad r_\perp = \sqrt{x^2 + y^2}, \tag{10.20}$$

$r_\perp$ is the distance from the rotation axis; i.e., it is the *moment arm*, λ is the longitude that measures the angle in the counter-clockwise direction from the positive x-axis, and $\dot{\lambda}$ is the time change of the longitude.

Figure 10.2: *Axial angular momentum, $L^z = m\, r_\perp^2\, (\dot{\lambda} + \Omega)$, is the moment of the zonal momentum around the sphere, with $r_\perp = \sqrt{x^2 + y^2}$ 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.*

10.4.4 Kinetic energy and Galilean transformation

As we saw in Section 9.1.5, changes in the kinetic energy of a particle equal to the *mechanical work* done on the particle as it moves along its trajectory

$$K(t_2) - K(t_1) = \int_{\boldsymbol{x}(t_1)}^{\boldsymbol{x}(t_2)} \boldsymbol{F} \cdot \mathrm{d}\boldsymbol{x} = \int_{t_1}^{t_2} \boldsymbol{F} \cdot \boldsymbol{V}\, \mathrm{d}t, \tag{10.21}$$

where $\boldsymbol{V}\, \mathrm{d}t = \mathrm{d}\boldsymbol{x}$ defines the vector increment along the trajectory. The integrand, $\boldsymbol{F} \cdot \boldsymbol{V}$, is the *power*, so that the time integral of the power equals to the change in kinetic energy.

The kinetic energy is *not* Galilean invariant since movement to another inertial reference frame leads to the kinetic energy change

$$\overline{\boldsymbol{V}} = \boldsymbol{V} + \boldsymbol{U} \Longrightarrow \overline{K} = K + \frac{m}{2}\, (2\, \boldsymbol{V} \cdot \boldsymbol{U} + \boldsymbol{U} \cdot \boldsymbol{U}), \tag{10.22}$$

where $\boldsymbol{U}$ is a constant boost velocity so that $\mathrm{d}\boldsymbol{U}/\mathrm{d}t = 0$. We do not expect kinetic energy to remain Galilean invariant since it is energy measured by the motion relative to the chosen reference frame. Even so, the time change of the kinetic energy in the new inertial frame is given by the power in the new frame

$$\frac{\mathrm{d}\overline{K}}{\mathrm{d}t} = \frac{\mathrm{d}K}{\mathrm{d}t} + m\, \boldsymbol{A} \cdot \boldsymbol{U} = \boldsymbol{F} \cdot \boldsymbol{V} + \boldsymbol{F} \cdot \boldsymbol{U} = \boldsymbol{F} \cdot \overline{\boldsymbol{V}}. \tag{10.23}$$

10.4.5 Cartesian expression in a rotating reference frame

We now consider the expression for the kinetic energy when introducing the velocity of the rotating reference frame. Writing the inertial velocity in the planetary Cartesian form

$$\boldsymbol{V} = \boldsymbol{V}_{\text{Cartesian}} + \boldsymbol{U}_{\text{solid}}, \tag{10.24}$$

leads to the kinetic energy

$$K = \frac{m}{2}\left[\boldsymbol{V}_{\text{Cartesian}} \cdot \boldsymbol{V}_{\text{Cartesian}} + 2\,\boldsymbol{V}_{\text{Cartesian}} \cdot \boldsymbol{U}_{\text{solid}} + \boldsymbol{U}_{\text{solid}} \cdot \boldsymbol{U}_{\text{solid}}\right] \tag{10.25}$$

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.

10.4.6 Kinetic energy using spherical coordinates: 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 7.1. Doing so for the solid body velocity leads to equation (9.56)

$$\boldsymbol{U}_{\text{solid}} = \Omega\, r \cos\phi\, \left(-\sin\lambda\, \hat{\boldsymbol{x}} + \cos\lambda\, \hat{\boldsymbol{y}}\right). \tag{10.26}$$

Likewise, the velocity components measured in the rotating frame are given by

$$\dot{X} = \frac{\mathrm{d}\left(r\cos\phi\cos\lambda\right)}{\mathrm{d}t} = \dot{r}\,\cos\phi\,\cos\lambda - r\dot{\phi}\,\sin\phi\,\cos\lambda - r\dot{\lambda}\,\cos\phi\,\sin\lambda \tag{10.27a}$$

$$\dot{Y} = \frac{\mathrm{d}\left(r\cos\phi\sin\lambda\right)}{\mathrm{d}t} = \dot{r}\,\cos\phi\,\sin\lambda - r\dot{\phi}\,\sin\phi\,\sin\lambda + r\dot{\lambda}\,\cos\phi\,\cos\lambda \tag{10.27b}$$

$$\dot{Z} = \frac{\mathrm{d}\left(r\sin\phi\right)}{\mathrm{d}t} = \dot{r}\,\sin\phi + r\dot{\phi}\,\cos\phi. \tag{10.27c}$$

Bringing terms together then leads to the kinetic energy in terms of spherical coordinates

$$K = \frac{m}{2}\left[\left(\dot{r}^2 + r^2\dot{\phi}^2 + \dot{\lambda}^2\, r^2\cos^2\phi\right) + \left(2\,\Omega\, r^2\dot{\lambda}\,\cos^2\phi\right) + \left(\Omega\, r\,\cos\phi\right)^2\right]. \tag{10.28}$$

10.4.7 Kinetic energy using spherical coordinates: Part II

An alternative means for deriving the kinetic energy in equation (10.28) makes use of the spherical coordinate form of the inertial velocity given by equation (9.52f), in which case

$$\boldsymbol{V} = \left(u + r_\perp\,\Omega\right)\hat{\boldsymbol{\lambda}} + v\,\hat{\boldsymbol{\phi}} + w\,\hat{\boldsymbol{r}}, \tag{10.29}$$

so that

$$K = \frac{m}{2}\left[\left(u + r_\perp\,\Omega\right)^2 + v^2 + w^2\right], \tag{10.30}$$

where $r_\perp = r\cos\phi$. Additionally, as discussed in Section 10.6, the axial angular momentum is given by

$$L^z = m\left(u + r_\perp\,\Omega\right), \tag{10.31}$$

and this property is a constant of the motion when there is azimuthal (zonal) symmetry. It is thus convenient in some contexts to write the kinetic energy can be written

$$K = \frac{\left(L^z\right)^2}{2\,m} + \frac{m}{2}\left(v^2 + w^2\right). \tag{10.32}$$

10.5 Mechanical energy conservation

Does the particle know anything about the time origin? 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 by manipulating the equations of motion. Many of the manipulations also occur when considering the mechanical conservation laws for a continuum fluid discussed in Chapter 18.

10.5.1 General expression for kinetic energy evolution

The time derivative of the kinetic energy is given by

$$\frac{\mathrm{d}K}{\mathrm{d}t} = m\,\boldsymbol{V}\cdot\frac{\mathrm{d}\boldsymbol{V}}{\mathrm{d}t} = m\,\boldsymbol{V}\cdot\boldsymbol{A} = -m\,\boldsymbol{V}\cdot\nabla\Phi_{\mathrm{e}}. \tag{10.33}$$

For the final equality we introduced the gravitational potential given that the particle only feels an external force from gravity as per equation (9.114).

10.5.2 Conservation of mechanical energy

The gravitational potential is given by (see equation (9.109))

$$\Phi_{\mathrm{e}} = g_{\mathrm{e}}\,r, \tag{10.34}$$

so that

$$\frac{\mathrm{d}K}{\mathrm{d}t} = -m\,\boldsymbol{V}\cdot\nabla\Phi_{\mathrm{e}} = -m\,g_{\mathrm{e}}\,\dot{r}. \tag{10.35}$$

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 (9.110))

$$\frac{\mathrm{d}P_{\mathrm{e}}}{\mathrm{d}t} = m\,g_{\mathrm{e}}\,\dot{r}, \tag{10.36}$$

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

$$\frac{\mathrm{d}(K+P_{\mathrm{e}})}{\mathrm{d}t} = 0, \tag{10.37}$$

where the mechanical energy is the sum of the inertial kinetic energy plus the gravitational potential energy

$$M = K + P_{\mathrm{e}} \tag{10.38a}$$

$$= \frac{m}{2}\,\boldsymbol{V}\cdot\boldsymbol{V} + m\,\Phi_{\mathrm{e}} \tag{10.38b}$$

$$= \frac{m}{2}\left[(u + r_{\perp}\,\Omega)^2 + v^2 + w^2\right] + m\,g_{\mathrm{e}}\,r \tag{10.38c}$$

$$= \frac{(L^z)^2}{2\,m} + \frac{m}{2}\left(v^2 + w^2\right) + m\,g_{\mathrm{e}}\,r. \tag{10.38d}$$

10.5.3 Comments and further study

Rotation plays no role in the evolution of either the kinetic energy or gravitational potential energy. That is, there are no appearance of the Coriolis and planetary centrifugal accelerations in the mechanical energy evolution equations. This result is not surprising since the Coriolis and planetary centrifugal accelerations arise through our subjective choice to view the motion from a rotating reference frame, and that choice plays no role in the evolution of total mechanical energy.

10.6 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 manipulations also occur when considering angular conservation for a continuous fluid (see Section 18.4).

10.6.1 Angular momentum

The angular momentum of the particle, computed with respect to the origin of the sphere, is given by

$$\boldsymbol{L} = m\,\boldsymbol{X} \wedge \boldsymbol{V}. \tag{10.39}$$

This is the moment of the linear momentum computed relative to the origin. We write the angular momentum computed along the polar axis as

$$\begin{aligned} L^z &= \boldsymbol{L} \cdot \hat{\boldsymbol{z}} && \text{(10.40a)}\\ &= m\,(\boldsymbol{X} \wedge \boldsymbol{V}) \cdot \hat{\boldsymbol{z}} && \text{(10.40b)}\\ &= m\,(\hat{\boldsymbol{z}} \wedge \boldsymbol{X}) \cdot \boldsymbol{V} && \text{(10.40c)}\\ &= m\,r\cos\phi\,(\hat{\boldsymbol{\lambda}} \cdot \boldsymbol{V}) && \text{(10.40d)}\\ &= m\,r_\perp\,(\hat{\boldsymbol{\lambda}} \cdot \boldsymbol{V}). && \text{(10.40e)} \end{aligned}$$

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)

$$r_\perp = r\,\cos\phi. \tag{10.41}$$

In deriving equation (10.40e), we made use of the identity

$$\hat{\boldsymbol{z}} \wedge \boldsymbol{X} = r_\perp\,\hat{\boldsymbol{\lambda}}, \tag{10.42}$$

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 (10.40e) in terms of the rotating frame quantities. To do so, introduce the inertial velocity written using spherical coordinates according to equation (9.52f), which yields

$$L^z = m\,r_\perp\,(\hat{\boldsymbol{\lambda}} \cdot \boldsymbol{V}) = m\,r_\perp^2\,(\dot{\lambda} + \Omega) = m\,r_\perp\,(u + r_\perp\,\Omega). \tag{10.43}$$

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.

10.6.2 Conservation of axial angular momentum

The time derivative of the axial angular momentum is given by

$$\frac{\mathrm{d}L^z}{\mathrm{d}t} = m\,\frac{\mathrm{d}}{\mathrm{d}t}\left[(\hat{\boldsymbol{z}}\wedge\boldsymbol{X})\cdot\boldsymbol{V}\right] \tag{10.44a}$$

$$= m\,(\hat{\boldsymbol{z}}\wedge\boldsymbol{V})\cdot\boldsymbol{V} + m\,(\hat{\boldsymbol{z}}\wedge\boldsymbol{X})\cdot\boldsymbol{A} \tag{10.44b}$$

$$= m\,(\hat{\boldsymbol{z}}\wedge\boldsymbol{X})\cdot\boldsymbol{A} \tag{10.44c}$$

$$= m\,r_\perp\,\hat{\boldsymbol{\lambda}}\cdot\boldsymbol{A}. \tag{10.44d}$$

To reach this result we noted that the polar axis direction, $\hat{\boldsymbol{z}}$, is time independent, and we used the identity (10.42) for the final step. The inertial acceleration arises just from the central-force gravitational field (equation (9.114))

$$\boldsymbol{A} = -\nabla\Phi_{\mathrm{e}} = -g_{\mathrm{e}}\,\hat{\boldsymbol{r}}. \tag{10.45}$$

Since $\hat{\boldsymbol{\lambda}}\cdot\hat{\boldsymbol{r}} = 0$, we have axial angular momentum conservation

$$\frac{\mathrm{d}L^z}{\mathrm{d}t} = 0. \tag{10.46}$$

10.7 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 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 10.3. 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

$$L^z = m\,l^z = m\,r_\perp^2\,(\dot{\lambda}+\Omega) = m\,r_\perp\,(u + r_\perp\,\Omega), \tag{10.47}$$

where

$$l^z = L^z/m \qquad \text{and} \qquad u = r_\perp\,\dot{\lambda} \tag{10.48}$$

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:

$$r_\perp\,\Omega > |u| \qquad \text{for most terrestrial motions,} \tag{10.49}$$

with $r_\perp\,\Omega = 465$ m s^{-1} at the equator.

10.7.1 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 $\delta l^z = 0$, which leads to

$$\delta\dot{\lambda} = -\frac{2\,l^z}{r_\perp^2}\,\frac{\delta r_\perp}{r_\perp} \Longleftrightarrow \delta u = -\Omega\,\delta r_\perp\left[1 + \frac{l^z}{\Omega\,r_\perp^2}\right], \tag{10.50}$$

where we set $\delta\Omega = 0$ since the earth's rotation rate is assumed to be fixed. Furthermore, we found it convenient to write the changes in terms of l^z since it is a constant of the motion. 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 acceleration ($\delta\dot{\lambda} > 0$ 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

$$\delta r_\perp = (\cos\phi)\,\delta r - (r\sin\phi)\,\delta\phi. \tag{10.51}$$

Assuming these changes occur over a small time increment δt allows us to write

$$\delta r_\perp/\delta t = w\cos\phi - v\sin\phi, \tag{10.52}$$

where

$$v = r\,\delta\phi/\delta t = r\,\dot{\phi} \qquad \text{and} \qquad w = \delta r/\delta t = \dot{r} \tag{10.53}$$

introduced the meridional and vertical components to the particle velocity according to equation (9.54). For geophysical fluid motion, changes in vertical distance δr are far smaller than the distance to the earth's center

$$\delta r \ll r. \tag{10.54}$$

Consequently, when $\phi \neq 0$, $\delta r_\perp$ is affected much more by meridional motion at constant radial position (second term on right hand side of equation (10.52)) than by vertical motion at constant latitude (first term). We return to this observation in Section 10.7.6 when discussing nearly horizontal motions, which are the basis for the shallow fluid approximation used to develop the primitive equations in Section 24.1.2.

10.7.2 Axial angular momentum conservation yields zonal Coriolis acceleration

Consider the zonal acceleration induced for a particle at rest whose angular momentum per mass equals $l^z = r_\perp^2\,\Omega$. In this case the zonal acceleration in equation (10.50) is

$$\delta\dot{\lambda} = -2\,\Omega\,\frac{\delta r_\perp}{r_\perp} \Longleftrightarrow \dot{u} = -2\,\Omega\,\delta r_\perp = -2\,\Omega\,(w\cos\phi - v\sin\phi). \tag{10.55}$$

The expression for $\dot{u}$ is precisely the same as the Coriolis acceleration appearing in the zonal momentum equation (9.130)

$$\dot{u} + \frac{u\,(w - v\tan\phi)}{r} = \underbrace{-2\,\Omega\,(w\cos\phi - v\sin\phi)}_{\text{Coriolis acceleration}}. \tag{10.56}$$

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 10.7.5 by deriving the zonal momentum equation from the axial angular momentum equation.

Figure 10.3: *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.*

10.7.3 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 (10.55) induces a zonal acceleration

$$\dot{u} = 2\,\Omega\, v \sin\phi, \tag{10.57}$$

which, as seen by equation (10.56), 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 10.3.

10.7.4 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 (10.55) induces a zonal acceleration

$$\dot{u} = -2\,\Omega\, w\, \cos\phi, \tag{10.58}$$

which, as seen by equation (10.56), 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.

10.7.5 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 (10.47), in which case

$$\frac{1}{m}\frac{\mathrm{d}L^z}{\mathrm{d}t} = \frac{\mathrm{d}[(r\cos\phi)^2\,(\dot{\lambda}+\Omega)]}{\mathrm{d}t} \tag{10.59a}$$

$$= 2\,(\dot{r}\cos\phi - r\,\dot{\phi}\sin\phi)\,(\dot{\lambda}\,r\cos\phi + r\,\Omega\cos\phi) + (r\cos\phi)^2\,\ddot{\lambda} \tag{10.59b}$$

$$= 2\,(w\cos\phi - v\sin\phi)\,(u + r\,\Omega\cos\phi) + (r\cos\phi)^2\,\ddot{\lambda}, \tag{10.59c}$$

where we introduced the (u, v, w) velocity components according to equation (9.54). With the zonal velocity $u = \dot{\lambda}\,r\cos\phi$, we have

$$r\cos\phi\,\ddot{\lambda} = \dot{u} + \frac{u}{r\cos\phi}\,(v\sin\phi - w\cos\phi)\,, \tag{10.60}$$

so that equation (10.59c) thus takes the form

$$\frac{1}{m}\frac{\mathrm{d}L^z}{\mathrm{d}t} = 2\,(w\cos\phi - v\sin\phi)\,(u + r\,\Omega\cos\phi) + (r\cos\phi)^2\,\ddot{\lambda} \tag{10.61a}$$

$$= (w\cos\phi - v\sin\phi)\,(u + 2\,r\,\Omega\cos\phi) + \dot{u}\,r\cos\phi. \tag{10.61b}$$

Setting $\mathrm{d}L^z/\mathrm{d}t = 0$ and rearranging then leads to a prognostic equation for the zonal velocity

$$\frac{\mathrm{d}u}{\mathrm{d}t} = \left[\frac{u}{r\cos\phi} + 2\,\Omega\right](v\sin\phi - w\cos\phi). \tag{10.62}$$

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 (10.47), in which case

$$\frac{1}{m}\frac{\mathrm{d}L^z}{\mathrm{d}t} = \frac{\mathrm{d}\,[u\,r\,\cos\phi + \Omega\,(r\,\cos\phi)^2]}{\mathrm{d}t} \tag{10.63a}$$

$$= \dot{u}\,r\,\cos\phi + u\,\dot{r}\,\cos\phi - u\,r\,\dot{\phi}\,\sin\phi + 2\,\Omega\,r\,\cos\phi\,(\dot{r}\,\cos\phi - r\,\dot{\phi}\,\sin\phi). \tag{10.63b}$$

Again, setting $\mathrm{d}L^z/\mathrm{d}t = 0$ and rearranging leads to the zonal velocity equation (10.62).

10.7.6 Nearly horizontal motions

We here consider two approximations relevant to large scale geophysical fluid motions.

1. The particle kinetic energy is dominated by horizontal motions on the sphere (i.e., motion at constant radial position).

2. Vertical excursions are much smaller than the earth radius.

When applied to a fluid, the first assumption leads to the hydrostatic approximation (Section 24.3), and the second assumption leads to the shallow fluid approximation (Section 24.1.2). 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

$$K \approx \frac{m}{2}\,[(u + r_\perp\,\Omega)^2 + v^2]. \tag{10.64}$$

The second assumption means that the axial angular momentum takes the approximate form

$$L^z \approx m\,R_\perp\,(u + \Omega\,R_\perp) = m\,R_\perp^2\,(\dot\lambda + \Omega), \tag{10.65}$$

where

$$r = R + z \approx R \qquad \text{and} \qquad R_\perp = R\cos\phi. \tag{10.66}$$

The approximate angular momentum (10.65) ignores contributions from vertical motion in changing the moment-arm. Indeed, as noted in Section 10.7.1, 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 (10.62) is modified to the form

$$\frac{\mathrm{d}u}{\mathrm{d}t} = v\left[\frac{u\tan\phi}{R} + f\right] \qquad \text{where} \qquad f = 2\,\Omega\,\sin\phi. \tag{10.67}$$

That is, we dropped the vertical velocity component, w, from the general form of the acceleration (10.62). Correspondingly, the meridional momentum equation takes the form

$$\frac{\mathrm{d}v}{\mathrm{d}t} = -u\left[\frac{u\tan\phi}{R} + f\right]. \tag{10.68}$$

These approximate forms for the zonal and meridional accelerations appear in the primitive equations of geophysical fluid dynamics presented in Section 24.1.

10.8 More facets of motion on the rotating sphere

In this section we expose a few more points to keep in mind regarding motion on a rotating sphere. These points build from the material in Section 10.7, but here also examine the meridional and vertical equations of motion, (9.131) and (9.132).

10.8.1 Constant zonal motion is not steady motion

Consider a particle moving zonally with $u \neq 0$. In the absence of any forces, constant zonal particle motion can presumably continue freely without acceleration, $\dot u = 0$ (see equation (10.56)). Indeed, this is the case for motion on a non-rotating plane. However, there are flaws to this argument for motion on a rotating sphere since $u =$ constant with $\dot u = 0$ is not a steady solution to the equations of motion. This point is also seen from our earlier discussions in Chapter 9 of particle kinematics on a rotating sphere.

The reason that $u =$ constant with $\dot u = 0$ is not a steady solution is that zonal motion leads to nonzero meridional and vertical Coriolis accelerations as well as metric accelerations. We see these accelerations by examining the meridional and vertical velocity equations (9.131) and (9.132)

$$\dot v + \frac{v\,w + u^2\tan\phi}{r} + 2\,\Omega\,u\sin\phi = 0 \tag{10.69}$$

$$\dot w - \frac{u^2 + v^2}{r} - 2\,\Omega\,u\cos\phi = -g. \tag{10.70}$$

So the presence of $u \neq 0$, even with $v = w = 0$, introduces accelerations to the meridional and vertical velocity equations. These meridional and vertical velocities in turn induce a nonzero zonal acceleration making $\dot{u} \neq 0$ and thus leading to a non-constant zonal velocity.

The two underlying reasons for these accelerations are (i) rotation and (ii) curvature of the sphere. Namely, a constant zonal velocity on the sphere, $\boldsymbol{v} = u\,\hat{\boldsymbol{\lambda}}$, is not a constant inertial or absolute velocity. In this case, the inertial velocity (velocity seen by an inertial reference frame fixed relative to the stars) is instead given by

$$\boldsymbol{V} = \boldsymbol{v} + \boldsymbol{\Omega} \wedge \boldsymbol{X} = u\,\hat{\boldsymbol{\lambda}} + \boldsymbol{\Omega} \wedge \boldsymbol{X}, \tag{10.71}$$

whose time derivative is nonzero

$$\frac{\mathrm{d}\boldsymbol{V}}{\mathrm{d}t} = \dot{u}\,\hat{\boldsymbol{\lambda}} + u\,\frac{\mathrm{d}\hat{\boldsymbol{\lambda}}}{\mathrm{d}t} + \boldsymbol{\Omega} \wedge \boldsymbol{V} = \dot{u}\,\hat{\boldsymbol{\lambda}} + u\,\frac{\mathrm{d}\hat{\boldsymbol{\lambda}}}{\mathrm{d}t} + \boldsymbol{\Omega} \wedge (u\,\hat{\boldsymbol{\lambda}} + \boldsymbol{\Omega} \wedge \boldsymbol{X}). \tag{10.72}$$

So even if $\dot{u}$ is initially zero, the zonal particle motion is associated with nonzero accelerations due to changes in the zonal unit vector ("metric acceleration"), as well as rotation (Coriolis and planetary centrifugal accelerations).

10.8.2 Coriolis acceleration in the meridional equation

What happens when zonal motion is perturbed in a meridional direction? Following the angular momentum constraint (10.55), 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 (9.131)

$$\dot{v} + \frac{v\,w + u^2 \tan\phi}{r} = \underline{-2\,\Omega\,u \sin\phi}, \tag{10.73}$$

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, $-2\,\Omega\,u \sin\phi < 0$, 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 10.3.

10.8.3 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 10.3 depicts some sample deflections.

10.9 Exercise

EXERCISE 10.1: WESTWARD MOTION FASTER THAN PLANETARY MOTION

Reconsider the discussion in Section 10.7.3, only now allow for the particle to move zonally westward

at a speed greater than the planetary rotation so that

$$\dot{\lambda} + \Omega < 0 \Longleftrightarrow u + \Omega\, r_\perp < 0. \tag{10.74}$$

(A) Discuss what happens to the particle as it is deflected poleward while conserving axial angular momentum.

(B) 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 $\Omega\, r_\perp$ for $\phi = \pi/4$ and $r = R_{\rm e}$? Note, we already provided the result for the equator just after equation (10.49).

Part III

Fluid kinematics

In this part of the book we develop the kinematics of classical non-relativistic fluid mechanics. Kinematics is concerned with the intrinsic properties of motion, including properties of the space-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.

The first chapter in this part of the book briefly argues for the relevance of a continuum approach to fluid mechanics (Chapter 11). A continuum approach is fundamental to nearly all areas of fluid mechanics, with this book sitting firmly within that realm.

Quite conveniently, the kinematic properties of non-rotating fluids also hold for the fluids on a uniformly rotating planet as considered in this book. The reason is that uniform rotation does not directly impart any strain to the fluid, where strain refers to the relative motion between fluid elements (Chapter 12). Uniform rotation does impart a planetary component to the vorticity of geophysical fluids, with important 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 motion. The Eulerian frame describes fluid motion relative to a frame fixed in the laboratory whereas the Lagrangian frame follows a fluid particle. The Eulerian frame is inertial (when the laboratory is not rotating), whereas the Lagrangian is non-inertial since fluid particles accelerate. Having two descriptions of the same motion provides a synergy that is missing with either alone, thus exemplifying a maxim: *if you can solve a problem more than one way, then do so!* It is thus very useful to have skills at moving between the two descriptions, with tools from mathematical transformation theory of Part I of great use. Developing the methods of Eulerian and Lagrangian kinematics forms the focus for Chapter 12, and Chapter 13 explores the Lagrangian kinematics of material lines, areas, and volumes.

mass conservation is part of fluid kinematics

The conservation of mass plays a central role in physics. For fluids, mass conservation constrains the fluid flow regardless what forces act on the fluid. Hence, mass conservation is traditionally included as part of fluid kinematics rather than fluid dynamics. Mass conservation, and its expression as volume conservation for incompressible flows, are the subjects of Chapters 14, 15 and 16. Furthermore, Chapter 17 explores the kinematics of fluid motion when described with the generalized vertical coordinates introduced in Chapter 8.

fluid kinematics + fluid dynamics = fluid mechanics

In a broad sense, kinematics deduces the acceleration whereas dynamics deduces the forces, with Newton linking the two into the equation of motion: $\boldsymbol{F} = m\,\boldsymbol{a}$. Furthermore, as discussed in Chapter 10, 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. 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.

11

The continuum hypothesis

Viewed macroscopically, a fluid deforms continuously when applying a force so that a fluid has no preferred shape. Correspondingly, a fluid cannot sustain a shearing stress in the absence of motion. 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 hypothesis*, which assumes that mathematical limits for fluid volumes tending to zero are reached on length and time scales very large compared to molecular scales. The temporal version of the continuum hypothesis 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.

Operationally, the continuum hypothesis allows us to make use of differential calculus for describing the mechanics of fluid motion. That is, the continuum hypothesis makes fluid mechanics a continuous space-time field theory. Correspondingly, the conservation laws describing fluid motion are partial differential equations.

READER'S GUIDE TO THIS CHAPTER

This chapter presents salient points concerning a continuum description of fluid mechanics. Our goal is to physically and consisely unpack the dictum "macroscopically small yet microscopically large", which summarizes the regime considered by the continuum and quasi-static hypotheses. Every chapter makes use of these two hypotheses thus warranting their presentation at the start of this book. For this presentation 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 geophysical fluid mechanics. No prior exposure to these treatments is necessary, nor do we dive into the details.

11.1 A variety of length scales **182**
11.1.1 Molecular and macroscopic length scales 182
11.1.2 Continuous fields rather than discrete molecules 182
11.1.3 Reynolds number and the continuum length scale 184
11.1.4 Resolution of measurements and simulations 185
11.1.5 Comments 185
11.2 Results from kinetic theory **186**
11.2.1 A mole and Avogadro's number 186
11.2.2 Ideal gas law 186
11.2.3 Molecular mean free path 187
11.2.4 Root mean square molecular speed 188
11.2.5 Basis for the quasi-static approximation 188
11.2.6 Macroscopically small and microscopically large 189
11.2.7 Whence a rigorous treatment? 189
11.2.8 Further study 189

11.1 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 in the environment and laboratory.

We here outline the essential features of the continuum hypothesis. This hypothesis supports our macroscopic description of a fluid in terms of continuous fields rather than discrete molecules. More details are offered in Section 11.2, although a full discussion is outside the subject of fluid mechanics, instead resting deep within the field of statistical physics.

11.1.1 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, whereas that same volume in water contains roughly 10^{19} water molecules. These numbers (justified in Section 11.2) 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 11.1 offers a schematic to illustrate these quite distinct length scales.

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 hypothesis* 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.

11.1.2 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

Figure 11.1: *Schematic to illustrate the three length scales considered when making the continuum hypothesis. The blue circles represent molecules with diameter $L_{molecule}$. On average, molecules are separated by the mean free path, $L_{mfp} \approx 1000\, L_{molecule}$. The smallest macroscopic length scale of interest for fluid mechanics is $L_{macro} \sim 10^{-3}$ m, which is roughly $L_{macro} = 10^4\, L_{mfp}$ 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 in water 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!*

molecule (see Section 11.2.4). 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 11.2.

The ratio of the mean free path to the macroscopic length scale is known as the Knudsen number

$$Kn = \frac{L_{\text{mfp}}}{L_{\text{macro}}}. \tag{11.1}$$

Large mean free paths occur for certain rarefied gases such as in the outer regions of the earth's atmosphere. Under these conditions, there are very few molecular collisions due to the tiny number density of molecules, thus supporting relatively large mean free paths. For our purposes, we are concerned only with fluid conditions where the mean free path is microscopic so that the Knudsen number is tiny

$$Kn \ll 1. \tag{11.2}$$

For tiny Knudsen numbers, we are led to make use of the continuum hypothesis.

When making the continuum hypothesis, we employ fluid properties that take values at each point within a space and time continuum, $(\boldsymbol{x}, t)$. For example, we make use of the mass density,

Figure 11.2: *The measurement of mass density of a fluid becomes erratic for volumes smaller than that determined by the molecular mean free path, $\delta V_{mfp} \sim L_{mfp}^3$. For fluid mechanics we are concerned with macroscopic length scales much larger than the mean free path, $L_{macro} \gg L_{mfp}$, in which case the density is a smooth function of space and time.*

$\rho(\boldsymbol{x}, t)$, fluid velocity, $\boldsymbol{v}(\boldsymbol{x}, t)$, pressure $p(\boldsymbol{x}, t)$, temperature $T(\boldsymbol{x}, t)$, tracer concentration, $C(\boldsymbol{x}, t)$, and other thermodynamic fields. These continuous fluid properties, or fields, formally represent the mean of molecular properties estimated over a linear dimension L_{macro} that is large microscopically but small macroscopically

$$L_{\text{molecule}} \ll L \ll L_{\text{macro}}. \tag{11.3}$$

As noted in Section 11.2.6, $L_{\text{macro}} \sim 10^{-3}$ m, which is much larger than the molecular mean free path.

11.1.3 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

$$Re_{\text{macro}} = \frac{U\, L_{\text{macro}}}{\nu} \sim 1. \tag{11.4}$$

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 are able to dissipate kinetic energy and to thus put a halt to the cascade of kinetic energy moving to the smaller scales through nonlinear turbulent processes. Without viscous effects, the turbulent cascade would lead to an ultraviolet energy catastrophe in which mechanical energy unboundedly accumulates

at the small scales. We are thus movitated 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))

$$\nu_{\text{air}} = \frac{1.7 \times 10^{-5} \text{ kg m}^{-1} \text{ s}^{-1}}{1.3 \text{ kg m}^{-3}} = 1.3 \times 10^{-5} \text{ m}^2 \text{ s}^{-1}, \tag{11.5}$$

and a typical fluid velocity fluctuation has a scale 10^{-1} m s^{-1}, so that

$$L_{\text{macro}} \approx 10^{-4} \text{ m} = 0.1 \text{ mm}. \tag{11.6}$$

Water has a kinematic viscosity (page 75 of *Gill* (1982))

$$\nu_{\text{water}} = \frac{10^{-3} \text{ kg m}^{-1} \text{ s}^{-1}}{1000 \text{ kg m}^{-3}} \approx 10^{-6} \text{ m}^2 \text{ s}^{-1}, \tag{11.7}$$

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 0.1 mm. 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 $10^{-4} - 10^{-3}$ mm.

11.1.4 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 grid cells with length scales $L_{\text{numerical}} \gg L_{\text{macro}}$. The equations describing motions at the measurement/simulation length scales involve correlations of nonlinear fluctuations occurring at the smaller (unmeasured) scales. The correlations play a role, sometimes a dominant role, in the evolution of the measured scales. The parameterization of these correlations in terms of measured/simulated motions constitutes the turbulence closure problem. We do not study turbulence closure theories 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

$$L_{\text{gradient}} = \frac{|\boldsymbol{v}|}{|\nabla \boldsymbol{v}|}, \tag{11.8}$$

where $\boldsymbol{v}$ is the velocity of a fluid element and $|\nabla \boldsymbol{v}|$ 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

$$2\,\pi\, L_{\text{measure}} \leq L_{\text{gradient}}. \tag{11.9}$$

This constraint means that to measure velocity fluctuations on a scale L_{gradient} requires a finer measurement sampling with $L_{\text{measure}} = L_{\text{gradient}}/(2\,\pi)$.

11.1.5 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.

11.2 Results from kinetic theory

If the reader is content to accept the continuum hypothesis 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 hypothesis. 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 11.2.8, we provide literature pointers for those wishing more rigor.

11.2.1 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

$$A_{\mathrm{v}} = 6.022 \times 10^{23} \text{ mole}^{-1}. \tag{11.10}$$

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.[1] The molar mass of dry air is thus

$$M_{\mathrm{air}} = 0.22 * 32 \text{ g mole}^{-1} + 0.78 * 28 \text{ g mole}^{-1} \approx 28.8 \text{ g mole}^{-1}. \tag{11.11}$$

Pure (fresh) water is comprised of two hydrogen atoms and one oxygen atom. The molar mass of pure water is thus given by

$$M_{\mathrm{water}} = 2 * 1 \text{ g mole}^{-1} + 16 \text{ g mole}^{-1} = 18 \text{ g mole}^{-1}. \tag{11.12}$$

11.2.2 Ideal gas law

The ideal gas law is given by

$$p\,V = n\,R\,T, \tag{11.13}$$

where p is the pressure, V is the volume, n is the number of moles, R is the universal gas constant, and T is the absolute or thermodynamic temperature (temperature relative to absolute zero). Measuring the temperature in Kelvin leads to the universal gas constant

$$R = 8.314 \text{ J mole}^{-1} \text{ K}^{-1} = 8.314 \text{ kg m}^2 \text{ s}^{-2} \text{ mole}^{-1} \text{ K}^{-1}, \tag{11.14}$$

where the second equality replaced the energy unit, Joule, by its MKS equivalent,

$$\text{J} = \text{kg m}^2 \text{ s}^{-2}. \tag{11.15}$$

[1] We 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.

Use of the ideal gas law (11.13) says that one mole of ideal gas at standard temperature ($T_{\text{stand}} = 0°\text{C} = 273.15$ K) and standard atmospheric pressure ($p_{\text{stand}} = 101.325 \times 10^3$ Pa) occupies the following volume

$$V = \frac{n\,R\,T_{\text{stand}}}{p_{\text{stand}}} \tag{11.16a}$$

$$= \frac{(1 \text{ mole})\,(8.314 \text{ kg m}^2 \text{ s}^{-2} \text{ mole}^{-1} \text{ K}^{-1})\,(273.15 \text{ K})}{101.325 \times 10^3 \text{ kg m}^{-1} \text{ s}^{-2}} \tag{11.16b}$$

$$\approx 2.25 \times 10^{-2} \text{ m}^3, \tag{11.16c}$$

where we introduced the MKS units for pressure (force per unit area)

$$\text{Pa} = \text{N m}^{-2} = \text{kg m}^{-1} \text{ s}^{-2}. \tag{11.17}$$

Hence, the number density (number of molecules) for a mole of ideal gas is given by

$$n_{\text{gas}} = \frac{\text{number per mole}}{\text{volume per mole}} \tag{11.18a}$$

$$= \frac{A_{\text{v}}}{V} \tag{11.18b}$$

$$= \frac{6.022 \times 10^{23}}{2.25 \times 10^{-2} \text{ m}^3} \tag{11.18c}$$

$$= 2.68 \times 10^{25} \text{ m}^{-3}. \tag{11.18d}$$

Specializing to air, we compute the mass density of air at standard temperature and pressure as

$$\rho_{\text{air}} = \frac{M_{\text{air}}}{V} = \frac{28.8 \times 10^{-3} \text{ kg}}{2.25 \times 10^{-2} \text{ m}^3} = 1.28 \text{ kg m}^{-3}, \tag{11.19}$$

where we set $M_{\text{air}} = 28.8 \times 10^{-3}$ kg according to equation (11.11). 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).

11.2.3 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

$$L_{\text{mfp}} = \frac{1}{\pi\,\sqrt{2}\,n_{\text{gas}}\,d^2} \tag{11.20}$$

where d is the diameter of the molecule. The mean diameter of air molecules is roughly

$$d_{\text{molecule air}} \approx 2 \times 10^{-10} \text{ m}. \tag{11.21}$$

The mean free path for air molecules is thus given by

$$L_{\text{mfp}} = \frac{1}{\pi\,\sqrt{2}\,n_{\text{gas}}\,d^2_{\text{molecule air}}} \tag{11.22a}$$

$$= \frac{1}{\pi\,\sqrt{2}\,(2.68 \times 10^{25} \text{ m}^{-3})\,(2 \times 10^{-10} \text{ m})^2} \tag{11.22b}$$

$$= 2 \times 10^{-7} \text{ m}. \tag{11.22c}$$

The mean free path for an air molecule is roughly 1000 times larger than the molecular diameter (e.g., Figure 11.1).

11.2.4 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

$$v_{\text{rms}} = \sqrt{\frac{3\,p}{\rho}} = \sqrt{\frac{3\,R\,T}{M}}. \tag{11.23}$$

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

$$v_{\text{rms}} = \sqrt{\frac{3\,p_{\text{stand}}}{\rho_{\text{air}}}} \tag{11.24a}$$

$$= \sqrt{\frac{3\,(101.325 \times 10^3 \text{ kg m}^{-1} \text{ s}^{-2})}{1.28 \text{ kg m}^{-3}}} \tag{11.24b}$$

$$= 487 \text{ m s}^{-1}. \tag{11.24c}$$

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.

11.2.5 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

$$t_{\text{collision}} = \frac{L_{\text{mfp}}}{v_{\text{rms}}} \tag{11.25}$$

The corresponding time for air is given by

$$t_{\text{air}} = \frac{2 \times 10^{-7} \text{ m}}{487 \text{ m s}^{-1}} = 4.1 \times 10^{-10} \text{ s}. \tag{11.26}$$

Inverting this number, we see that there are roughly $t_{\text{air}}^{-1} = 2.5 \times 10^9$ 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 22 and 23 when studying thermodynamics, with particular emphasis on the implications of the quasi-static approximation presented in Section 22.1.2.

11.2.6 Macroscopically small and microscopically large

For environmental measurements of the atmosphere and ocean, or for conventional measurements in fluid 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

$$L_{\text{macro}} = 10^{-3} \text{ m}. \tag{11.27}$$

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^3_{macro} 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^3_{macro} contains

$$N_{\text{air molecules}} = V\, n_{\text{gas}} = (10^{-3} \text{ m})^3\, (2.68 \times 10^{25} \text{ m}^{-3}) \approx 3 \times 10^{16} \text{ air molecules}. \tag{11.28}$$

To compute the number of water molecules in this same volume, we first use the water mass density of $\rho \approx 10^3$ kg m^{-3} to determine the water mass in this region

$$M_{\text{water}} = \rho_{\text{water}}\, V = (1000 \text{ kg m}^{-3})\, (10^{-9} \text{ m}^3) = 10^{-6} \text{ kg}. \tag{11.29}$$

Water has a molar mass of 0.018 kg mole^{-1}, so a volume of $(10^{-3}$ m$)^3$ contains[2]

$$N_{\text{water molecules}} = \left(\frac{10^{-6} \text{ kg}}{0.018 \text{ kg mole}^{-1}} \right) \times 6.022 \times 10^{23} \text{ molecules mole}^{-1} = 3 \times 10^{19} \text{ water molecules}. \tag{11.30}$$

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.

11.2.7 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.

11.2.8 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 hypothesis reflecting that given here, see the terse 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.* (2012). Chapter 1 of *Salmon* (1998) offers an even more thorough treatment, touching on elements from kinetic theory and details for how to coarse grain

[2] The calculation on page 9 of *Griffies* (2004) has a factor of 10^6 error.

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.

12

Fundamentals of fluid kinematics

Fluid motion is very complex. Hence, it is very useful to use 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 and seamlessly 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).

12.1 Strong and weak formulations **192**
12.2 A variety of fluid systems **193**
12.2.1 Fluid particle 193
12.2.2 Material fluid parcel 193
12.2.3 Finite sized material fluid region 194
12.2.4 Fluid element 194
12.2.5 Finite sized fluid region 194
12.2.6 Comments 194
12.3 Lagrangian and Eulerian reference frames **195**

12.4 Material and position coordinates **196**
12.4.1 Trajectories of fluid particles 196
12.4.2 Material coordinates 197
12.4.3 Transforming between material and position coordinates 197
12.4.4 Transformation matrix and its Jacobian 198
12.4.5 A discrete algorithm to compute the transformation matrix 198
12.4.6 Jacobian as the ratio of volumes 199
12.4.7 Further study 201
12.5 Lagrangian and Eulerian time derivatives **201**
12.5.1 Infinitesimal space-time increment of a function 201
12.5.2 Total time derivative of a function 201
12.5.3 Eulerian: evolution measured in the laboratory frame 202
12.5.4 Lagrangian: evolution measured in the material frame 202
12.5.5 Lagrangian time derivative formulated from the material frame . . . 202
12.5.6 Sample material time derivative operations 203
12.5.7 Summarizing some terminology for equations 205
12.6 Galilean invariance **205**
12.6.1 Galilean transformation 206
12.6.2 Transformation matrix 206
12.6.3 Transforming the differential operators 206
12.7 Invariance of the material time derivative **207**
12.7.1 Invariance based on definition of the material time derivative 207
12.7.2 Invariance for a rotating reference frame 208
12.7.3 Comments 210
12.8 Fluid flow lines **210**
12.8.1 Material pathlines from fluid particle trajectories 210
12.8.2 Fluid streamlines and streamtubes 211
12.8.3 Distinguishing streamlines from pathlines for unsteady flow 211
12.8.4 Fluid streaklines 212
12.8.5 An analytic example of flow lines 213
12.8.6 Further study 215
12.9 Exercises **215**

12.1 Strong and weak formulations

The continuum hypothesis (Chapter 11) 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. Indeed, a connection between the weak and strong formulations is afforded to smooth fluid flows by the Leibniz-Reynolds transport theorem derived in Section

Figure 12.1: *Schematic for the fluid systems used in our considerations of fluid kinematics. Left panel: fluid particle that tests the fluid flow field, $\boldsymbol{v}$, without altering it. Fluid particles have no mass and no extension. Their trajectories 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 deformable and tiny regions of the fluid that move with the fluid velocity. Right panel: fluid element, which is a generalization of the fluid parcel and is comprised of a fixed mass but with matter exchanged across its boundary. The fluid element moves with the barycentric velocity (see Section 15.1), which is the center of mass velocity. Finite extensions of the material parcel are referred to as material fluid regions, and finite extensions of fluid elements are referred to as fluid regions.*

15.3.4. 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)).

12.2 A variety of fluid systems

As part of this continuum description, we make use of conceptual systems for framing the kinematic and dynamic description of the motion. These systems are used throughout this book.

12.2.1 Fluid particle

A *fluid particle* is a non-extensive massless point that follows the fluid flow, with the flow specified by the fluid's velocity field (left panel in Figure 12.1). A fluid particle is uniquely specified by its material coordinate and time (we discuss material coordinates in Section 12.4.1).

A fluid particle does not refer to a molecule. Rather, a fluid particle is a mathematical construct afforded to us by the continuum hypothesis. The fluid particle has zero mass and zero extent, so that it does not affect the flow. Rather, the sole purpose of a fluid particle is to sample the fluid flow at an arbitrary point in the fluid continuum, and fluid particle trajectories define the Lagrangian reference frame (Section 12.3). A fluid particle is thus directly analogous to *test particles* in Newtonian gravitation that are used to map gravitational field lines, and test charges in electromagnetism for mapping the electromagnetic field. The path or trajectory of a fluid particle represents an *integral curve* of the flow field. We return to this point when discussing fluid flow lines in Section 12.8.

12.2.2 Material fluid parcel

A *material fluid parcel* is an infinitesimal deformable fluid region that follows the fluid flow as specified by the velocity field (middle panel in Figure 12.1). A material parcel maintains a fixed mass, a fixed matter content, and a fixed specific entropy. Hence, it does not exchange matter, heat, or entropy with other fluid parcels. Furthermore, it does not experience irreversible exchange of momentum arising from friction. Its only interaction with adjacent fluid parcels is through reversible mechanical exchanges from pressure. It is thus a closed thermodynamic system yet open to 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, $\delta V > 0$, that deforms with the flow. Its mass is written $\delta M = \rho \, \delta V$, with ρ the mass density. Think of a material fluid parcel as a tiny perfectly insulated slippery bag full of gas or liquid. Even so, we never have reason in this book to specify the boundary of a fluid parcel. Rather, we only make use of the conceptual framework provided by fluid parcels as part of a Lagrangian formulation of fluid kinematics and dynamics.

12.2.3 Finite sized material fluid region

A material region is a finite volume generalization of a material fluid parcel. Conversely, a material fluid parcel is the infinitesimal limit of a material fluid region. That is, a material region is comprised of fixed mass and fixed matter content. Hence, as the material region moves through the fluid there is zero exchange of matter across its boundary. In contrast to the material fluid parcel, we are generally concerned with details of the boundary to material fluid regions.

12.2.4 Fluid element

A fluid element is an infinitesimal and deformable fluid region of fixed mass yet non-fixed matter and non-fixed specific entropy (right panel in Figure 12.1). For a homogeneous fluid comprised of a single matter constituent and no irreversible processes, then a fluid element is identical to a material fluid parcel. However, there is a distinction for non-homogeneous fluids, such as the ocean (e.g., fresh water, salts, biogeochemical tracers) and the atmosphere (e.g., nitrogen, oxygen, water, dust, carbon dioxide, trace chemical species). The exchange of matter across the boundary of a fluid element arises from the irreversible diffusive mixing of trace constituents within the fluid (Sections 15.1 and 47.2). As detailed in Section 15.1, diffusive matter exchange leaves the mass of the fluid element unchanged since the fluid element velocity is determined by its center of mass. As for a material fluid parcel, we have no reason in this book to specify the boundary of a fluid element. Instead, fluid elements are conceptual physical systems of use to formulate the kinematic and dynamic equations.

12.2.5 Finite sized fluid region

A fluid region is the most general subsystem within a fluid, whereby we consider an arbitrary finite region whose boundaries are open to the exchange of matter, mechanical forces, and thermodynamic properties with the surrounding environment. Here, we specify the boundary of the region and provide details for transport of properties across that boundary.

12.2.6 Comments

The fluid particle's sole purpose is to determine trajectories and the associated fluid pathlines. The material fluid parcel has the added feature of nonzero volume and an associated kinematic description leading to the continuity equation (Section 14.1). Mechanics of material fluid parcels are the focus of perfect fluid mechanics, where the fluid matter is comprised of a single homogeneous constituent and there are no irreversible processes.

Following page 3 of *Olbers et al.* (2012), we introduced the fluid element as the next most general infinitesimal fluid system beyond a material fluid parcel. Fluid elements allow us to consider real fluids with more than one matter constituent, with such multi-component fluids generally allowing for matter and other properties to be irreversibly exchanged between the elements. Much of the

kinematics in this chapter holds for both material fluid parcels and fluid elements. We thus typically refer in this chapter to "fluid parcels" for brevity. In Chapter 14 and elsewhere, we generally make the distinction when studying the kinematics of multi-constituent fluids.

Many authors do not distinguish between material fluid parcels and fluid elements. For our purposes, we make the distinction since the perfect fluid mechanics of material parcels is sufficient for only a limited number of fluid systems considered in this book. Fluid elements are required to formulate the kinematics and dynamics of multi-component fluids, in which irreversible exchanges occur between fluid elements thus leading to the diffusion of matter, heat, momentum, and other properties. Such exchanges are fundamental features of geophysical fluid mechanics.

We can make use of a fluid particle for both homogeneous and multi-component fluids. Again, the fluid particle tracks the trajectories of fluid as defined by the velocity field. For the multi-component fluid the velocity field is the barycentric velocity (see Section 15.1), which is the center of mass velocity. So fluid particles in a multi-component fluid track the barycentric velocity.

12.3 Lagrangian and Eulerian reference frames

There are two reference frames commonly used as the basis for describing motion of a fluid continuum.

- Lagrangian or material reference frame: This reference frame is defined by that of moving material fluid particles. 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.

- Eulerian or laboratory reference frame: The second reference frame is based on observing the fluid from a fixed spatial position, $\boldsymbol{x}$, 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 fluid mechanics is on fluid properties determined as a function of position $\boldsymbol{x}$ 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, we can retain our focus on classical fluid kinematics.

Figure 12.2: *Sample trajectory of a fluid particle with endpoints $\boldsymbol{X}(t)$ and $\boldsymbol{X}(t+\delta t)$. The trajectory passes through the point $\boldsymbol{x} = \boldsymbol{X}(t)$ at time t and $\boldsymbol{x} + \delta\boldsymbol{x} = \boldsymbol{X}(t+\delta t)$ 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.*

12.4 Material and position coordinates

A material description is suggested by the Lagrangian reference frame, whereby fluid particles are labeled with a material coordinate. This description complements the Eulerian, 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.

12.4.1 Trajectories of fluid particles

In describing the motion of a classical point particle (Chapter 9), we specify its spatial position according to a time dependent position vector $\boldsymbol{X}$ 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 $\boldsymbol{x}$, in which case we write

$$\boldsymbol{x} = \boldsymbol{X}(t) \qquad \text{point particle.} \tag{12.1}$$

A sample trajectory is shown in Figure 12.2. 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 $\boldsymbol{X}(t)$, whose instantaneous space position is denoted by the lowercase $\boldsymbol{x}$. This convention aims to distinguish functions, such as $\boldsymbol{X}(t)$, from the value of these functions at a point in time, $\boldsymbol{x}$.

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

$$\boldsymbol{x} = \boldsymbol{X}(n, t). \tag{12.2}$$

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, $\boldsymbol{a}$, which is referred to as the *material* coordinate. At time t, the position of a fluid element labelled by the material coordinate, $\boldsymbol{a}$, is written

$$\boldsymbol{x} = \boldsymbol{X}(\boldsymbol{a}, t) \qquad \text{continuum of matter.} \tag{12.3}$$

The continuous vector, $\boldsymbol{a}$, labels a point of matter within the continuum fluid.

12.4.2 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

$$\boldsymbol{a} = \boldsymbol{X}(t = t_0). \tag{12.4}$$

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 21). 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

$$\boldsymbol{a} = (X, Y, \theta)_{t=t_0}, \tag{12.5}$$

where we write θ as a measure of the specific entropy (or potential temperature; see Section 22.9). 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 Lagrangian coordinates. The mathematics and physics of these *generalized vertical coordinates* are detailed in Chapters 8, 17, and 38.

12.4.3 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 $\boldsymbol{a}$ to continuously move from an initial position $\boldsymbol{x}_0 = \boldsymbol{X}(\boldsymbol{a}, t_0)$ to another position $\boldsymbol{x} = \boldsymbol{X}(\boldsymbol{a}, t)$ at time $t > t_0$. The point transformation is defined by the vector relation

$$\boldsymbol{x} = \boldsymbol{X}(\boldsymbol{a}, t), \tag{12.6}$$

which is written in component form as[1]

$$x^m = X^m(a^i, t). \tag{12.7}$$

In fluid mechanics, the relation (12.6) defines a single-valued and invertible transformation from material coordinates $(\boldsymbol{a}, t)$ to position coordinates $(\boldsymbol{x}, t)$. That is, for each material coordinate $\boldsymbol{a}$ and time t, there is a unique spatial point $\boldsymbol{x}$, with this point specified by the trajectory $\boldsymbol{X}(\boldsymbol{a}, t)$. Conversely, for each space-time point $(\boldsymbol{x}, t)$ there exists a unique material coordinate found by inverting equation (12.6)[2]

$$\boldsymbol{a} = \boldsymbol{A}(\boldsymbol{x}, t). \tag{12.8}$$

[1] We choose tensor labels m, n, p for spatial coordinates and trajectories, and i, j, k for material coordinates.

[2] The use of $\boldsymbol{A}$ for the inverse function in equation (12.8) should not be confused with the acceleration, also written as $\boldsymbol{A}$ elsewhere in this book. We will not have much use for equation (12.8), thus minimizing the opportunity for confusion.

In this equation, $\boldsymbol{A}$ is the inverse function that specifies the material coordinate $\boldsymbol{a}$ given $(\boldsymbol{x}, t)$. 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 hypothesis 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.

12.4.4 Transformation matrix and its Jacobian

In our analysis of fluid motions, we make routine use of the position coordinates of an Eulerian description and material coordinates of a Lagrangian description. We here introduce a tensor framework to transform from one description to the other. Fundamental to that framework is the transformation matrix. Recall that in Section 7.2.1 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 12.6 and 12.7. Here, we develop the transformation matrix between position coordinates, termed $\boldsymbol{x}$-space, and material coordinates, termed $\boldsymbol{a}$-space.

The fluid particle trajectories, $\boldsymbol{X}(\boldsymbol{a}, t)$, are fundamental to this transformation. Namely, the trajectories as given by equation (12.6) provide a transformation between position coordinates $(\boldsymbol{x}, t)$ and material coordinates $(\boldsymbol{a}, t)$. The transformation is invertible so long as the Jacobian of the transformation matrix remains nonzero.

The transformation matrix is given by the matrix of partial derivatives, and we choose to organize this matrix according to the following convention

$$F_i^m = \frac{\partial X^m}{\partial a^i} \equiv \begin{bmatrix} \partial X^1/\partial a^1 & \partial X^1/\partial a^2 & \partial X^1/\partial a^3 \\ \partial X^2/\partial a^1 & \partial X^2/\partial a^2 & \partial X^2/\partial a^3 \\ \partial X^3/\partial a^1 & \partial X^3/\partial a^2 & \partial X^3/\partial a^3 \end{bmatrix}. \tag{12.9}$$

As defined by equation (12.9), the upper label, m, denotes the row and the lower label, i, is the column. As seen in Section 13.2, 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. The Jacobian of the transformation matrix is the determinant, which can be written in either of the following manners

$$\det(F_i^m) = \frac{\partial \boldsymbol{X}}{\partial \boldsymbol{a}} = \det \begin{bmatrix} \partial X^1/\partial a^1 & \partial X^1/\partial a^2 & \partial X^1/\partial a^3 \\ \partial X^2/\partial a^1 & \partial X^2/\partial a^2 & \partial X^2/\partial a^3 \\ \partial X^3/\partial a^1 & \partial X^3/\partial a^2 & \partial X^3/\partial a^3 \end{bmatrix}. \tag{12.10}$$

We make use of the notation $\partial \boldsymbol{X}/\partial \boldsymbol{a}$ for the Jacobian as it offers a useful means to distinguish between the determinant of the transformation (12.9), and the determinant of the inverse transformation, written as $\partial \boldsymbol{a}/\partial \boldsymbol{X}$.

12.4.5 A discrete algorithm to compute the transformation matrix

To help further our understanding of the transformation matrix (12.9), 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

$$\boldsymbol{X}(t) = X^1(t)\,\hat{\boldsymbol{x}} + X^2(t)\,\hat{\boldsymbol{y}}, \tag{12.11}$$

and use a Cartesian representation for the material coordinate

$$\boldsymbol{a} = a^1\,\hat{\boldsymbol{x}} + a^2\,\hat{\boldsymbol{y}}. \tag{12.12}$$

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

$$\boldsymbol{x}(e, f) = x(e, f)\,\hat{\boldsymbol{x}} + y(e, f)\,\hat{\boldsymbol{y}}. \tag{12.13}$$

Initialize fluid particles at each of the lattice grid points,

$$\boldsymbol{X}(e, f; t = 0) = \boldsymbol{x}(e, f) = \boldsymbol{a}(e, f), \tag{12.14}$$

with the discrete material coordinates defined by the initial positions. Then time step the trajectories using the velocity field to compute the particle pathlines $\boldsymbol{X}[\boldsymbol{a}(e, f); t]$ as illustrated in Figure 12.3. 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

$$F_i^m = \begin{bmatrix} F_1^1 & F_2^1 \\ F_1^2 & F_2^2 \end{bmatrix} \approx \begin{bmatrix} \frac{X^1(e+1,f;t)-X^1(e-1,f;t)}{X^1(e+1,f;0)-X^1(e-1,f;0)} & \frac{X^1(e,f+1;t)-X^1(e,f-1;t)}{X^2(e,f+1;0)-X^2(e,f-1;0)} \\ \frac{X^2(e+1,f;t)-X^2(e-1,f;t)}{X^1(e+1,f;0)-X^1(e-1,f;0)} & \frac{X^2(e,f+1;t)-X^2(e,f-1;t)}{X^2(e,f+1;0)-X^2(e,f-1;0)} \end{bmatrix}. \tag{12.15}$$

If the grid is regular in both directions, then the initial positions have a separation given by the grid spacing in which case

$$F_i^m \approx \begin{bmatrix} \frac{X^1(e+1,f;t)-X^1(e-1,f;t)}{\Delta} & \frac{X^1(e,f+1;t)-X^1(e,f-1;t)}{\Delta} \\ \frac{X^2(e+1,f;t)-X^2(e-1,f;t)}{\Delta} & \frac{X^2(e,f+1;t)-X^2(e,f-1;t)}{\Delta} \end{bmatrix}. \tag{12.16}$$

This algorithm illustrates how the transformation matrix provides a measure of trajectory spreading as fluid particles move away from their initial positions. That is, it measures the "deformation" of the fluid trajectories.

12.4.6 Jacobian as the ratio of volumes

We here establish the Jacobian as the ratio of volume elements written in position space and material space. This property holds at each point within the continuum fluid, and thus 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

$$\mathrm{d}V(\boldsymbol{x}) = \mathrm{d}x\,\mathrm{d}y\,\mathrm{d}z. \tag{12.17}$$

The $\boldsymbol{x}$ argument is introduced on the left hand side to distinguish this volume, which has dimensions L^3, from the volume written in material coordinates

$$\mathrm{d}V(\boldsymbol{a}) = \mathrm{d}a\,\mathrm{d}b\,\mathrm{d}c, \tag{12.18}$$

where $\boldsymbol{a} = (a^1, a^2, a^3) = (a, b, c)$ are coordinates in material space. Note that $\mathrm{d}V(\boldsymbol{a})$ 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 (12.5)).

Figure 12.3: *Illustrating the computational algorithm of Section 12.4.5 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 betwen the node points.*

The two volumes are related by the Jacobian of transformation between the two coordinate systems

$$\mathrm{d}V(\boldsymbol{x}) = \frac{\partial \boldsymbol{X}}{\partial \boldsymbol{a}}\,\mathrm{d}V(\boldsymbol{a}). \tag{12.19}$$

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

$$\frac{\partial \boldsymbol{X}}{\partial \boldsymbol{a}} = \frac{\mathrm{d}V(\boldsymbol{x})}{\mathrm{d}V(\boldsymbol{a})}. \tag{12.20}$$

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

$$\delta V(\boldsymbol{x}) = \frac{\partial \boldsymbol{X}}{\partial \boldsymbol{a}}\,\delta V(\boldsymbol{a}) \Rightarrow \frac{\partial \boldsymbol{X}}{\partial \boldsymbol{a}} = \frac{\delta V(\boldsymbol{x})}{\delta V(\boldsymbol{a})}. \tag{12.21}$$

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

$$\frac{\partial \boldsymbol{X}}{\partial \boldsymbol{X}(0)} = \frac{\delta V(\boldsymbol{x})}{\delta V(0)}. \tag{12.22}$$

12.4.7 Further study

This classic video from the *National Committee for Fluid Mechanics Films*, featuring Prof. Lumley, offers insightful visualizations to help understand Eulerian and Lagrangian fluid descriptions.

12.5 Lagrangian and Eulerian time derivatives

As noted in Section 12.4.2, 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, $\boldsymbol{x}$, whereas the material frame computes time derivatives by fixing the material coordinate, $\boldsymbol{a}$. 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.

12.5.1 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

$$\Pi = \Pi(\boldsymbol{x}, t). \tag{12.23}$$

The difference between $\Pi(\boldsymbol{x}, t)$ and $\Pi(\boldsymbol{x} + \mathrm{d}\boldsymbol{x}, t + \mathrm{d}t)$ delivers the differential increment, computed to leading order via a Taylor series expansion

$$\mathrm{d}\Pi = \Pi(\boldsymbol{x} + \mathrm{d}\boldsymbol{x}, t + \mathrm{d}t) - \Pi(\boldsymbol{x}, t) \tag{12.24a}$$

$$= \mathrm{d}t\, \partial_t \Pi + \mathrm{d}\boldsymbol{x} \cdot \nabla \Pi. \tag{12.24b}$$

In this equation, $\mathrm{d}t$ is the infinitesimal time increment, and $\mathrm{d}\boldsymbol{x}$ is the vector of infinitesimal space increments. For example, making use of Cartesian coordinates leads to the increment

$$\mathrm{d}\boldsymbol{x} = \hat{\boldsymbol{x}}\, \mathrm{d}x + \hat{\boldsymbol{y}}\, \mathrm{d}y + \hat{\boldsymbol{z}}\, \mathrm{d}z. \tag{12.25}$$

We ignore higher order terms in equation (12.24b) since the space and time increments are infinitesimal.

12.5.2 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 (12.24b) by the infinitesimal time increment

$$\frac{\mathrm{d}\Pi}{\mathrm{d}t} = \frac{\partial \Pi}{\partial t} + \frac{\mathrm{d}\boldsymbol{x}}{\mathrm{d}t} \cdot \nabla \Pi. \tag{12.26}$$

The first term measures the time derivative of Π at the point $\boldsymbol{x}$. The second term accounts for changes in Π arising from movement of the reference frame relative to a point $\boldsymbol{x}$ according to the velocity $\mathrm{d}\boldsymbol{x}/\mathrm{d}t$. Expression (12.26) holds in general since the velocity of the moving frame is arbitrary. We next specialize to the two common cases in fluid mechanics.

12.5.3 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

$$\text{Eulerian time derivative} = \frac{\partial \Pi(\boldsymbol{x}, t)}{\partial t}. \tag{12.27}$$

This result follows from specializing the total time derivative in equation (12.26) to the case of fixed spatial points, so that $\mathrm{d}\boldsymbol{x}/\mathrm{d}t = 0$. 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 frame measuring a zero time change for fluid properties.

12.5.4 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

$$\text{Lagrangian time derivative} = \frac{\mathrm{D}\Pi}{\mathrm{D}t} = \frac{\partial \Pi}{\partial t} + \boldsymbol{v} \cdot \nabla \Pi. \tag{12.28}$$

The second equality follows by setting $\mathrm{d}\boldsymbol{x}/\mathrm{d}t = \boldsymbol{v}$ in equation (12.26) since we are sampling points along the fluid particle trajectory $\boldsymbol{x} = \boldsymbol{X}(\boldsymbol{a}, t)$. The operator $\partial/\partial t$ is the Eulerian time derivative from equation (12.27), whereas $\boldsymbol{v} \cdot \nabla$ is referred to as the *advection* operator. Use of the capital D for the material time operator

$$\frac{\mathrm{D}}{\mathrm{D}t} = \frac{\partial}{\partial t} + \boldsymbol{v} \cdot \nabla \tag{12.29}$$

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 (12.26).

Equation (12.29) provides an Eulerian expression (right hand side) to the material time derivative $\mathrm{D}/\mathrm{D}t$. There are two Eulerian contributions: the local (fixed space point) time tendency $\partial/\partial t$ and advection, $\boldsymbol{v} \cdot \nabla$. 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 12.4 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.

12.5.5 Lagrangian time derivative formulated from the material frame

Rather than start from the total time derivative (12.26), 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 sometimes useful to introduce the shorthand notation

$$\Pi[\boldsymbol{X}(\boldsymbol{a}, t), t] \equiv \Pi^L(\boldsymbol{a}, t) \qquad \Leftarrow \text{sampling } \Pi \text{ on a trajectory } \boldsymbol{x} = \boldsymbol{X}(\boldsymbol{a}, t) \text{ at time } t. \tag{12.30}$$

Figure 12.4: *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, $\Pi^L(\boldsymbol{a}, t)$ is the function Π evaluated on a fluid particle trajectory $\boldsymbol{x} = \boldsymbol{X}(\boldsymbol{a}, t)$. That is, Π^L is the Lagrangian version of the function. For example, the Lagrangian velocity is given by

$$\boldsymbol{v}^L(\boldsymbol{a}, t) \equiv \boldsymbol{v}[\boldsymbol{X}(\boldsymbol{a}, t), t] = \left[\frac{\partial \boldsymbol{X}(\boldsymbol{a}, t)}{\partial t}\right]_{\boldsymbol{a}}. \tag{12.31}$$

The $\boldsymbol{a}$ subscript emphasizes that the partial time derivative is computed while holding $\boldsymbol{a}$ fixed.

The time derivative along a fluid particle trajectory is the material derivative. We introduce finite differences along the trajectory to estimate the material time derivative

$$\left[\frac{\partial \Pi[\boldsymbol{X}(\boldsymbol{a}, t), t]}{\partial t}\right]_{\boldsymbol{a}} = \lim_{\Delta t \to 0} \left[\frac{\Pi[\boldsymbol{X}(\boldsymbol{a}, t + \Delta t/2), t + \Delta t/2] - \Pi[\boldsymbol{X}(\boldsymbol{a}, t - \Delta t/2), t - \Delta t/2]}{\Delta t}\right]. \tag{12.32}$$

Expanding the numerator in a Taylor series, and keeping just the leading terms, yields

$$\left[\frac{\partial \Pi[\boldsymbol{X}(\boldsymbol{a}, t), t]}{\partial t}\right]_{\boldsymbol{a}} = \lim_{\Delta t \to 0} \left[\frac{\Pi[\boldsymbol{X}(\boldsymbol{a}, t + \Delta t/2), t + \Delta t/2] - \Pi[\boldsymbol{X}(\boldsymbol{a}, t - \Delta t/2), t - \Delta t/2]}{\Delta t}\right] \tag{12.33a}$$

$$= \left[\left(\frac{\partial}{\partial t}\right)_{\boldsymbol{X}} + \left(\frac{\partial \boldsymbol{X}(\boldsymbol{a}, t)}{\partial t}\right)_{\boldsymbol{a}} \cdot \nabla\right] \Pi[\boldsymbol{X}(\boldsymbol{a}, t), t] \tag{12.33b}$$

$$= \left[\left(\frac{\partial}{\partial t}\right)_{\boldsymbol{X}} + \boldsymbol{v}[\boldsymbol{X}(\boldsymbol{a}, t), t] \cdot \nabla\right] \Pi[\boldsymbol{X}(\boldsymbol{a}, t), t]. \tag{12.33c}$$

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 $\boldsymbol{X}(\boldsymbol{a}, t) = \boldsymbol{x}$ allows us to dispense with the trajectory notation to recover the more succinct expression (12.28). Even so, it is important to keep in mind the underlying trajectory basis for the material time derivative.

12.5.6 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 linear phase function $\Gamma(\boldsymbol{x}, t) = \boldsymbol{k} \cdot \boldsymbol{x} - \omega\, t$, where ω is a constant angular frequency and $\boldsymbol{k}$ 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 3.1) and wave equation (Section 3.6). Now consider the material time derivative of the phase function

$$\frac{\mathrm{D}(\boldsymbol{k} \cdot \boldsymbol{x} - \omega\, t)}{\mathrm{D}t} = \left[\frac{\partial}{\partial t} + \boldsymbol{v} \cdot \nabla\right](\boldsymbol{k} \cdot \boldsymbol{x} - \omega\, t) \tag{12.34a}$$

$$= \boldsymbol{v} \cdot \nabla\,(\boldsymbol{k} \cdot \boldsymbol{x}) - \frac{\partial(\omega\, t)}{\partial t} \tag{12.34b}$$

$$= \boldsymbol{v} \cdot \boldsymbol{k} - \omega. \tag{12.34c}$$

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

$$\frac{\mathrm{D}\Pi}{\mathrm{D}t} = 0. \tag{12.35}$$

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

$$\frac{\partial \Pi}{\partial t} = -\boldsymbol{v} \cdot \nabla \Pi. \tag{12.36}$$

Recall the case of a wave characteristic whereby $\partial\Pi/\partial t = \omega$. Wave characteristics are material if the fluid particle velocity satisfies $\boldsymbol{k} \cdot \boldsymbol{v} = \omega$. In one space dimension and for linear non-rotating gravity waves (Section 31.2), 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, $|v| = \omega/k$. This situation is very unlikely for most waves, whose speeds are much larger than fluid particle speeds. 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, with highly nonlinear behavior when the Froude number reaches unity (see Exercise 29.3).

We garner geometric insight into relation (12.36) by introducing the unit normal vector to the surface of constant Π

$$\hat{\boldsymbol{n}} = \frac{\nabla \Pi}{|\nabla \Pi|}. \tag{12.37}$$

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

$$\frac{\partial \Pi/\partial t}{|\nabla \Pi|} = -\boldsymbol{v} \cdot \hat{\boldsymbol{n}}. \tag{12.38}$$

That is, the fluid particle velocity, $\boldsymbol{v}$, is matched precisely to the velocity of the moving surface of constant Π. As detailed in Section 14.4.2, 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 $\boldsymbol{v}^{(s)}$. 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 $\boldsymbol{v}^{(s)} \neq \boldsymbol{v}$. Returning to the general expression (12.26) for the total time derivative, we have the time derivative operator as measured in this non-material moving reference frame

$$\frac{\mathrm{D}^{(s)}}{\mathrm{D}t} = \frac{\partial}{\partial t} + \boldsymbol{v}^{(s)} \cdot \nabla. \tag{12.39}$$

A function that remains constant within this general moving frame thus satisfies

$$\frac{\mathrm{D}^{(s)}\Pi}{\mathrm{D}t} = 0 \Rightarrow \frac{\partial \Pi}{\partial t} = -\boldsymbol{v}^{(s)} \cdot \nabla \Pi. \tag{12.40}$$

Introducing the normal direction $\hat{\boldsymbol{n}} = |\nabla \Pi|^{-1} \nabla \Pi$ leads to

$$\frac{\partial \Pi / \partial t}{|\nabla \Pi|} = -\boldsymbol{v}^{(s)} \cdot \hat{\boldsymbol{n}}, \tag{12.41}$$

which is an analog to the material constancy condition (12.38).

12.5.7 Summarizing some terminology for equations

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 placed on the fluid velocity in an incompressible fluid (see Section 16.1). 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; i.e., the property is a material constant.

12.6 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 9.1.2. 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 12.7, where we transform space and time derivative operators between an inertial frame and a rotating frame.

12.6.1 Galilean transformation

A Galilean transformation is given by the linear space-time transformation

$$\bar{t} = t \qquad \text{and} \qquad \overline{\boldsymbol{x}} = \boldsymbol{x} + \boldsymbol{U}t \qquad \text{and} \qquad \overline{\boldsymbol{v}} = \boldsymbol{v} + \boldsymbol{U}. \tag{12.42}$$

By convention, we say that the barred coordinates are those measured in the moving 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 $\boldsymbol{U}$. The inverse transformation is trivially given by

$$t = \bar{t} \qquad \text{and} \qquad \boldsymbol{x} = \overline{\boldsymbol{x}} - \boldsymbol{U}\bar{t} \qquad \text{and} \qquad \boldsymbol{v} = \overline{\boldsymbol{v}} - \boldsymbol{U}. \tag{12.43}$$

12.6.2 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 6.8. 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)

$$\frac{\partial}{\partial x^{\overline{\alpha}}} = \frac{\partial x^\alpha}{\partial x^{\overline{\alpha}}} \frac{\partial}{\partial x^\alpha}, \tag{12.44}$$

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

$$\frac{\partial x^{\overline{\alpha}}}{\partial x^\alpha} = \begin{bmatrix} \partial x^{\bar{0}}/\partial x^0 & \partial x^{\bar{0}}/\partial x^1 \\ \partial x^{\bar{1}}/\partial x^0 & \partial x^{\bar{1}}/\partial x^1 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ U & 1 \end{bmatrix}, \tag{12.45}$$

and the inverse is

$$\frac{\partial x^\alpha}{\partial x^{\overline{\alpha}}} = \begin{bmatrix} 1 & 0 \\ -U & 1 \end{bmatrix}. \tag{12.46}$$

The Jacobian determinant of the transformation matrix is unity, so that the Galilean transformation always has an inverse.

12.6.3 Transforming the differential operators

Given the transformation matrix, we can compute the Eulerian time derivative as measured in the moving frame is given by

$$\frac{\partial}{\partial x^{\bar{0}}} = \frac{\partial x^0}{\partial x^{\bar{0}}} \frac{\partial}{\partial x^0} + \frac{\partial x^1}{\partial x^{\bar{0}}} \frac{\partial}{\partial x^1} = \frac{\partial}{\partial x^0} - U \frac{\partial}{\partial x^1} = \frac{\partial}{\partial t} - U \frac{\partial}{\partial x}. \tag{12.47}$$

In words, this identity says that the time derivative computed between two inertial reference frames differs due to an advective term arising from the relative motion of the two inertial observers. The space derivatives are related by

$$\frac{\partial}{\partial x^{\bar{1}}} = \frac{\partial x^0}{\partial x^{\bar{1}}} \frac{\partial}{\partial x^0} + \frac{\partial x^1}{\partial x^{\bar{1}}} \frac{\partial}{\partial x^1} = \frac{\partial}{\partial x^1}, \tag{12.48}$$

so that the space derivative operator remains form invariant under a Galilean transformation. This result holds also for the other two space dimensions. Hence, the material time derivative operator is form invariant under a Galilean transformation

$$\frac{\mathrm{D}}{\mathrm{D}t} = \frac{\partial}{\partial t} + \boldsymbol{v} \cdot \nabla \tag{12.49a}$$

$$= \frac{\partial}{\partial \overline{t}} + \boldsymbol{U} \cdot \overline{\nabla} + (\overline{\boldsymbol{v}} - \boldsymbol{U}) \cdot \overline{\nabla} \tag{12.49b}$$

$$= \frac{\partial}{\partial \overline{t}} + \overline{\boldsymbol{v}} \cdot \overline{\nabla} \tag{12.49c}$$

$$= \frac{\mathrm{D}}{\mathrm{D}\overline{t}}\,, \tag{12.49d}$$

where we used the shorthand

$$\overline{\boldsymbol{v}} \cdot \overline{\nabla} = \overline{u}\,\frac{\partial}{\partial x^{\overline{1}}} + \overline{v}\,\frac{\partial}{\partial x^{\overline{2}}} + \overline{w}\,\frac{\partial}{\partial x^{\overline{3}}}. \tag{12.50}$$

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 $\mathrm{D}\Pi/\mathrm{d}t$ 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

12.7 Invariance of the material time derivative

In our discussion of Galilean invariance in Section 12.6, 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.

12.7.1 Invariance based on definition of the material time derivative

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

$$\frac{\mathrm{D}}{\mathrm{D}t} = \frac{\partial}{\partial t} + \boldsymbol{v} \cdot \nabla. \tag{12.51}$$

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.

12.7.2 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 7.1. Introduce Cartesian coordinates for the inertial frame, with corresponding basis vectors $(\hat{\overline{\boldsymbol{x}}}, \hat{\overline{\boldsymbol{y}}}, \hat{\overline{\boldsymbol{x}}})$. Let these inertial frame unit vectors be related to rotating frame unit vectors according to

$$\hat{\overline{\boldsymbol{x}}} = \hat{\boldsymbol{x}} \cos\vartheta - \hat{\boldsymbol{y}} \sin\vartheta \tag{12.52a}$$

$$\hat{\overline{\boldsymbol{y}}} = \hat{\boldsymbol{x}} \sin\vartheta + \hat{\boldsymbol{y}} \cos\vartheta \tag{12.52b}$$

$$\hat{\overline{\boldsymbol{z}}} = \hat{\boldsymbol{z}}, \tag{12.52c}$$

and let time be the same in the two reference frames. The angle ϑ measures the counter-clockwise angle between the inertial frame direction $\hat{\overline{\boldsymbol{x}}}$ and the moving frame direction $\hat{\boldsymbol{x}}$, with this angle a linear function of time

$$\vartheta = \Omega\, t. \tag{12.53}$$

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)

$$\overline{t} = t \tag{12.54a}$$

$$\overline{x} = x \cos\vartheta - y \sin\vartheta \tag{12.54b}$$

$$\overline{y} = x \sin\vartheta + y \cos\vartheta \tag{12.54c}$$

$$\overline{z} = z. \tag{12.54d}$$

The inverse transformation can be easily found

$$t = \overline{t} \tag{12.55a}$$

$$x = \overline{x} \cos\vartheta + \overline{y} \sin\vartheta \tag{12.55b}$$

$$y = -\overline{x} \sin\vartheta + \overline{y} \cos\vartheta \tag{12.55c}$$

$$z = \overline{z}. \tag{12.55d}$$

We are now prepared to make use of the transformation formalism generally considered in Section 6.8, and specifically applied for the Galilean transformation in Section 12.6. 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

$$\frac{\partial x^{\overline{\alpha}}}{\partial x^{\alpha}} = \begin{bmatrix} \partial x^{\overline{0}}/\partial x^0 & \partial x^{\overline{0}}/\partial x^1 & \partial x^{\overline{0}}/\partial x^2 & \partial x^{\overline{0}}/\partial x^3 \\ \partial x^{\overline{1}}/\partial x^0 & \partial x^{\overline{1}}/\partial x^1 & \partial x^{\overline{1}}/\partial x^2 & \partial x^{\overline{1}}/\partial x^3 \\ \partial x^{\overline{2}}/\partial x^0 & \partial x^{\overline{2}}/\partial x^1 & \partial x^{\overline{2}}/\partial x^2 & \partial x^{\overline{2}}/\partial x^3 \\ \partial x^{\overline{3}}/\partial x^0 & \partial x^{\overline{3}}/\partial x^1 & \partial x^{\overline{3}}/\partial x^2 & \partial x^{\overline{3}}/\partial x^3 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ -\Omega\,\overline{y} & \cos\vartheta & -\sin\vartheta & 0 \\ \Omega\,\overline{x} & \sin\vartheta & \cos\vartheta & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}. \tag{12.56}$$

Similarly, the inverse transformation is given by

$$\frac{\partial x^\alpha}{\partial x^{\overline{\alpha}}} = \begin{bmatrix} \partial x^0/\partial x^{\overline{0}} & \partial x^0/\partial x^{\overline{1}} & \partial x^0/\partial x^{\overline{2}} & \partial x^0/\partial x^{\overline{3}} \\ \partial x^1/\partial x^{\overline{0}} & \partial x^1/\partial x^{\overline{1}} & \partial x^1/\partial x^{\overline{2}} & \partial x^1/\partial x^{\overline{3}} \\ \partial x^2/\partial x^{\overline{0}} & \partial x^2/\partial x^{\overline{1}} & \partial x^2/\partial x^{\overline{2}} & \partial x^2/\partial x^{\overline{3}} \\ \partial x^3/\partial x^{\overline{0}} & \partial x^3/\partial x^{\overline{1}} & \partial x^3/\partial x^{\overline{2}} & \partial x^3/\partial x^{\overline{3}} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ \Omega\, y & \cos\vartheta & \sin\vartheta & 0 \\ -\Omega\, x & -\sin\vartheta & \cos\vartheta & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}. \tag{12.57}$$

The derivative operators transform according to

$$\frac{\partial}{\partial x^\alpha} = \frac{\partial x^{\overline{\alpha}}}{\partial x^\alpha}\,\frac{\partial}{\partial x^{\overline{\alpha}}}, \tag{12.58}$$

in which case

$$\frac{\partial}{\partial t} = \frac{\partial}{\partial \overline{t}} + (\boldsymbol{\Omega} \wedge \overline{\boldsymbol{x}}) \cdot \overline{\nabla} \tag{12.59a}$$

$$\frac{\partial}{\partial x} = \cos\vartheta\,\frac{\partial}{\partial \overline{x}} + \sin\vartheta\,\frac{\partial}{\partial \overline{y}} \tag{12.59b}$$

$$\frac{\partial}{\partial y} = -\sin\vartheta\,\frac{\partial}{\partial \overline{x}} + \cos\vartheta\,\frac{\partial}{\partial \overline{y}} \tag{12.59c}$$

$$\frac{\partial}{\partial z} = \frac{\partial}{\partial \overline{z}}. \tag{12.59d}$$

The velocity vector components transform according to

$$v^\alpha = \frac{\partial x^\alpha}{\partial x^{\overline{\alpha}}}\, v^{\overline{\alpha}}, \tag{12.60}$$

so that

$$v^0 = v^{\overline{0}} \tag{12.61a}$$

$$u = \Omega\, y + \overline{u}\cos\vartheta + \overline{v}\sin\vartheta \tag{12.61b}$$

$$v = -\Omega\, x - \overline{u}\sin\vartheta + \overline{v}\cos\vartheta \tag{12.61c}$$

$$w = \overline{w}. \tag{12.61d}$$

Bringing these result together leads to the transformation of the horizontal advection operator

$$u\,\frac{\partial}{\partial x} + v\,\frac{\partial}{\partial y} = (\overline{\boldsymbol{u}} - \boldsymbol{\Omega} \wedge \overline{\boldsymbol{x}}) \cdot \overline{\nabla}. \tag{12.62}$$

Combining this result with the transformed Eulerian time derivative leads to the material time derivative

$$\frac{\mathrm{D}}{\mathrm{D}t} = \frac{\partial}{\partial t} + \boldsymbol{v} \cdot \nabla \tag{12.63a}$$

$$= \frac{\partial}{\partial t} + u\,\frac{\partial}{\partial x} + v\,\frac{\partial}{\partial y} + w\,\frac{\partial}{\partial z} \tag{12.63b}$$

$$= \frac{\partial}{\partial \overline{t}} + (\boldsymbol{\Omega} \wedge \overline{\boldsymbol{x}}) \cdot \overline{\nabla} + (\overline{\boldsymbol{u}} - \boldsymbol{\Omega} \wedge \overline{\boldsymbol{x}}) \cdot \overline{\nabla} + \overline{w}\,\frac{\partial}{\partial \overline{z}} \tag{12.63c}$$

$$= \frac{\partial}{\partial \overline{t}} + \overline{\boldsymbol{v}} \cdot \overline{\nabla}. \tag{12.63d}$$

As advertised, the operator is form invariant under time dependent transformations to a non-inertial reference frame.

12.7.3 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.

12.8 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 mostly for pathlines and streamlines, though also introduce streaklines for completeness.

12.8.1 Material pathlines from fluid particle trajectories

As introduced in Section 12.4.1, a fluid particle traces out a *trajectory* as it moves through space (Figure 12.2). 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.

Mathematically, a fluid particle trajectory is a curve $\boldsymbol{x} = \boldsymbol{X}(\boldsymbol{a}, t)$ in space that is traced by fixing the material coordinate, $\boldsymbol{a}$, and letting time advance. Trajectories are computed by time integrating the ordinary differential equation

$$\frac{\partial \boldsymbol{X}(\boldsymbol{a}, t)}{\partial t} = \boldsymbol{v}[\boldsymbol{X}(\boldsymbol{a}, t), t] \tag{12.64a}$$

$$\boldsymbol{X}(\boldsymbol{a}, t = 0) = \boldsymbol{a}, \tag{12.64b}$$

where

$$\boldsymbol{v}[\boldsymbol{X}(\boldsymbol{a}, t), t] = \boldsymbol{v}^L(\boldsymbol{a}, t) \tag{12.65}$$

is Lagrangian velocity of the fluid particle (see equation (12.31)), 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 the laboratory, we can insert tiny trace particles into the fluid to offer a means for flow visualization. A time exposed photograph of the trace particles provides a visualization of fluid pathlines. Trace particles provide an increasingly accurate estimate of fluid particle pathlines if the trace particles do not disperse through diffusion (see Chapter 47). Another example offers further experience with pathliness, where here we consider cars moving at night. A time exposed photograph reveals pathlines for the cars as formed by their lights. Like cars, the material pathlines in a fluid can intersect, cross, and become quite complex, particularly when the flow is turbulent.

Figure 12.5: *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 (12.67)). 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.*

12.8.2 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

$$\mathrm{d}\boldsymbol{x} = \hat{\boldsymbol{x}}\,\mathrm{d}x + \hat{\boldsymbol{y}}\,\mathrm{d}y + \hat{\boldsymbol{z}}\,\mathrm{d}z \tag{12.66}$$

be an infinitesimal increment along a streamline written using Cartesian coordinates. The family of streamlines at a given time t satisfy the tangent constraint

$$\boldsymbol{v} \wedge \mathrm{d}\boldsymbol{x} = 0, \tag{12.67}$$

which is equivalent to

$$\frac{\mathrm{d}x}{u(\boldsymbol{x},t)} = \frac{\mathrm{d}y}{v(\boldsymbol{x},t)} = \frac{\mathrm{d}z}{w(\boldsymbol{x},t)}. \tag{12.68}$$

Alternatively, we can introduce a pseudo-time parameter, s, that determines a position along a streamline. Streamlines are the curves $\boldsymbol{x} = \boldsymbol{X}(s;\boldsymbol{a},t)$ computed with $(\boldsymbol{a},t)$ held fixed, but with the pseudo-time varied

$$\frac{\partial \boldsymbol{X}(s;\boldsymbol{a},t)}{\partial s} = \boldsymbol{v}[\boldsymbol{X}(s;\boldsymbol{a},t),t] \tag{12.69a}$$

$$\boldsymbol{X}(s=0;\boldsymbol{a},t) = \boldsymbol{a}. \tag{12.69b}$$

Again, both the material coordinate $\boldsymbol{a}$ and time t are held fixed when determining streamlines, so that $(\boldsymbol{a},t)$ 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 12.5). 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 pathlines. A streamtube is therefore a material tube for steady flow, in which case no fluid particles cross the streamtube boundary.

12.8.3 Distinguishing streamlines from pathlines for unsteady flow

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 (12.69a) is not equal to the time, t, used to compute fluid particle trajectories in equation (12.64a). Consequently, the condition $\boldsymbol{v} \cdot \hat{\boldsymbol{n}} = 0$ satisfied at each instance by a streamline does not ensure that fluid particles do not cross streamlines. The reason is that a material line moves with the fluid in such a way that

$$\boldsymbol{v} \cdot \hat{\boldsymbol{n}} = \boldsymbol{v}^{\text{line}} \cdot \hat{\boldsymbol{n}} \qquad \text{material lines,} \tag{12.70}$$

where $\boldsymbol{v}^{\text{line}}$ is the velocity of a point on the material line. The material line thus moves so that no fluid particles cross it. Only when the flow is steady, so that $\boldsymbol{v}^{\text{line}} \cdot \hat{\boldsymbol{n}} = 0$, will material lines and streamlines be equal. That is, the streamline constraint $\boldsymbol{v} \cdot \hat{\boldsymbol{n}} = 0$ is not a material constraint when $\boldsymbol{v}^{\text{line}} \cdot \hat{\boldsymbol{n}} \neq 0$. The key point is that streamlines do not probe the time behaviour of the flow, so they do not know whether the velocity is steady or unsteady.

The distinction between streamlines and pathlines for unsteady flow is sometimes not clearly stated in the literature. For example, on top of page 77 of *Williams and Elder* (1989), they state that since the flow is instantaneously tangent to a streamtube boundary, there can be no flow across the streamtube. *Kundu et al.* (2012) make a similar statement on their page 84. However, as emphasized above, streamlines equal to material pathlines only for steady flow in which the velocity has a zero time tendency.

12.8.4 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 12.6). 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 $\boldsymbol{y}$ is a curve going from $\boldsymbol{y}$ to $\boldsymbol{X}(\boldsymbol{y}, t)$, the position reached by the particle initialized at $t = 0$ at the point $\boldsymbol{y}$. A particle is on the streakline if it passed the fixed point $\boldsymbol{y}$ at some time between 0 and t. If this time was s, then the material coordinate of the particle would be given by $\boldsymbol{a}(\boldsymbol{y}, s)$ (see equation (12.8) relating the material coordinate to its corresponding laboratory position). Furthermore, at time t, this particle is at $\boldsymbol{x}$, so that the equation of the streakline at time t is

$$\boldsymbol{x} = \boldsymbol{X}[\boldsymbol{a}(\boldsymbol{y}, s), t] \qquad 0 \leq s \leq t. \tag{12.71}$$

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 $s \leq \tilde{t}$ in the function $\boldsymbol{X}(s; \boldsymbol{a}, \tilde{t})$. We determine the curves $\boldsymbol{x} = \boldsymbol{X}(s; \boldsymbol{a}, \tilde{t})$ by solving the following set of initial value problems for trajectories with initial conditions imposed at $t = s$ rather than $t = 0$

$$\frac{\partial \boldsymbol{X}(s; \boldsymbol{a}, t)}{\partial t} = \boldsymbol{v}[\boldsymbol{X}(s; \boldsymbol{a}, t), t] \tag{12.72a}$$

$$\boldsymbol{X}(t = s; \boldsymbol{a}, t) = \boldsymbol{a}. \tag{12.72b}$$

Note that $\boldsymbol{a}$ 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.

Figure 12.6: *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 $s < t$.*

12.8.5 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

$$u = \frac{x}{\tau + t} \tag{12.73a}$$

$$v = \frac{y}{\tau} \tag{12.73b}$$

$$w = 0, \tag{12.73c}$$

where $\tau > 0$ is a constant with the dimensions of time.

Pathlines

Pathlines are determined by solving the trajectory equations

$$\frac{\mathrm{d}X(t)}{\mathrm{d}t} = \frac{X(t)}{\tau + t} \tag{12.74a}$$

$$\frac{\mathrm{d}Y(t)}{\mathrm{d}t} = \frac{Y(t)}{\tau} \tag{12.74b}$$

$$\frac{\mathrm{d}Z(t)}{\mathrm{d}t} = 0, \tag{12.74c}$$

which are found to be

$$X(t) = X_0\,(1 + t/\tau) \tag{12.75a}$$

$$Y(t) = Y_0\,e^{t/\tau} \tag{12.75b}$$

$$Z(t) = Z_0, \tag{12.75c}$$

where $\boldsymbol{X}(t = 0) = \boldsymbol{X}_0$. Sample trajectories are shown in Figure 12.7 over time $t \in [0, 2]$. We can eliminate time to yield a curve in the horizontal (x, y) plane

$$y = Y_0\,e^{(x - X_0)/X_0}. \tag{12.76}$$

Figure 12.7: *Left panel: sample pathlines* $X(t) = X_0\,(1+t/\tau)$ *and* $Y(t) = Y_0\,e^{t/\tau}$ *(see equations (12.75a) and (12.75b)) 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 (12.78a) and (12.78b). 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 (12.81a) and (12.81b). 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.*

Streamlines

Streamlines are determined by solving the differential equations

$$\frac{\mathrm{d}X(s;t)}{\mathrm{d}s} = \frac{X(s;t)}{\tau + t} \tag{12.77a}$$

$$\frac{\mathrm{d}Y(s;t)}{\mathrm{d}s} = \frac{Y(s;t)}{\tau} \tag{12.77b}$$

$$\frac{\mathrm{d}Z(s;t)}{\mathrm{d}s} = 0, \tag{12.77c}$$

where time, t, is a fixed parameter whereas the pseudo-time, s, is varied. Integration renders the streamlines

$$X(s;t) = X_0\,e^{s/(\tau+t)} \tag{12.78a}$$

$$Y(s;t) = Y_0\,e^{s/\tau} \tag{12.78b}$$

$$Z(s;t) = Z_0. \tag{12.78c}$$

Sample streamlines are shown in Figure 12.7. Note that we can eliminate the pseudo-time s to render a curve in the horizontal (x, y) plane

$$y = Y_0 \left[\frac{x}{X_0}\right]^{(\tau+t)/\tau} \tag{12.79a}$$

$$z = Z_0. \tag{12.79b}$$

Streaklines

For streaklines, invert the trajectory expressions (12.75a)-(12.75b) to find the material coordinates $\boldsymbol{a}(\boldsymbol{y}, s)$ in the form

$$a_1 = \frac{y_1}{1 + s/\tau} \tag{12.80a}$$

$$a_2 = y_2\, e^{-s/\tau} \tag{12.80b}$$

$$a_3 = y_3. \tag{12.80c}$$

We next evaluate the trajectory expressions (12.75a)-(12.75b) with $\boldsymbol{a}$ as the initial positions to find the streaklines

$$X(s; \boldsymbol{a}, t) = \frac{y_1\,(1 + t/\tau)}{1 + s/\tau} \tag{12.81a}$$

$$Y(s; \boldsymbol{a}, t) = y_2\, e^{(t-s)/\tau} \tag{12.81b}$$

$$Z(s; \boldsymbol{a}, t) = y_3. \tag{12.81c}$$

Figure 12.7 illustrates the streakline for a particular point $(X, Y) = (1, 1)$.

12.8.6 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.* (2012), and online lecture notes on fluid kinematics from Professor McIntyre of Cambridge University.

12.9 Exercises

EXERCISE 12.1: 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

$$\frac{\mathrm{D}(\partial_m \Pi)}{\mathrm{D}t} = -\partial_m \boldsymbol{v} \cdot \nabla \Pi. \tag{12.82}$$

For example, if $\mathrm{D}\Pi/\mathrm{D}t = 0$, then the zonal partial derivative $\partial_x \Pi$ has a material time derivative given by

$$\frac{\mathrm{D}(\partial \Pi/\partial x)}{\mathrm{D}t} = -\frac{\partial \boldsymbol{v}}{\partial x} \cdot \nabla \Pi. \tag{12.83}$$

Hint: use Cartesian tensors for convenience.

(b) What is the material time derivative of $\nabla \Pi$ for the case that Π is not materially constant?

13

Material fluid objects

In this chapter we study the kinematics of material objects such as lines, surfaces, and volumes moving within a continuous fluid. The discussion moves seamlessly between Lagrangian and Eulerian descriptions, with the two offering complementary insights. Some attention is given to the kinematics of two-dimensional flow due to the relative mathematical ease and the associated intuition that proves useful for general geophysical flows.

READER'S GUIDE FOR THIS CHAPTER

This chapter builds from the kinematics of Chapter 12. It is of fundamental interest to a variety of kinematic aspects of fluid motion, with particular application to the kinematics of mixing and stirring of trace matter 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. Extension to arbitrary coordinates follow the methods of Chapter 6.

13.1 Increments in material and position space **218**
13.1.1 Eulerian/position space differential 218
13.1.2 Lagrangian/material space differential 218
13.1.3 Duality between Eulerian and Lagrangian perspectives 219
13.2 Evolution of a material line element **219**
13.2.1 Deformation gradient tensor 219
13.2.2 Cauchy-Green strain tensor 220
13.2.3 Material evolution of a line element 220
13.2.4 Velocity gradient tensor 221
13.2.5 Stretching and tilting of a material line element 221
13.2.6 Rate of strain tensor 222
13.2.7 Rotation tensor 223
13.2.8 Comments and further study 223
13.3 Evolution of a material area element **223**
13.3.1 Material area in three-dimensional flow 223
13.3.2 Material area in two-dimensional flow 225
13.4 Volume and the Jacobian of transformation **225**
13.4.1 Material parcel volume 225
13.4.2 Evolution of the Jacobian of transformation 226
13.5 Kinematics of two-dimensional flow **227**

13.5.1 Diverging flow 227
13.5.2 Rotational flow with nonzero vorticity 227
13.5.3 Flow with nonzero deformation 227
13.5.4 Further study 229
13.6 Exercises **229**

13.1 Increments in material and position space

In this section we summarize the mathematics associated with the differential increment of a function, exploring the increment in both position/Eulerian space and in material/Lagrangian space. We make use of the resulting dual expressions throughout this chapter.

13.1.1 Eulerian/position space differential

In Section 12.5.1, 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

$$\mathrm{d}\Pi(\boldsymbol{x},t) = \Pi(\boldsymbol{x}+\mathrm{d}\boldsymbol{x},t) - \Pi(\boldsymbol{x},t) \tag{13.1a}$$
$$= (\mathrm{d}\boldsymbol{x}\cdot\nabla)\Pi. \tag{13.1b}$$

The operator

$$\mathrm{d}\boldsymbol{x}\cdot\nabla = \mathrm{d}x_m\,\frac{\partial}{\partial x_m} \tag{13.2}$$

is a scalar since it remains form invariant when switching to another set of Cartesian position coordinates.[1]

13.1.2 Lagrangian/material space differential

Consider the same function Π evaluated on a material particle trajectory, and write this "Lagrangian" function as

$$\Pi^L(\boldsymbol{a},t) = \Pi[\boldsymbol{X}(\boldsymbol{a},t),t]. \tag{13.3}$$

In words, the Lagrangian version of a function is obtained by evaluating that function on a fluid particle trajectory. We use the notation $\Pi^L(\boldsymbol{a},t)$ as a shorthand, which is defined by this equality.

Consider an infinitesimal increment of $\Pi^L(\boldsymbol{a},t)$ within material coordinate space. This increment represents the difference of Π when evaluated on two separate fluid particles labelled by $\boldsymbol{a}$ and $\boldsymbol{a}+\delta\boldsymbol{a}$. Note that we use the δ symbol to signal material increments. Taking a Taylor series and truncating to leading order yields

$$\delta\Pi^L(\boldsymbol{a},t) = \Pi[\boldsymbol{X}(\boldsymbol{a}+\delta\boldsymbol{a},t),t] - \Pi[\boldsymbol{X}(\boldsymbol{a},t),t] \tag{13.4a}$$
$$= \Pi^L(\boldsymbol{a}+\delta\boldsymbol{a},t) - \Pi^L(\boldsymbol{a},t) \tag{13.4b}$$
$$= (\delta\boldsymbol{a}\cdot\nabla_{\boldsymbol{a}})\Pi^L(\boldsymbol{a},t). \tag{13.4c}$$

The operator

$$\delta\boldsymbol{a}\cdot\nabla_{\boldsymbol{a}} = \delta a_j\,\frac{\partial}{\partial a_j} \tag{13.5}$$

[1] This form invariance also holds when using curvilinear coordinates if we make use of the general tensor analysis formalism of Chapter 6.

Figure 13.1: *The ends of an infinitesimal material line element are defined by the trajectories of two fluid particles, $\boldsymbol{X}(\boldsymbol{a},t)$ and $\boldsymbol{X}(\boldsymbol{a}+\delta\boldsymbol{a},t)$. All points between the endpoints are part of the material line. Kinematics of the line element are determined by properties of the velocity gradient tensor discussed in Section 13.2.4.*

is a scalar since it remains form invariant when switching to another set of Cartesian material coordinates.[2] We use the notation $\nabla_{\boldsymbol{a}}$ to emphasize that the gradient operator is in material space rather than position space.

13.1.3 Duality between Eulerian and Lagrangian perspectives

By construction, the value of a function at a postion $\boldsymbol{x}$ (Eulerian perspective) equals to the function evaluated on a moving fluid particle (Lagrangian perspective) when the trajectory passes through $\boldsymbol{x}$. Mathematically, this identity takes the form

$$\Pi^L(\boldsymbol{a},t) = \Pi(\boldsymbol{x},t) \qquad \text{if } \boldsymbol{X}(\boldsymbol{a},t) = \boldsymbol{x}. \tag{13.6}$$

Likewise, if the infinitesimal increment in space, $\delta\boldsymbol{x}$, equals to the vector increment of the two fluid particles,

$$\delta\boldsymbol{X}(\boldsymbol{a},t) = \boldsymbol{X}(\boldsymbol{a}+\delta\boldsymbol{a},t) - \boldsymbol{X}(\boldsymbol{a},t), \tag{13.7}$$

then the functional increments are identical

$$\delta\Pi^L(\boldsymbol{a},t) = \delta\Pi(\boldsymbol{x},t) \qquad \text{if } \delta\boldsymbol{X}(\boldsymbol{a},t) = \delta\boldsymbol{x}, \tag{13.8}$$

where

$$\delta\Pi(\boldsymbol{x},t) = \Pi(\boldsymbol{x}+\delta\boldsymbol{x},t) - \Pi(\boldsymbol{x},t). \tag{13.9}$$

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.

13.2 Evolution of a material line element

Material line elements, and their generalizations to area and volume elements, are geometric objects that follow fluid particles. We initialize a material line element by drawing a line in the fluid and then following the fluid particles contained on the initial line. The material line element is stretched and folded by the fluid flow. We here develop the rudimentary kinematics of material line elements.

13.2.1 Deformation gradient tensor

A material line element is a small line marked in the fluid and whose motion follows that of fluid particles. Assume the line element endpoints are particles with trajectories $\boldsymbol{X}(\boldsymbol{a},t)$ and $\boldsymbol{X}(\boldsymbol{a}+\delta\boldsymbol{a},t)$ (see Figure 13.1). At time t, the vector displacement between these two particles is written

$$\delta\boldsymbol{X}(\boldsymbol{a},t) = \boldsymbol{X}(\boldsymbol{a}+\delta\boldsymbol{a},t) - \boldsymbol{X}(\boldsymbol{a},t). \tag{13.10}$$

[2] As for the position space, this invariance also holds when using curvilinear material coordinates if we make use of the general tensor analysis formalism of Chapter 6.

Expanding this expression to leading order yields

$$\delta \boldsymbol{X}(\boldsymbol{a},t) = \boldsymbol{X}(\boldsymbol{a}+\delta\boldsymbol{a},t) - \boldsymbol{X}(\boldsymbol{a},t) \approx (\delta\boldsymbol{a}\cdot\nabla_{\boldsymbol{a}})\,\boldsymbol{X}(\boldsymbol{a},t), \tag{13.11}$$

where again $\nabla_{\boldsymbol{a}}$ is the gradient operator acting on the material coordinates. Writing this equation in component form leads to

$$\delta X_m = \delta a_j\,\frac{\partial X_m}{\partial a_j}. \tag{13.12}$$

As in Chapter 12, we assign the labels for m, n, p position/Eulerian coordinates, $\boldsymbol{x}$, and the labels i, j, k for material/Lagrangian coordinates, $\boldsymbol{a}$.

The components

$$F_{mj} \equiv \frac{\partial X_m}{\partial a_j} \tag{13.13}$$

appearing in equation (13.12) form elements of the transformation matrix linking position space to material space. We already encountered this tensor in Section 12.4.3 (see equation (12.9)). In the continuum mechanics literature, the tensor (13.13) is known as the *displacement gradient* tensor or the *deformation gradient* tensor.

13.2.2 Cauchy-Green strain tensor

The deformation gradient tensor plays a role in measuring the length of material line elements. We see this role by considering the squared length of a line element

$$\delta\boldsymbol{X}\cdot\delta\boldsymbol{X} = \frac{\partial\boldsymbol{X}}{\partial a_i}\cdot\frac{\partial\boldsymbol{X}}{\partial a_j}\,\delta a_i\,\delta a_j \equiv C_{ij}\,\delta a_i\,\delta a_j. \tag{13.14}$$

Algebraically, this expression is a quadratic form, and the symmetric tensor

$$C_{ij} \equiv \frac{\partial\boldsymbol{X}}{\partial a_i}\cdot\frac{\partial\boldsymbol{X}}{\partial a_j} = F_{mi}\,F_{mj} \tag{13.15}$$

is the metric tensor that provides the means to measure distance along an infinitesimal material line element. This metric tensor is called the *Cauchy-Green strain* tensor in the continuum mechanics literature.

13.2.3 Material evolution of a line element

Now consider the material time derivative of the material line element

$$\frac{\partial[\delta\boldsymbol{X}(\boldsymbol{a},t)]}{\partial t} = \frac{\partial\boldsymbol{X}(\boldsymbol{a}+\delta\boldsymbol{a},t)}{\partial t} - \frac{\partial\boldsymbol{X}(\boldsymbol{a},t)}{\partial t} \tag{13.16a}$$

$$= \boldsymbol{v}^L(\boldsymbol{a}+\delta\boldsymbol{a},t) - \boldsymbol{v}^L(\boldsymbol{a},t) \tag{13.16b}$$

$$\equiv \delta\boldsymbol{v}^L(\boldsymbol{a},t). \tag{13.16c}$$

In these equations, we introduced the Lagrangian velocity

$$\boldsymbol{v}^L(\boldsymbol{a},t) = \boldsymbol{v}[\boldsymbol{X}(\boldsymbol{a},t),t] \tag{13.17}$$

as per equation (12.31) and the discussion in Section 13.1.2.

As for the line element manipulations in Section 13.2.1, we can massage the expression (13.16c) by performing a Taylor series expansion and truncating to leading order

$$\frac{\partial[\delta \boldsymbol{X}(\boldsymbol{a},t)]}{\partial t} = \delta \boldsymbol{v}^L(\boldsymbol{a},t) = (\delta \boldsymbol{a} \cdot \nabla_{\boldsymbol{a}})\, \boldsymbol{v}^L(\boldsymbol{a},t). \tag{13.18}$$

Alternatively, we can choose to evaluate this expression using an Eulerian perspective (see Section 13.1.3), in which case

$$\frac{\mathrm{D}(\delta \boldsymbol{x})}{\mathrm{D}t} = \delta \boldsymbol{v}(\boldsymbol{x},t) = (\delta \boldsymbol{x} \cdot \nabla_{\boldsymbol{x}})\, \boldsymbol{v}(\boldsymbol{x},t). \tag{13.19}$$

13.2.4 Velocity gradient tensor

Writing the Eulerian result (13.19) in component form leads to

$$\frac{\mathrm{D}(\delta x_m)}{\mathrm{D}t} = \delta x_n \frac{\partial v_m}{\partial x_n}. \tag{13.20}$$

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 line element 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

$$\frac{\partial v_m}{\partial x_n} = \frac{1}{2}\left[\frac{\partial v_m}{\partial x_n} + \frac{\partial v_n}{\partial x_m}\right] + \frac{1}{2}\left[\frac{\partial v_m}{\partial x_n} - \frac{\partial v_n}{\partial x_m}\right] \tag{13.21a}$$

$$\equiv \mathbb{S}_{mn} + \mathbb{A}_{mn}, \tag{13.21b}$$

where[3]

$$\mathbb{S}_{mn} = \frac{1}{2}\left[\frac{\partial v_m}{\partial x_n} + \frac{\partial v_n}{\partial x_m}\right] = \mathbb{S}_{nm} \qquad \text{rate of strain tensor} \tag{13.22a}$$

$$\mathbb{A}_{mn} = \frac{1}{2}\left[\frac{\partial v_m}{\partial x_n} - \frac{\partial v_n}{\partial x_m}\right] = -\mathbb{A}_{nm} \qquad \text{rotation tensor.} \tag{13.22b}$$

As seen in the following, these tensors affect the line element evolution in very distinct manners.

13.2.5 Stretching and tilting of a material line element

Consider a line element that is initially aligned with the vertical axis

$$\delta \boldsymbol{x}_{t=0} = \hat{\boldsymbol{z}}\, \delta Z_0. \tag{13.23}$$

Consequently, the initial evolution of this material line element takes on the form

$$\frac{\mathrm{D}(\delta x)}{\mathrm{D}t} = \underbrace{\delta Z_0 \left[\frac{\partial u}{\partial z}\right]}_{\text{tilting}} \qquad \frac{\mathrm{D}(\delta y)}{\mathrm{D}t} = \underbrace{\delta Z_0 \left[\frac{\partial v}{\partial z}\right]}_{\text{tilting}} \qquad \frac{\mathrm{D}(\delta z)}{\mathrm{D}t} = \underbrace{\delta Z_0 \left[\frac{\partial w}{\partial z}\right]}_{\text{stretching}}. \tag{13.24}$$

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 line element onto the horizontal plane. That

[3]The rate of strain tensor, $\mathbb{S}_{mn}$, is sometimes called the *deformation* tensor in the fluid dynamics literature. We avoid this nomenclature to avoid confusion with the *deformation gradient* tensor defined by equation (13.13).

is, these terms *tilt* the line element. Additionally, in the presence of a vertical derivative in the vertical velocity, the line element 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 34. 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.

13.2.6 Rate of strain tensor

Recall the expression (13.14) for the squared length of a line element

$$\delta \boldsymbol{X} \cdot \delta \boldsymbol{X} = \frac{\partial \boldsymbol{X}}{\partial a_i} \cdot \frac{\partial \boldsymbol{X}}{\partial a_j} \delta a_i \, \delta a_j. \tag{13.25}$$

Its material time derivative is given by

$$\left[\frac{\partial (\delta \boldsymbol{X} \cdot \delta \boldsymbol{X})}{\partial t} \right]_{\boldsymbol{a}} = 2 \, \frac{\partial \boldsymbol{v}^L}{\partial a_i} \cdot \frac{\partial \boldsymbol{X}}{\partial a_j} \, \delta a_i \, \delta a_j. \tag{13.26}$$

We can express this result using Eulerian $\boldsymbol{x}$-coordinates by making use of the duality described in Section 13.1.3, which leads to

$$\frac{\partial \boldsymbol{v}^L}{\partial a_i} \, \delta a_i = \frac{\partial \boldsymbol{v}}{\partial x_n} \, \delta x_n \tag{13.27a}$$

$$\frac{\partial \boldsymbol{X}}{\partial a_j} \, \delta a_j = \delta \boldsymbol{x}, \tag{13.27b}$$

so that

$$\frac{\mathrm{D}(\delta \boldsymbol{x} \cdot \delta \boldsymbol{x})}{\mathrm{D}t} = 2 \, \frac{\partial v_m}{\partial x_n} \, \delta x_n \, \delta x_m. \tag{13.28}$$

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

$$\frac{1}{2} \frac{\mathrm{D}(\delta \boldsymbol{x} \cdot \delta \boldsymbol{x})}{\mathrm{D}t} = \mathbb{S}_{mn} \, \delta x_n \, \delta x_m. \tag{13.29}$$

Consequently, the rate of strain tensor, $\mathbb{S}_{mn}$, determines the rate at which a material line element changes its length. When the rate of strain tensor vanishes, then the line element retains a constant length. We can understand the result (13.29) 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 (13.29).

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 line elements oriented according to the principle axes (eigenvectors) expand/contract under the impacts of straining motion in the fluid. According to equation (13.29), 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 13.4, the sum of these eigenvectors (trace of the rate of strain tensor) measures the rate that a volume element changes through the divergence of the velocity

$$\mathbb{S}_{mm} = \nabla \cdot \boldsymbol{v}. \tag{13.30}$$

13.2.7 Rotation tensor

The rotation tensor

$$\mathbb{A}_{mn} = \frac{1}{2}\left[\frac{\partial v_m}{\partial x_n} - \frac{\partial v_n}{\partial x_m}\right] \tag{13.31}$$

is anti-symmetric: $\mathbb{A}_{mn} = -\mathbb{A}_{nm}$. Its components are related to the vorticity vector $\boldsymbol{\omega} = \nabla \wedge \boldsymbol{v}$ according to

$$\mathbb{A}_{mn} = -\epsilon_{mnp}\,\omega_p/2 \Longleftrightarrow \mathbb{A} = \frac{1}{2}\begin{bmatrix} 0 & -\omega_3 & \omega_2 \\ \omega_3 & 0 & -\omega_1 \\ -\omega_2 & \omega_1 & 0 \end{bmatrix}. \tag{13.32}$$

The contribution of the rotation matrix to material evolution of the line element is given by

$$\left[\frac{\mathrm{D}(\delta x_m)}{\mathrm{D}t}\right]_{\mathrm{rot}} = \mathbb{A}_{mn}\,\delta x_n = -(\epsilon_{mnp}\,\omega_p/2)\,\delta x_n \Longrightarrow \left[\frac{\mathrm{D}(\delta \boldsymbol{x})}{\mathrm{D}t}\right]_{\mathrm{rot}} = \frac{1}{2}\left(\boldsymbol{\omega} \wedge \delta \boldsymbol{x}\right). \tag{13.33}$$

This relation is in the form of a pure rotation of the vector $\delta\boldsymbol{x}$ generated by the vector $\boldsymbol{\omega}/2$ (recall the discussion of rotations in Section 9.4). We thus conclude that the anti-symmetric rotation tensor, $\mathbb{A}_{mn}$, provides a rigid body rotation to a fluid line element (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 32.1.

13.2.8 Comments and further study

The above discussion of how fluid motion impacts on a material line element 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 the equation (13.20) to read

$$v_m(\boldsymbol{x}, t) = v_m(\boldsymbol{x}_0, t) + \mathbb{S}_{mn}\,\delta x_n + \mathbb{A}_{mn}\,\delta x_n, \tag{13.34}$$

which can be written in the dyadic form

$$\boldsymbol{v}(\boldsymbol{x}, t) = \boldsymbol{v}(\boldsymbol{x}_0, t) + \mathbb{S}\cdot\delta\boldsymbol{x} + \mathbb{A}\cdot\delta\boldsymbol{x}. \tag{13.35}$$

Figure 13.2 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), with a brief summary in Section 1.1 of *Olbers et al.* (2012).

13.3 Evolution of a material area element

We extend the discussion of material line elements in Section 13.2 to a material area element such as that shown in Figure 13.3. We consider area elements in both three-dimensional and two-dimensional flows.

13.3.1 Material area in three-dimensional flow

Following from the geometric interpretation of the vector product in Section 1.4.4, we here define a material area element by (see Figure 13.3)

$$\delta\boldsymbol{S} = \delta\boldsymbol{A} \wedge \delta\boldsymbol{B} \Rightarrow \delta\mathcal{S}_m = \epsilon_{mnp}\,\delta A_n\,\delta B_p \tag{13.36}$$

Figure 13.2: *Schematic illustrating the decomposition of the manners that fluid flow can modify a material volume. First the sphere can be deformed without rotation, with this process encompassed by the rate of strate 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 are meant to represent the principle axes of the body, so that deformation of the sphere corresponds to expansion or contraction along the principle axes directions.*

Figure 13.3: *A material area defined by the cross product of two material line elements, $\delta \boldsymbol{S} = \delta \boldsymbol{A} \wedge \delta \boldsymbol{B}$. In the special case of $\delta \boldsymbol{A} = \hat{\boldsymbol{x}}\, \delta x$ and $\delta \boldsymbol{B} = \hat{\boldsymbol{y}}\, \delta y$, then $\delta \boldsymbol{S} = \delta x\, \delta y\, \hat{\boldsymbol{z}}$.*

where $\delta \boldsymbol{A}$ and $\delta \boldsymbol{B}$ are non-parallel infinitesimal material line elements. The area projected onto the direction $\hat{\boldsymbol{n}}$ is given by

$$\hat{\boldsymbol{n}} \cdot \delta \boldsymbol{S} = \hat{\boldsymbol{n}} \cdot (\delta \boldsymbol{A} \wedge \delta \boldsymbol{B}). \tag{13.37}$$

The evolution of the material area element is given by

$$\frac{\mathrm{D}(\delta \boldsymbol{S})}{\mathrm{D}t} = \frac{\mathrm{D}(\delta \boldsymbol{A})}{\mathrm{D}t} \wedge \delta \boldsymbol{B} + \delta \boldsymbol{A} \wedge \frac{\mathrm{D}(\delta \boldsymbol{B})}{\mathrm{D}t} \tag{13.38a}$$

$$= [(\delta \boldsymbol{A} \cdot \nabla)\, \boldsymbol{v}] \wedge \delta \boldsymbol{B} + \delta \boldsymbol{A} \wedge [(\delta \boldsymbol{B} \cdot \nabla)\, \boldsymbol{v}], \tag{13.38b}$$

where the second equality made use of the line element evolution equation (13.19). To proceed we expose indices and make use of some tensor identities

$$\frac{\mathrm{D}(\delta \mathcal{S}_m)}{\mathrm{D}t} = \epsilon_{mnp}\, [(\delta A_q\, \partial_q)\, v_n]\, \delta B_p + \epsilon_{mnp}\, \delta A_n\, [(\delta B_q\, \partial_q)\, v_p] \tag{13.39a}$$

$$= \epsilon_{mnp}\, [\delta A_q\, \delta B_p\, \partial_q v_n + \delta A_n\, \delta B_q\, \partial_q\, v_p] \tag{13.39b}$$

$$= \epsilon_{mnp}\, \partial_q v_n\, [\delta A_q\, \delta B_p - \delta A_p\, \delta B_q] \tag{13.39c}$$

$$= \epsilon_{mnp}\, \partial_q v_n\, \epsilon_{rqp}\, \delta \mathcal{S}_r \tag{13.39d}$$

$$= (\delta_{mr}\, \delta_{nq} - \delta_{mq}\, \delta_{nr}) \partial_q v_n\, \delta \mathcal{S}_r \tag{13.39e}$$

$$= (\nabla \cdot \boldsymbol{v})\, \delta \mathcal{S}_m - (\partial_m \boldsymbol{v}) \cdot \delta \boldsymbol{S}. \tag{13.39f}$$

To reach this result we made use of the following identities

$$\delta A_q\, \delta B_p - \delta A_p\, \delta B_q = \epsilon_{rqp}\, \delta \mathcal{S}_r \tag{13.40a}$$

$$\epsilon_{mnp}\, \epsilon_{rqp} = \delta_{mr}\, \delta_{nq} - \delta_{mq}\, \delta_{nr} \tag{13.40b}$$

Figure 13.4: *A parallelepiped defined by three material line elements, with volume (to within a sign) given by* $\delta V = (\delta \boldsymbol{A} \wedge \delta \boldsymbol{B}) \cdot \delta \boldsymbol{C}$. *See also the discussion surrounding Figure 1.4.*

where δ_{mn} is the Kronecker (or identity) tensor.

13.3.2 Material area in two-dimensional flow

Now consider a material area element for two-dimensional fluid flow with velocity, $\boldsymbol{v} = (u, v, 0)$, and $\delta \boldsymbol{A} = \hat{\boldsymbol{x}}\, \delta x$, $\delta \boldsymbol{B} = \hat{\boldsymbol{y}}\, \delta y$, with zero dependence on z. In this case, the area of an infinitesimal material region is

$$\delta \mathcal{S} = (\delta \boldsymbol{A} \wedge \delta \boldsymbol{B}) \cdot \hat{\boldsymbol{z}} = \delta x\, \delta y, \tag{13.41}$$

and its evolution is given by

$$\frac{\mathrm{D}(\delta \mathcal{S})}{\mathrm{D}t} = (\delta \boldsymbol{B} \wedge \hat{\boldsymbol{z}}) \cdot (\delta \boldsymbol{A} \cdot \nabla)\, \boldsymbol{u} + (\hat{\boldsymbol{z}} \wedge \delta \boldsymbol{A}) \cdot (\delta \boldsymbol{B} \cdot \nabla)\, \boldsymbol{u} \tag{13.42a}$$

$$= \delta x\, \delta y\, \nabla \cdot \boldsymbol{u}, \tag{13.42b}$$

so that

$$\frac{1}{\delta \mathcal{S}} \frac{\mathrm{D}(\delta \mathcal{S})}{\mathrm{D}t} = \nabla \cdot \boldsymbol{u}. \tag{13.43}$$

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 general result (13.39f) to the case of two-dimensional flow with no dependence on the vertical direction.

13.4 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.

13.4.1 Material parcel volume

Consider a material region with a volume δV spanned by the infinitesimal material line elements $\delta \boldsymbol{A}$, $\delta \boldsymbol{B}$, and $\delta \boldsymbol{C}$ (see Figure 13.4). To within a sign the volume is given by

$$\delta V = (\delta \boldsymbol{A} \wedge \delta \boldsymbol{B}) \cdot \delta \boldsymbol{C}. \tag{13.44}$$

Making use of the line element evolution equation (13.19) renders

$$\frac{\mathrm{D}(\delta V)}{\mathrm{D}t} = (\delta \boldsymbol{B} \wedge \delta \boldsymbol{C}) \cdot (\delta \boldsymbol{A} \cdot \nabla)\boldsymbol{v} + (\delta \boldsymbol{C} \wedge \delta \boldsymbol{A}) \cdot (\delta \boldsymbol{B} \cdot \nabla)\boldsymbol{v} + (\delta \boldsymbol{A} \wedge \delta \boldsymbol{B}) \cdot (\delta \boldsymbol{C} \cdot \nabla)\boldsymbol{v}. \tag{13.45}$$

Now specialize to the case where the parcel is a parallepiped oriented according to the coordinate axes

$$\delta \boldsymbol{A} = \hat{\boldsymbol{x}}\, \delta x \qquad \delta \boldsymbol{B} = \hat{\boldsymbol{y}}\, \delta y \qquad \delta \boldsymbol{C} = \hat{\boldsymbol{z}}\, \delta z, \tag{13.46}$$

so that

$$\delta V = \delta x\, \delta y\, \delta z. \tag{13.47}$$

Plugging into equation (13.45) leads to

$$\frac{1}{\delta V} \frac{\mathrm{D}(\delta V)}{\mathrm{D}t} = \nabla \cdot \boldsymbol{v}. \tag{13.48}$$

This result is a three-dimensional generalization of the material area equation derived in Section 13.3.2.

We offer an alternative derivation of equation (13.48) in Section 14.1, 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 ($\nabla \cdot \boldsymbol{v} > 0$). We think of a diverging velocity field as "pushing out" the material parcel boundary, thus increasing its volume. In contrast, the volume of a material parcel decreases where the fluid velocity converges ($\nabla \cdot \boldsymbol{v} < 0$)

$$\nabla \cdot \boldsymbol{v} > 0 \Rightarrow \text{material volume increases in diverging flow} \Rightarrow \text{parcel expands} \tag{13.49a}$$

$$\nabla \cdot \boldsymbol{v} < 0 \Rightarrow \text{material volume decreases in converging flow} \Rightarrow \text{parcel contracts.} \tag{13.49b}$$

13.4.2 Evolution of the Jacobian of transformation

Recall the discussion in Section 12.4.6, where we showed that the Jacobian of transformation between material space $(\boldsymbol{a}, t)$ and position space $(\boldsymbol{x}, t)$ is related to the ratio of the volume elements written in the two coordinate systems. In particular, equation (12.21) is given by

$$\delta V(\boldsymbol{x}) = \frac{\partial \boldsymbol{X}}{\partial \boldsymbol{a}}\, \delta V(\boldsymbol{a}) \Rightarrow \frac{\partial \boldsymbol{X}}{\partial \boldsymbol{a}} = \frac{\delta V(\boldsymbol{x})}{\delta V(\boldsymbol{a})}. \tag{13.50}$$

The material coordinate volume element $\delta V(\boldsymbol{a})$ is time independent when following the flow. Consequently, the material evolution of the Jacobian is given by

$$\frac{\mathrm{D}}{\mathrm{D}t} \frac{\partial \boldsymbol{X}}{\partial \boldsymbol{a}} = \frac{1}{\delta V(\boldsymbol{a})} \frac{\mathrm{D}(\delta V(\boldsymbol{x}))}{\mathrm{D}t} \tag{13.51a}$$

$$= \frac{\delta V(\boldsymbol{x})}{\delta V(\boldsymbol{a})}\, \nabla \cdot \boldsymbol{v} \tag{13.51b}$$

$$= \frac{\partial \boldsymbol{X}}{\partial \boldsymbol{a}}\, \nabla \cdot \boldsymbol{v}. \tag{13.51c}$$

The second equality made use of the result (13.48), 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

$$\left[\frac{\partial \boldsymbol{X}}{\partial \boldsymbol{a}}\right]^{-1} \frac{\mathrm{D}}{\mathrm{D}t} \left[\frac{\partial \boldsymbol{X}}{\partial \boldsymbol{a}}\right] = \nabla \cdot \boldsymbol{v}. \tag{13.52}$$

This equation is identical to the parcel volume equation (13.48), which is expected given the relation between the Jacobian and the parcel volume. In Exercise 13.2, we derive this result using the explicit expression for the Jacobian in terms of the ϵ-tensor.

13.5 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. In so doing we expose kinematic properties commonly used to characterise two-dimensional flow. Generalizations to three-dimensions are available with a bit more mathematical formalism.

Our starting point is Figure 13.5, 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 13.2.4, here written for the two-dimensional flow with horizontal velocity components, (u, v)

$$\frac{\partial v_m}{\partial x_n} = \begin{bmatrix} \dfrac{\partial u}{\partial x} & \dfrac{\partial u}{\partial y} \\ \dfrac{\partial v}{\partial x} & \dfrac{\partial v}{\partial y} \end{bmatrix} = \begin{bmatrix} \dfrac{\partial u}{\partial x} & \dfrac{1}{2}\left(\dfrac{\partial u}{\partial y} + \dfrac{\partial v}{\partial x}\right) \\ \dfrac{1}{2}\left(\dfrac{\partial u}{\partial y} + \dfrac{\partial v}{\partial x}\right) & \dfrac{\partial v}{\partial y} \end{bmatrix} + \frac{\zeta}{2}\begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} = \mathbb{S}_{mn} + \mathbb{A}_{mn}, \quad (13.53)$$

where

$$\zeta = \hat{\boldsymbol{z}} \cdot \nabla \wedge \boldsymbol{v} = \frac{\partial v}{\partial x} - \frac{\partial u}{\partial y} \quad (13.54)$$

is the vertical component to the vorticity.

13.5.1 Diverging flow

Recall from Section 13.3.2 that the area of a fluid element in two-dimensional flow changes according to the divergence. The upper left panel of Figure 13.5 thus illustrates equation (13.43)

$$\frac{1}{\delta\mathcal{S}} \frac{\mathrm{D}(\delta\mathcal{S})}{\mathrm{D}t} = \nabla \cdot \boldsymbol{u} = \mathbb{S}_{mm}, \quad (13.55)$$

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 $\nabla \cdot \boldsymbol{u} > 0$, leads to an expansion of the area. The opposite occurs for a converging flow, where the area compresses.

13.5.2 Rotational flow with nonzero vorticity

The upper right panel of Figure 13.5 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.

13.5.3 Flow with nonzero deformation

The lower left panel of Figure 13.5 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, $\nabla \cdot \boldsymbol{u} = 0$, and has zero

Figure 13.5: *Illustrating the varieties of shape changes for a square material fluid region. Upper left: purely divergent flow, whereby* $\nabla \cdot \boldsymbol{u} > 0$ *yet with zero vorticity, thus leading to a uniform increase in the area. Upper right: rotational flow with nonzero vorticity,* $\zeta = \hat{\boldsymbol{z}} \cdot (\nabla \wedge \boldsymbol{u})$*, 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* $\boldsymbol{u} = \hat{\boldsymbol{z}} \wedge \nabla\psi$*, which is generated by the streamfunction* $\psi = -\gamma\, x\, y$*. Lower right: pure shearing strain flow with zero divergence and zero vorticity, as as realized by the flow* $\boldsymbol{u} = \hat{\boldsymbol{z}} \wedge \nabla\psi$ *generated by the streamfunction* $\psi = -(\gamma/2)\,(x^2 - y^2)$*. This figure is based on Figure 2.4 of Hoskins and James (2014).*

vorticity, $\zeta = 0$, so that the area remains constant and the orientation is fixed. However, it has shear that acts to deform the area. More precisely, the non-divergent deformational flow as determined by

$$\boldsymbol{u} = \hat{\boldsymbol{z}} \wedge \nabla\psi, \tag{13.56}$$

with the streamfunction, $\psi = -\gamma\, x\, y$ 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:

$$\text{tension strain} = S_{\mathrm{T}} = \mathbb{S}_{11} - \mathbb{S}_{22} = \frac{\partial u}{\partial x} - \frac{\partial v}{\partial y} \tag{13.57a}$$

$$\text{shearing strain} = S_{\mathrm{S}} = 2\,\mathbb{S}_{12} = \frac{\partial u}{\partial y} + \frac{\partial v}{\partial x}. \tag{13.57b}$$

In the literature, 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 $\psi = -\gamma\, x\, y$, we have

$$S_{\mathrm{T}} = 2\,\gamma \qquad S_{\mathrm{S}} = 0, \tag{13.58}$$

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

$$\psi = -(\gamma/2)\,(x^2 - y^2) \qquad u = -\gamma\, y \qquad v = -\gamma\, x \qquad S_{\mathrm{T}} = 0 \qquad S_{\mathrm{S}} = -2\,\gamma, \tag{13.59}$$

as depicted by the right panel of Figure 13.6. This pure shearing flow leads to the deformation of the fluid square shown in the lower right panel of Figure 13.5.

Figure 13.6: *Two-dimensional non-divergent and irrotational flow with nonzero deformation/strain. Left panel: pure tension strain as determined by the streamfunction, $\psi = -\gamma\, x\, y$, 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.*

13.5.4 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).

13.6 Exercises

EXERCISE 13.1: 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

$$\boldsymbol{v} = \boldsymbol{U} + \boldsymbol{\Omega} \wedge \boldsymbol{x}, \tag{13.60}$$

where $\boldsymbol{\Omega}$ is a constant angular rotation rate and $\boldsymbol{U}$ is a constant velocity. That is, a fluid velocity equal to a constant rotation plus translation renders zero strain. Hint: if $\mathbb{S}_{mn} = 0$, what does that imply about the velocity field? You may also wish to make use of the general decomposition (13.35).

EXERCISE 13.2: EVOLUTION OF THE JACOBIAN USING ϵ-TENSOR GYMNASTICS
There is another way to derive the identity (13.52) 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

$$\frac{\partial \boldsymbol{X}}{\partial \boldsymbol{a}} = \frac{1}{3!}\, \epsilon_{mnp}\, \epsilon_{ijk}\, \frac{\partial X_m}{\partial a_i}\, \frac{\partial X_n}{\partial a_j}\, \frac{\partial X_p}{\partial a_k}. \tag{13.61}$$

Take the material derivative of this expression and show that we get the same expression as equation (13.52). Hint: make use of the identity

$$\frac{\mathrm{D}}{\mathrm{D}t}\, \frac{\partial X_m}{\partial a_i} = \frac{\partial V_m}{\partial a_i}, \tag{13.62}$$

which holds since the material time derivative is taken with the material coordinates, $\boldsymbol{a}$, held fixed.

14

Mass conservation

We assume that mass is neither created nor destroyed within the fluid domain. This assumption holds independently of the forces acting on the fluid, and so it forms 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 15. Spatial positions and trajectories are represented using Cartesian coordinates to simplify the maths, with generalizations following the tensor analysis methods of Chapter 6. We build on our understanding of the Eulerian and Lagrangian kinematic descriptions developed in Chapter 12. There are many facets to mass conservation presented in this chapter, some presented for pedagogical reasons and others presented since they appear in many places later in this book.

14.1 Material fluid parcels . . . **232**
14.1.1 Lagrangian expression for mass conservation . . . 232
14.1.2 Alternative derivation based on the Jacobian . . . 232
14.1.3 Summary of parcel kinematic equations . . . 233
14.2 Eulerian fluid regions . . . **233**
14.2.1 Differential expression . . . 233
14.2.2 Finite volume expression . . . 234
14.2.3 Arbitrary Eulerian region . . . 235
14.3 Material fluid regions . . . **235**
14.3.1 Evolution of volume . . . 236
14.3.2 Mass conservation . . . 236
14.3.3 Mass conservation using Lagrangian methods . . . 237
14.3.4 Reynolds transport theorem . . . 238
14.3.5 Comments on notation . . . 239
14.4 Kinematic boundary conditions . . . **239**
14.4.1 Static material surface . . . 239
14.4.2 Moving material surface . . . 240
14.4.3 Dynamic and permeable surface . . . 244
14.5 Mass budget for a column of ocean fluid . . . **246**
14.6 Exercises . . . **248**

14.1 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.

14.1.1 Lagrangian expression for mass conservation

The mass of an infinitesimal fluid parcel is written

$$\delta M = \rho\, \delta V, \tag{14.1}$$

where δV is the volume and

$$\rho = \frac{\delta M}{\delta V} \tag{14.2}$$

is the mass density. By definition, the fluid parcel has a constant mass, so that its material time derivative vanishes

$$\frac{\mathrm{D}(\delta M)}{\mathrm{D}t} = 0. \tag{14.3}$$

Equation (14.3) 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

$$\frac{\mathrm{D}(\delta M)}{\mathrm{D}t} = \frac{\mathrm{D}\left(\rho\, \delta V\right)}{\mathrm{D}t} \tag{14.4a}$$

$$= \delta M \left[\frac{1}{\rho} \frac{\mathrm{D}\rho}{\mathrm{D}t} + \frac{1}{\delta V} \frac{\mathrm{D}(\delta V)}{\mathrm{D}t} \right]. \tag{14.4b}$$

Making use of equation (14.23) derived in Section 14.3.1 for the material evolution of the parcel volume then leads to

$$\frac{1}{\delta M} \frac{\mathrm{D}(\delta M)}{\mathrm{D}t} = \frac{1}{\rho} \frac{\mathrm{D}\rho}{\mathrm{D}t} + \nabla \cdot \boldsymbol{v}. \tag{14.5}$$

Setting $\mathrm{D}(\delta M)/\mathrm{D}t = 0$ then leads to the mass conservation equation, also known as the continuity equation

$$\frac{1}{\rho} \frac{\mathrm{D}\rho}{\mathrm{D}t} = -\nabla \cdot \boldsymbol{v}. \tag{14.6}$$

The parcel volume contracts in regions where the velocity converges (we prove that property in Section 14.3.1). As seen by the mass continuity equation (14.6), such regions are also where the parcel density increases. The opposite occurs for regions where the velocity diverges.

14.1.2 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 14.3.3, and further

discuss it for fluid parcels in equation (13.51c). Anticipating those results here allows us to write the material evolution for the fluid parcel mass in the form

$$\frac{\mathrm{D}}{\mathrm{D}t}\left[\rho\,\delta V(\boldsymbol{x})\right] = \frac{\mathrm{D}}{\mathrm{D}t}\left[\rho\,\frac{\partial \boldsymbol{X}}{\partial \boldsymbol{a}}\,\delta V(\boldsymbol{a})\right] \tag{14.7a}$$

$$= \left[\frac{\mathrm{D}\rho}{\mathrm{D}t} + \rho\,\nabla\cdot\boldsymbol{v}\right]\left[\frac{\partial \boldsymbol{X}}{\partial \boldsymbol{a}}\right]\delta V(\boldsymbol{a}) \tag{14.7b}$$

$$= \left[\frac{\mathrm{D}\rho}{\mathrm{D}t} + \rho\,\nabla\cdot\boldsymbol{v}\right]\delta V(\boldsymbol{x}). \tag{14.7c}$$

We recover the mass conservation equation (14.6) when noting that the mass of a material parcel is constant.

14.1.3 Summary of parcel kinematic equations

Let us now summarize the variety of differential evolution equations for mass, volume, and density as viewed from a material reference frame

$$\frac{\mathrm{D}(\delta M)}{\mathrm{D}t} = 0 \qquad \text{parcel mass} \tag{14.8a}$$

$$\frac{1}{\delta V}\frac{\mathrm{D}(\delta V)}{\mathrm{D}t} = \nabla\cdot\boldsymbol{v} \qquad \text{parcel volume} \tag{14.8b}$$

$$\frac{1}{\rho}\frac{\mathrm{D}\rho}{\mathrm{D}t} = -\nabla\cdot\boldsymbol{v} \qquad \text{parcel density.} \tag{14.8c}$$

Note that the parcel volume equation is derived in Section 14.3.1 below. To help remember the proper signs, recall that as the fluid diverges from a point ($\nabla\cdot\boldsymbol{v} > 0$), it expands the boundaries of the material parcel and so increases the parcel volume as per equation (14.8b). This process in turn causes the material parcel density to decrease ($-\nabla\cdot\boldsymbol{v} < 0$) as per equation (14.8c).

14.2 Eulerian fluid regions

We now 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.

14.2.1 Differential expression

The Eulerian form of mass continuity results from introducing the Eulerian expression for the material time derivative operator (equation (12.28)) into the mass continuity equation (14.6). The resulting *flux-form* Eulerian mass continuity equation is

$$\frac{\partial\rho}{\partial t} + \nabla\cdot(\rho\,\boldsymbol{v}) = 0. \tag{14.9}$$

This equation 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

$$\frac{\partial\rho}{\partial t} = -\nabla\cdot(\rho\,\boldsymbol{v}). \tag{14.10}$$

The flux

$$\text{mass flux} = \rho\,\boldsymbol{v} \qquad \text{dimensions } M\ L^{-2}\ T^{-1} \tag{14.11}$$

Figure 14.1: *A finite sized cube (also known as a cell) of fixed dimensions and position (an Eulerian cube) 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.*

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.

14.2.2 Finite volume expression

Now consider a finite sized cube region that is fixed in space; i.e., an Eulerian region such as shown in Figure 14.1. The mass contained within the cube is given by

$$\Delta M = \rho\,\Delta V = \rho\,\Delta x\,\Delta y\,\Delta z, \tag{14.12}$$

where the cube volume, $\Delta V = \Delta x\,\Delta y\,\Delta z$, 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 14.1, the accumulation of mass within the cube through this transport is determined by the difference in mass transport crossing the two adjacent cell faces

$$\text{mass change from meridional transport} = (\Delta x\,\Delta z)\,[(v\,\rho)_{y-\Delta y/2} - (v\,\rho)_{y+\Delta y/2}]. \tag{14.13}$$

Expanding the difference into a Taylor series leads to

$$\text{mass change from meridional transport} \approx -(\Delta x\,\Delta y\,\Delta z)\,\frac{\partial(v\,\rho)}{\partial y}. \tag{14.14}$$

The same analysis for the zonal and vertical directions leads to the mass budget for the cube

$$\frac{\partial(\rho\,\Delta V)}{\partial t} = -\Delta V\left[\frac{\partial(u\,\rho)}{\partial x} + \frac{\partial(v\,\rho)}{\partial y} + \frac{\partial(w\,\rho)}{\partial z}\right]. \tag{14.15}$$

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 continuity equation (14.9), $\partial\rho/\partial t = -\nabla\cdot(\rho\,\boldsymbol{v})$.

Figure 14.2: *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, $\mathrm{d}\mathcal{S}$, is oriented according to the outward normal, $\hat{\boldsymbol{n}}$.*

14.2.3 Arbitrary Eulerian region

The previous discussion for a cube can be generalized by making use of the Divergence theorem (Section 2.7). For that purpose, consider an arbitrary static and simply closed region within the fluid such as in Figure 14.2. Integrating the continuity equation (14.9) over that region leads to

$$\int_{\mathcal{R}} \frac{\partial \rho}{\partial t}\,\mathrm{d}V = -\int_{\mathcal{R}} \nabla\cdot(\rho\,\boldsymbol{v})\,\mathrm{d}V. \tag{14.16}$$

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

$$\frac{\mathrm{d}}{\mathrm{d}t}\int_{\mathcal{R}} \rho\,\mathrm{d}V = -\oint_{\partial\mathcal{R}} \rho\,\boldsymbol{v}\cdot\hat{\boldsymbol{n}}\,\mathrm{d}\mathcal{S}, \tag{14.17}$$

where $\hat{\boldsymbol{n}}$ is the outward normal vector along the closed boundary of the region, and $\mathrm{d}\mathcal{S}$ 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.

14.3 Material fluid regions

We now extend the kinematics of material fluid parcels to finite sized material fluid regions. 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 14.2. 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 14.3.4.

14.3.1 Evolution of volume

Consider a finite material region, $\mathcal{R}(\boldsymbol{v})$, whose volume is given by the integral

$$V = \int_{\mathcal{R}(\boldsymbol{v})} \mathrm{d}V, \tag{14.18}$$

with $\mathrm{d}V$ 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

$$\mathcal{R}(\boldsymbol{v}) = \text{region following flow}, \tag{14.19}$$

with the $\boldsymbol{v}$ 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

$$\frac{\mathrm{d}}{\mathrm{d}t} \int_{\mathcal{R}(\boldsymbol{v})} \mathrm{d}V = \oint_{\partial\mathcal{R}(\boldsymbol{v})} \boldsymbol{v} \cdot \hat{\boldsymbol{n}} \, \mathrm{d}\mathcal{S}, \tag{14.20}$$

where $\hat{\boldsymbol{n}}$ is the outward normal on the region's closed boundary, $\mathrm{d}\mathcal{S}$ is the area element on the boundary, and

$$\boldsymbol{v} \cdot \hat{\boldsymbol{n}} \, \mathrm{d}\mathcal{S} = \text{volume transport (volume per time) at the boundary } \partial\mathcal{R}. \tag{14.21}$$

Use of Gauss's divergence theorem then leads to the equivalent expression

$$\frac{\mathrm{d}}{\mathrm{d}t} \int_{\mathcal{R}(\boldsymbol{v})} \mathrm{d}V = \int_{\mathcal{R}(\boldsymbol{v})} \nabla \cdot \boldsymbol{v} \, \mathrm{d}V. \tag{14.22}$$

We now take the limit as the material region becomes a material parcel, in which case we recover the differential expression

$$\frac{1}{\delta V} \frac{\mathrm{D}(\delta V)}{\mathrm{D}t} = \nabla \cdot \boldsymbol{v}, \tag{14.23}$$

where we make use of $\mathrm{D}/\mathrm{D}t$ since the infinitesimal volume is moving with the fluid. This equation is also derived in Section 13.4.1 using different methods.

14.3.2 Mass conservation

The mass of fluid contained in a finite material region is given by

$$M = \int_{\mathcal{R}(\boldsymbol{v})} \rho \, \mathrm{d}V. \tag{14.24}$$

As a material fluid region, it maintains a constant mass as it moves through the fluid

$$\frac{\mathrm{d}}{\mathrm{d}t} \int_{\mathcal{R}(\boldsymbol{v})} \rho \, \mathrm{d}V = 0. \tag{14.25}$$

Just as for the volume in Section 14.3.1, taking the limit as the material region becomes infinitesimally small, the region mass conservation statement (14.25) becomes the parcel mass conservation statement (14.3)

$$\frac{\mathrm{D}(\delta M)}{\mathrm{D}t} = \frac{\mathrm{D}(\rho \, \delta V)}{\mathrm{D}t} = 0. \tag{14.26}$$

14.3.3 Mass conservation using Lagrangian methods

Rather than take the limit as the finite material region $\mathcal{R}(\boldsymbol{v})$ 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 (14.25). As part of this discussion we introduce two coordinate representations of the mass within a material region. The first is the Cartesian $\boldsymbol{x}$-space representation

$$\int_{\mathcal{R}} \rho \, \mathrm{d}V = \int_{x_1(t)}^{x_2(t)} \int_{y_1(t)}^{y_2(t)} \int_{z_1(t)}^{z_2(t)} \rho \, \mathrm{d}x \, \mathrm{d}y \, \mathrm{d}z \equiv \int_{\mathcal{R}[\boldsymbol{v}(\boldsymbol{x},t)]} \rho \, \mathrm{d}V(\boldsymbol{x}). \tag{14.27}$$

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 makes us of a notation that emphasizes the region bounds and volume element are moving with the fluid velocity, $\boldsymbol{v}$, as specified using the $\boldsymbol{x}$-space representation, $\boldsymbol{v}(\boldsymbol{x}, t)$.

The second representation of the mass in a material region makes use of $\boldsymbol{a}$-space material coordinates. For this representation we perform a coordinate transformation from the $\boldsymbol{x}$-space representation to the $\boldsymbol{a}$-space representation, with this coordinate transformation necessitating the Jacobian of transformation. To capture the gist of this transformation let us consider the one-dimensional case in which

$$\int_{x_1(t)}^{x_2(t)} \rho \, \mathrm{d}x = \int_{a[x_1(t)]}^{a[x_2(t)]} \rho \, \frac{\partial X}{\partial a} \, \mathrm{d}a = \int_{a_1}^{a_2} \rho \, \frac{\partial X}{\partial a} \, \mathrm{d}a. \tag{14.28}$$

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 one-to-one 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

$$\int_{\mathcal{R}[\boldsymbol{v}(\boldsymbol{x},t)]} \rho \, \mathrm{d}V(\boldsymbol{x}) = \int_{\mathcal{R}[\boldsymbol{v}(\boldsymbol{a})]} \rho \, \frac{\partial \boldsymbol{X}}{\partial \boldsymbol{a}} \, \mathrm{d}V(\boldsymbol{a}). \tag{14.29}$$

Again, we distinguish the $\boldsymbol{x}$ and $\boldsymbol{a}$ arguments for the material region, with the $\boldsymbol{x}$ representation also requiring a time argument whereas the material $\boldsymbol{a}$ representation has no time dependence.

We now make use of the equality (14.29) 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

$$\frac{\mathrm{d}}{\mathrm{d}t} \left[\int_{\mathcal{R}[\boldsymbol{v}(\boldsymbol{x},t)]} \rho \, \mathrm{d}V(\boldsymbol{x}) \right] = \frac{\mathrm{d}}{\mathrm{d}t} \left[\int_{\mathcal{R}[\boldsymbol{v}(\boldsymbol{a})]} \rho \, \frac{\partial \boldsymbol{X}}{\partial \boldsymbol{a}} \, \mathrm{d}V(\boldsymbol{a}) \right] \tag{14.30a}$$

$$= \int_{\mathcal{R}[\boldsymbol{v}(\boldsymbol{a})]} \frac{\mathrm{D}}{\mathrm{D}t} \left[\rho \, \frac{\partial \boldsymbol{X}}{\partial \boldsymbol{a}} \mathrm{d}V(\boldsymbol{a}) \right] \tag{14.30b}$$

$$= \int_{\mathcal{R}[\boldsymbol{v}(\boldsymbol{x},t)]} \left[\frac{\mathrm{D}\rho}{\mathrm{D}t} + \rho \, \nabla \cdot \boldsymbol{v} \right] \mathrm{d}V(\boldsymbol{x}) \tag{14.30c}$$

$$= \int_{\mathcal{R}[\boldsymbol{v}(\boldsymbol{x},t)]} \left[\frac{1}{\rho} \frac{\mathrm{D}\rho}{\mathrm{D}t} + \nabla \cdot \boldsymbol{v} \right] \rho \, \mathrm{d}V(\boldsymbol{x}). \tag{14.30d}$$

When expressing the integral bounds using $\boldsymbol{a}$-space coordinates, the integral bounds have no time dependence, thus allowing us to 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, $\mathrm{D}/\mathrm{D}t$, since it is a time derivative computed by following fluid particles. The third equality made use of equation (14.7c) and converted back to $\boldsymbol{x}$-space. As the material region $\mathcal{R}$ has a materially constant mass, we recover the mass continuity equation (14.6) by setting the integrand in equation (14.30d) to zero.

14.3.4 Reynolds transport theorem

On first encounter, the method from Section 14.3.3 that involve moving between Eulerian ($\boldsymbol{x}$-space) and Lagrangian ($\boldsymbol{a}$-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. Indeed, the method is formalized by the *Reynolds transport theorem* discussed in this section.

Manipulations leading to the mass conservation statement (14.30d) can be generalized by considering the material time derivative of a mass-weighted field ψ (e.g., a tracer concentration as in Section 15.1)

$$\frac{\mathrm{D}(\psi\,\rho\,\delta V)}{\mathrm{D}t} = \frac{\mathrm{D}\psi}{\mathrm{D}t}\,\rho\,\delta V + \psi\,\frac{\mathrm{D}(\rho\,\delta V)}{\mathrm{D}t} \tag{14.31a}$$

$$= \rho\,\delta V\left[\frac{\mathrm{D}\psi}{\mathrm{D}t} + \frac{\psi}{\rho}\frac{\mathrm{D}\rho}{\mathrm{D}t} + \psi\,\nabla\cdot\boldsymbol{v}\right] \tag{14.31b}$$

$$= \delta V\left[\frac{\partial(\rho\,\psi)}{\partial t} + \nabla\cdot(\rho\,\psi\,\boldsymbol{v})\right]. \tag{14.31c}$$

The first equality used the product rule, which holds for material time derivatives. Mass conservation means that the material derivative $\mathrm{D}(\rho\,\delta V)/\mathrm{D}t = 0$. However, we choose to write mass conservation in the form of equation (14.7c), which allows us to introduce the flux-form Eulerian expression after replacing the material time derivative with its Eulerian form from equation (12.28). Another, perhaps more direct means, to derive this result is to write

$$\rho\,\frac{\mathrm{D}\psi}{\mathrm{D}t} = \rho\left[\frac{\partial\psi}{\partial t} + \boldsymbol{v}\cdot\nabla\psi\right] \tag{14.32a}$$

$$= \rho\left[\frac{\partial\psi}{\partial t} + \boldsymbol{v}\cdot\nabla\psi\right] + \psi\left[\frac{\partial\rho}{\partial t} + \nabla\cdot(\boldsymbol{v}\,\rho)\right] \tag{14.32b}$$

$$= \frac{\partial(\rho\,\psi)}{\partial t} + \nabla\cdot(\rho\,\boldsymbol{v}\,\psi). \tag{14.32c}$$

Following the discussion in Section 14.3.3, we can extend the material parcel result (14.31c) 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*

$$\frac{\mathrm{d}}{\mathrm{d}t}\int_{\mathcal{R}(\boldsymbol{v})}\psi\,\rho\,\mathrm{d}V = \int_{\mathcal{R}(\boldsymbol{v})}\frac{\mathrm{D}\psi}{\mathrm{D}t}\rho\,\mathrm{d}V \tag{14.33a}$$

$$= \int_{\mathcal{R}(\boldsymbol{v})}\left[\frac{\partial(\rho\,\psi)}{\partial t} + \nabla\cdot(\rho\,\psi\,\boldsymbol{v})\right]\mathrm{d}V \tag{14.33b}$$

$$= \int_{\mathcal{R}(\boldsymbol{v})}\frac{\partial(\rho\,\psi)}{\partial t}\,\mathrm{d}V + \oint_{\partial\mathcal{R}(\boldsymbol{v})}\rho\,\psi\,\boldsymbol{v}\cdot\hat{\boldsymbol{n}}\,\mathrm{d}\mathcal{S}, \tag{14.33c}$$

where the final equality follows from Gauss's divergence theorem. Additionally, we returned to the notation $\mathcal{R}(\boldsymbol{v})$ for material region as introduced in Section 14.3.1. This notation is sufficient to designate that the region is following fluid parcels as per the fluid velocity field, $\boldsymbol{v}$. Note that for the surface integral term, $\boldsymbol{v} \cdot \hat{\boldsymbol{n}}$ generally does not vanish. Rather, it is given by $\boldsymbol{v} \cdot \hat{\boldsymbol{n}} = \boldsymbol{v}^{(s)} \cdot \hat{\boldsymbol{n}}$, where $\boldsymbol{v}^{(s)}$ is the velocity of a point on the boundary of the material region. We detail this result in Section 14.4.2 when discussing the kinematics of a moving material surface.

14.3.5 Comments on notation

In this book we write $\mathrm{d}/\mathrm{d}t$ 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, $\mathcal{R}(\boldsymbol{v})$. This notation designates that all points in the domain, $\mathcal{R}$, move with the fluid velocity, $\boldsymbol{v}$. However, many authors choose an alternative notation by using the material time derivative, $\mathrm{D}/\mathrm{D}t$, when acting on an integral over a material region. We thus have the following equality across the two notational conventions

$$\frac{\mathrm{D}}{\mathrm{D}t} \int_{\mathcal{R}} \psi \, \rho \, \mathrm{d}V = \frac{\mathrm{d}}{\mathrm{d}t} \int_{\mathcal{R}(\boldsymbol{v})} \psi \, \rho \, \mathrm{d}V. \tag{14.34}$$

The use of one convention versus the other is a matter of taste. We follow Section 2.1 of *Batchelor* (1967) by restricting the $\mathrm{D}/\mathrm{D}t$ operator to act only on space-time fields, such as $\psi(\boldsymbol{x}, t)$. Hence, the $\mathrm{D}/\mathrm{D}t$ operator is not used when acting on integrals over spatial regions. Following this convention leads us to write $\mathcal{R}(\boldsymbol{v})$ for a region that moves with the fluid flow and to retain $\mathrm{d}/\mathrm{d}t$ when acting on the integral over that region.

Admittedly, the $\mathcal{R}(\boldsymbol{v})$ notation is rarely used in the literature, with authors generally dropping the $\boldsymbol{v}$ 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, $\mathcal{R}(\boldsymbol{v})$, to denote that a material region is being considered. This usage hopefully helps the reader follow the mathematical flow rather than swimming in a turbulent sea driven by confused notation.

14.4 Kinematic boundary conditions

When a fluid encounters a boundary, either at the edge of the fluid region or within the fluid itself, the fluid must accommodate the boundary. Alternatively, the boundary must accomodate the fluid. 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.

14.4.1 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 14.3)

$$\boldsymbol{v} \cdot \hat{\boldsymbol{n}} = 0 \qquad \text{no-normal flow condition on static material boundary.} \tag{14.35}$$

Figure 14.3: *Illustrating the no-normal flow boundary condition maintained for a solid boundary, on which $\boldsymbol{v} \cdot \hat{\boldsymbol{n}} = 0$ (equation (14.35)). 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 (14.36)).*

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 12.8.2, where $\boldsymbol{v} \cdot \hat{\boldsymbol{n}} = 0$ 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 (14.35). Doing so helps to develop a geometric formalism especially useful for the more complicated moving boundary conditions in Sections 14.4.2 and 14.4.3. For this purpose, introduce a coordinate expression for the boundary according to

$$s(x, y, z) = z - \eta_\mathrm{b}(x, y) = 0 \qquad \text{static material boundary,} \tag{14.36}$$

with $z = \eta_\mathrm{b}(x, y)$ the vertical position of the boundary. The outward normal vector at the boundary is thus given by

$$\hat{\boldsymbol{n}} = -\frac{\nabla s}{|\nabla s|} = -\frac{\nabla(z - \eta_\mathrm{b})}{|\nabla(z - \eta_\mathrm{b})|} = -\frac{\hat{\boldsymbol{z}} - \nabla\eta_\mathrm{b}}{\sqrt{1 + (\nabla\eta_\mathrm{b})^2}}. \tag{14.37}$$

Consequently, the no-flux boundary condition (14.35) takes the form

$$w - \boldsymbol{u} \cdot \nabla\eta_\mathrm{b} = 0 \qquad \text{at } z = \eta_\mathrm{b}(x, y), \tag{14.38}$$

where the velocity is decomposed into its horizontal and vertical components, $\boldsymbol{v} = (\boldsymbol{u}, w)$. 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 $\nabla\eta_\mathrm{b} = 0$, then the kinematic boundary condition reduces to $w = 0$.

14.4.2 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. However, as discussed in Chapter 15, 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.

Figure 14.4: *Illustrating the boundary condition for a moving material surface, on which $\boldsymbol{v}\cdot\hat{\boldsymbol{n}} = \boldsymbol{v}^{(s)}\cdot\hat{\boldsymbol{n}}$ (equation (14.39)). 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 (14.40)), in which case the normal direction is given by $\hat{\boldsymbol{n}} = |\nabla s|^{-1}\nabla s$ (equation (14.41)).*

General expression of the kinematic boundary condition

To ensure no flow crosses the material surface, the surface must have a velocity that matches that of the fluid. More precisely, the normal component of the velocity for a point stuck to the surface at a point must match the normal component of the fluid at that point on the surface. We are thus led to the kinematic boundary condition for a moving material surface

$$(\boldsymbol{v} - \boldsymbol{v}^{(s)})\cdot\hat{\boldsymbol{n}} = 0 \qquad \text{moving material boundary condition.} \tag{14.39}$$

We illustrate this boundary condition in Figure 14.4, where $\boldsymbol{v}^{(s)}$ is the velocity of a point fixed on the moving material surface and $\boldsymbol{v}$ is the velocity of the fluid particles.

The boundary condition (14.39) does not mean $\boldsymbol{v}$ and $\boldsymbol{v}^{(s)}$ are identical. It only says that their normal components are the same when evaluated on the material surface. As a Corollary, we see that $\boldsymbol{v}\cdot\hat{\boldsymbol{n}}$ is not generally zero so that a moving material boundary does *not* coincide with a flow streamline (see discussion in Sections 12.8.2 and 12.8.3).

Specialized expression of the boundary condition

Now specialize the kinematic condition (14.39) to the case of a material surface specified by a function that takes on a constant value on the surface

$$s(\boldsymbol{x},t) = s_0. \tag{14.40}$$

Correspondingly, the surface normal vector is given by

$$\hat{\boldsymbol{n}} = |\nabla s|^{-1}\nabla s. \tag{14.41}$$

From Section 12.5.6, we know that a point fixed on an arbitrary surface has a velocity that satisfies (see equation (12.41))

$$\frac{\partial s}{\partial t} + \boldsymbol{v}^{(s)}\cdot\nabla s = 0 \qquad \text{on an iso-surface } s(\boldsymbol{x},t) = s_0. \tag{14.42}$$

Use of the identity

$$\frac{\partial s}{\partial t} = \frac{\mathrm{D}s}{\mathrm{D}t} - \boldsymbol{v}\cdot\nabla s \tag{14.43}$$

renders

$$\frac{\mathrm{D}s}{\mathrm{D}t} - \boldsymbol{v}\cdot\nabla s + \boldsymbol{v}^{(s)}\cdot\nabla s = \frac{\mathrm{D}s}{\mathrm{D}t} + (\boldsymbol{v}^{(s)} - \boldsymbol{v})\cdot\nabla s = 0. \tag{14.44a}$$

Since $(\boldsymbol{v}^{(s)}-\boldsymbol{v})\cdot\nabla s = 0$ from the boundary condition (14.39), we are left with the material constancy condition

$$\frac{\mathrm{D}s}{\mathrm{D}t} = \frac{\partial s}{\partial t} + \boldsymbol{v}\cdot\nabla s = 0 \qquad \text{on material surface } s(\boldsymbol{x},t) = s_0. \tag{14.45}$$

Consequently, no net 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

The expression (14.45) of the kinematic boundary condition is quite useful for many applications. For example, 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

$$s(x,y,z,t) = z - \eta(x,y,t) = s_0. \tag{14.46}$$

The function $\eta(x,y,t)$ is the vertical deviation of the interface relative to the horizontal. The kinematic boundary condition (14.45) thus takes the form

$$\frac{\mathrm{D}s}{\mathrm{D}t} = \frac{\mathrm{D}(z-\eta)}{\mathrm{D}t} = 0. \tag{14.47}$$

Hence, the vertical velocity component at the interface equals to the material time derivative of the interface displacement

$$\frac{\mathrm{D}z}{\mathrm{D}t} = \frac{\mathrm{D}\eta}{\mathrm{D}t} \Rightarrow w = \frac{\partial\eta}{\partial t} + \boldsymbol{u}\cdot\nabla\eta \qquad \text{material b.c. at interface } z = \eta(x,y,t). \tag{14.48}$$

This boundary condition can be equivalently written in the form

$$\boldsymbol{v}\cdot\hat{\boldsymbol{n}} = \frac{\partial\eta/\partial t}{\sqrt{1+|\nabla\eta|^2}}, \tag{14.49}$$

where

$$\hat{\boldsymbol{n}} = \frac{\nabla\left(z-\eta\right)}{|\nabla\left(z-\eta\right)|} = \frac{-\nabla\eta + \hat{\boldsymbol{z}}}{\sqrt{1+|\nabla\eta|^2}} \tag{14.50}$$

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 $\mathrm{D}s/\mathrm{D}t = 0$ is a key kinematic result. We thus offer yet another derivation to help solidify its meaning. As before, define the surface according to

$$s(\boldsymbol{x},t) = z - \eta(x,y,t) = 0, \tag{14.51}$$

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 14.5). The vertical position of the surface at the new time is determined by

$$s(\boldsymbol{x}+\delta\boldsymbol{x}, t+\delta t) = 0, \tag{14.52}$$

Figure 14.5: *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 $\boldsymbol{X}(t)$ and $\boldsymbol{X}(t+\delta t)$. The equation $s(\boldsymbol{x},t)=z-\eta(x,y,t)=0$ 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 $s(\boldsymbol{x},t)=s(\boldsymbol{x}+\delta\boldsymbol{x},t+\delta t)=0$.*

where $\boldsymbol{X}(t+\delta t)=\boldsymbol{x}+\delta\boldsymbol{x}$ is the displaced position of a point on the surface that started at $\boldsymbol{X}(t)=\boldsymbol{x}$. Expanding equation (14.52) in a Taylor series to leading order yields

$$s(\boldsymbol{x},t)+\delta\boldsymbol{x}\cdot\nabla s+\delta t\,\partial_t s=0. \tag{14.53}$$

Since $s(\boldsymbol{x},t)=0$ from equation (14.51) we thus have

$$\frac{\partial s}{\partial t}+\frac{\delta\boldsymbol{x}}{\delta t}\cdot\nabla s=\frac{\partial s}{\partial t}+\frac{\delta\boldsymbol{x}}{\delta t}\cdot\hat{\boldsymbol{n}}\,|\nabla s|=0, \tag{14.54}$$

where $\hat{\boldsymbol{n}}=|\nabla s|^{-1}\,\nabla s$ 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 $s(\boldsymbol{x},t)$

$$\frac{\partial s}{\partial t}=-\frac{\delta\boldsymbol{x}}{\delta t}\cdot\hat{\boldsymbol{n}}\,|\nabla s|. \tag{14.55}$$

In contrast, any tangential displacement along an s-isosurface leaves $s(\boldsymbol{x},t)$ 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, $\hat{\boldsymbol{n}}\,(\boldsymbol{v}^{(s)}\cdot\hat{\boldsymbol{n}})$. 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

$$\delta\boldsymbol{x}=\delta t\,\hat{\boldsymbol{n}}\,(\boldsymbol{v}^{(s)}\cdot\hat{\boldsymbol{n}}). \tag{14.56}$$

Making use of this result in equation (14.54) leads to the differential equation satisfied by a point fixed on the moving surface

$$\frac{\partial s}{\partial t}+\boldsymbol{v}^{(s)}\cdot\nabla s=0, \tag{14.57}$$

where we used the identity

$$\frac{\delta\boldsymbol{x}}{\delta t}\cdot\nabla s=\hat{\boldsymbol{n}}\,(\boldsymbol{v}^{(s)}\cdot\hat{\boldsymbol{n}})\cdot\nabla s=\boldsymbol{v}^{(s)}\cdot\nabla s. \tag{14.58}$$

Assuming the surface is material means that

$$\boldsymbol{v}^{(s)}\cdot\hat{\boldsymbol{n}}=\boldsymbol{v}\cdot\hat{\boldsymbol{n}}, \tag{14.59}$$

Figure 14.6: *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* $\rho\,(\boldsymbol{v} - \boldsymbol{v}^{(s)}) \cdot \hat{\boldsymbol{n}}\,\mathrm{d} = -\mathcal{Q}_m$ *(equation (14.60)). In the special case of an ocean free surface with no overturns, this boundary condition reduces to the surface kinematic boundary condition (14.73).*

so that motion of the surface normal to itself is identical to that of the fluid in the same direction. The identity in equation (14.57) renders the material invariance condition $\mathrm{D}s/\mathrm{D}t = 0$.

14.4.3 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

$$\rho\,(\boldsymbol{v} - \boldsymbol{v}^{(s)}) \cdot \hat{\boldsymbol{n}}\,\mathrm{d}\mathcal{S} = -\mathcal{Q}_{\mathrm{m}}\,\mathrm{d}\mathcal{S} \qquad \text{moving non-material boundary condition.} \tag{14.60}$$

In this equation, $\mathrm{d}\mathcal{S}$ is an infinitesimal area element on the surface, and $\mathcal{Q}_{\mathrm{m}}$ is the mass per time per surface area 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 (14.42) for $\boldsymbol{v}^{(s)} \cdot \hat{\boldsymbol{n}}$ holds for a point on an arbitrary surface so that

$$\boldsymbol{v}^{(s)} \cdot \hat{\boldsymbol{n}} = -\frac{\partial s/\partial t}{|\nabla s|}. \tag{14.61}$$

Furthermore, the projection of the fluid velocity onto the normal direction can be written

$$\frac{\mathrm{D}s}{\mathrm{D}t} = \frac{\partial s}{\partial t} + \boldsymbol{v} \cdot \nabla s \Longrightarrow \boldsymbol{v} \cdot \hat{\boldsymbol{n}} = \frac{1}{|\nabla s|}\left(\frac{\mathrm{D}s}{\mathrm{D}t} - \frac{\partial s}{\partial t}\right). \tag{14.62}$$

Bringing these results together leads to

$$\rho\,(\boldsymbol{v} - \boldsymbol{v}^{(s)}) \cdot \hat{\boldsymbol{n}}\,\mathrm{d}\mathcal{S} = \frac{\rho\,\mathrm{d}\mathcal{S}}{|\nabla s|}\,\frac{\mathrm{D}s}{\mathrm{D}t}. \tag{14.63}$$

This equation says that the net 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 net transport across the surface (see discussion in Section 14.4.2).

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

$$\frac{\partial s}{\partial z} \neq 0, \tag{14.64}$$

so that we can further massage the boundary condition (14.63) by writing the area factor in the form

$$\frac{\mathrm{d}\mathcal{S}}{|\nabla s|} = \frac{\mathrm{d}\mathcal{S}}{\sqrt{(\partial s/\partial x)^2 + (\partial s/\partial y)^2 + (\partial s/\partial z)^2}} \tag{14.65a}$$

$$= \frac{\mathrm{d}\mathcal{S}}{|\partial s/\partial z| \sqrt{[(\partial s/\partial x)/(\partial s/\partial z)]^2 + [(\partial s/\partial y)/(\partial s/\partial z)]^2 + 1}} \tag{14.65b}$$

$$= \frac{\mathrm{d}\mathcal{S}}{|\partial s/\partial z| \sqrt{1 + \tan^2 \vartheta}} \tag{14.65c}$$

$$= \left|\frac{\partial z}{\partial s}\right| |\cos \vartheta| \, \mathrm{d}\mathcal{S} \tag{14.65d}$$

$$= \left|\frac{\partial z}{\partial s}\right| \mathrm{d}A. \tag{14.65e}$$

The equality (14.65c) introduced the angle, ϑ, between the boundary surface and the horizontal plane. The squared slope of this surface given by

$$\tan^2 \vartheta = \frac{\nabla_z s \cdot \nabla_z s}{(\partial s/\partial z)^2} = \nabla_s z \cdot \nabla_s z \tag{14.66}$$

with

$$\nabla_z = \hat{\boldsymbol{x}} \left[\frac{\partial}{\partial x}\right]_{y,z} + \hat{\boldsymbol{y}} \left[\frac{\partial}{\partial y}\right]_{x,z} \tag{14.67}$$

the horizontal gradient operator on constant z surfaces, and

$$\nabla_s = \hat{\boldsymbol{x}} \left[\frac{\partial}{\partial x}\right]_{y,s} + \hat{\boldsymbol{y}} \left[\frac{\partial}{\partial y}\right]_{x,s} \tag{14.68}$$

the horizontal gradient operator on constant s surfaces.[1] The equality (14.65d) made use of a trigonometric identity, and the equality (14.65e) introduced the horizontal projection of the area,

$$\mathrm{d}A = |\cos \vartheta| \, \mathrm{d}\mathcal{S}. \tag{14.69}$$

These results bring the kinematic boundary condition (14.63) into the form

$$\rho \left(\boldsymbol{v} - \boldsymbol{v}^{(s)}\right) \cdot \hat{\boldsymbol{n}} \, \mathrm{d}\mathcal{S} = -\mathcal{Q}_{\mathrm{m}} \, \mathrm{d}\mathcal{S} \tag{14.70a}$$

$$= \rho \frac{\mathrm{D}s}{\mathrm{D}t} \left|\frac{\partial z}{\partial s}\right| \mathrm{d}A \tag{14.70b}$$

$$\equiv -Q_{\mathrm{m}} \, \mathrm{d}A. \tag{14.70c}$$

As defined, the flux Q_{m} is the net mass per time per horizontal area crossing the boundary surface

$$Q_{\mathrm{m}} = -\rho \left(\boldsymbol{v} - \boldsymbol{v}^{(s)}\right) \cdot \hat{\boldsymbol{n}} \frac{\mathrm{d}\mathcal{S}}{\mathrm{d}A} = -\rho \frac{\mathrm{D}s}{\mathrm{D}t} \left|\frac{\partial z}{\partial s}\right|. \tag{14.71}$$

The minus sign is a convention that we motivate through the ocean free surface example in the following.

[1] We discuss such operators in Chapter 8 when studying the mathematics of generalized vertical coordinates.

Kinematic boundary condition at the ocean free surface

Consider the ocean free surface located at

$$s(x, y, z, t) = z - \eta(x, y, t) = 0 \qquad \text{ocean free surface.} \tag{14.72}$$

For this boundary, $\partial s/\partial z = 1$, so that the boundary condition (14.70b) takes on the form

$$\rho \left[\frac{\mathrm{D}(z-\eta)}{\mathrm{D}t} \right] = -Q_{\mathrm{m}} \Rightarrow w + \rho^{-1} Q_{\mathrm{m}} = \frac{\partial \eta}{\partial t} + \boldsymbol{u} \cdot \nabla \eta. \tag{14.73}$$

To motivate the sign convention in equation (14.70c), consider the special case of a flat free surface and a resting fluid with $\boldsymbol{v} = 0$. Adding mass to the ocean raises the free surface, so that $\partial \eta/\partial t > 0$. Hence, our sign convention means that $Q_{\mathrm{m}} > 0$ corresponds to mass added to the ocean.

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 symbol[2]

$$s = \sigma(x, y, z, t) \tag{14.74}$$

for a particular isopycnal, σ. The mass transport crossing the isopycnal is written

$$Q_{\mathrm{m}} = \rho \, \frac{\mathrm{D}\sigma}{\mathrm{D}t} \left| \frac{\partial z}{\partial \sigma} \right| \equiv \rho \, w^{(\dot{\sigma})}, \tag{14.75}$$

where we introduced the *diapycnal transport velocity*

$$w^{(\dot{\sigma})} \equiv \frac{\mathrm{D}\sigma}{\mathrm{D}t} \left| \frac{\partial z}{\partial \sigma} \right|. \tag{14.76}$$

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.

14.5 Mass budget for a column of ocean fluid

We close this chapter by deriving the mass budget for a column of ocean fluid such as that shown in Figure 14.7. This derivation requires much of the formalism discussed earlier, thus serving as a useful close to this chapter. In outline form, the derivation proceeds by vertically integrating the mass continuity equation (14.6) over the depth of an ocean column, from $z = \eta_{\mathrm{b}}(x, y)$ 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.

[2] In this book, we use σ as an arbitrary generalized vertical coordinate, here chosen to be an isopycnal.

Figure 14.7: *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, $\rho\, u$ 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 also 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. Mass flo*

Vertically integrating the continuity equation (14.6) for a compressible fluid renders

$$-\int_{\eta_{\mathrm{b}}}^{\eta} \frac{1}{\rho} \frac{\mathrm{D}\rho}{\mathrm{D}t}\, \mathrm{d}z = \int_{\eta_{\mathrm{b}}}^{\eta} \nabla \cdot \boldsymbol{v}\, \mathrm{d}z \tag{14.77a}$$

$$= w(\eta) - w(\eta_{\mathrm{b}}) + \int_{\eta_{\mathrm{b}}}^{\eta} \nabla_z \cdot \boldsymbol{u}\, \mathrm{d}z \tag{14.77b}$$

$$= w(\eta) - w(\eta_{\mathrm{b}}) + \nabla_z \cdot \left[\int_{\eta_{\mathrm{b}}}^{\eta} \boldsymbol{u}\, \mathrm{d}z\right] - \boldsymbol{u}(\eta) \cdot \nabla_z \eta + \boldsymbol{u}(\eta_{\mathrm{b}}) \cdot \nabla_z \eta_{\mathrm{b}} \tag{14.77c}$$

$$= [w(\eta) - \boldsymbol{u}(\eta) \cdot \nabla_z \eta] - [w(\eta_{\mathrm{b}}) - \boldsymbol{u}(\eta_{\mathrm{b}}) \cdot \nabla_z \eta_{\mathrm{b}}] + \nabla_z \cdot \left[\int_{\eta_{\mathrm{b}}}^{\eta} \boldsymbol{u}\, \mathrm{d}z\right], \tag{14.77d}$$

where we made use of Leibniz's Rule to move the horizontal divergence outside of the integral. Also note that $\nabla \cdot \boldsymbol{u} = \nabla_z \cdot \boldsymbol{u}$ since $\boldsymbol{u}$ is the horizontal velocity. Likewise for $\nabla \eta_{\mathrm{b}}$ and $\nabla \eta$ since η_{b} and η are both spatially two-dimensional functions.

Use of the surface kinematic boundary condition (14.73) and no-normal flow bottom boundary condition yield

$$\frac{\partial \eta}{\partial t} = \frac{Q_{\mathrm{m}}}{\rho(\eta)} - \nabla \cdot \boldsymbol{U} - \int_{\eta_{\mathrm{b}}}^{\eta} \frac{1}{\rho} \frac{\mathrm{D}\rho}{\mathrm{D}t}\, \mathrm{d}z \tag{14.78}$$

where

$$\boldsymbol{U} = \int_{\eta_{\mathrm{b}}}^{\eta} \boldsymbol{u}\, \mathrm{d}z \tag{14.79}$$

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 (14.78) by unpacking the physical processes leading to the material evolution of density.

14.6 Exercises

EXERCISE 14.1: CENTER OF MASS MOTION

Consider a material fluid region, $\mathcal{R}(\boldsymbol{v})$, with constant mass written as

$$M = \int_{\mathcal{R}(\boldsymbol{v})} \rho \, \mathrm{d}V. \tag{14.80}$$

(a) Show mathematically that the centre of mass for the region moves with the region's total linear momentum

$$\frac{\mathrm{d}}{\mathrm{d}t}\left[\frac{1}{M}\int_{\mathcal{R}(\boldsymbol{v})} \rho \, \boldsymbol{x} \, \mathrm{d}V\right] = \frac{1}{M}\int_{\mathcal{R}(\boldsymbol{v})} \rho \, \boldsymbol{v} \, \mathrm{d}V. \tag{14.81}$$

Precisely describe the reasoning behind each step. Note: a very 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

$$\frac{\mathrm{d}}{\mathrm{d}t}\left[\int_{\mathcal{R}(\boldsymbol{v})} \rho \, \boldsymbol{v} \, \mathrm{d}V\right] = \int_{\mathcal{R}(\boldsymbol{v})} \frac{\mathrm{D}\boldsymbol{v}}{\mathrm{D}t} \, \rho \, \mathrm{d}V. \tag{14.82}$$

Precisely describe the reasoning behind each step. Note: a very brief solution is sufficient, so long as the reasoning is sound.

EXERCISE 14.2: MASS BUDGET FOR A FLUID COLUMN

We here derive the equation for mass conservation over a column of fluid, such as a seawater column extending from the ocean bottom to its surface. This exercise shares much with Section 14.5, but we come at the problem differently and arrive at a slightly different (though equivalent) form for column mass balance. Figure 14.7 provides a schematic of the setup. This is a long problem to state, but the solution is actually quite minimal.

The mass within an arbitrary fluid region is given by

$$M = \int \rho \, \mathrm{d}V. \tag{14.83}$$

Consider the fluid mass within the column shown in Figure 14.7. In this column, the vertical sidewalls are fixed in time, the bottom surface, $z = \eta_{\mathrm{b}}(x, y)$, 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

$$M = \int \mathrm{d}x \, \mathrm{d}y \int_{\eta_{\mathrm{b}}(x,y)}^{\eta(x,y,t)} \rho \, \mathrm{d}z, \tag{14.84}$$

where the horizontal (x, y) integrals extend over the horizontal area of the column. Mass conservation for this column means that the change in mass arises just through boundary fluxes, so that

$$\frac{\mathrm{d}M}{\mathrm{d}t} = -\int \rho\, \Delta \boldsymbol{v} \cdot \hat{\boldsymbol{n}}\, \mathrm{d}\mathcal{S}, \tag{14.85}$$

where $\hat{\boldsymbol{n}}$ is the outward normal to the surface of the fluid region, $\mathrm{d}\mathcal{S}$ 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

$$\Delta \boldsymbol{v} = \boldsymbol{v} - \boldsymbol{v}^{(\mathrm{s})} \tag{14.86}$$

is the velocity of the fluid relative to the velocity of the boundary.

Mass transported in the zonal direction $(\hat{\boldsymbol{x}})$ that crosses the column's vertical boundary at x is given by

$$F^{(x)}(x,t) = \int \mathrm{d}y \int_{\eta_{\mathrm{b}}(x,y)}^{\eta(x,y,t)} u(x,y,z,t)\, \rho(x,y,z,t)\, \mathrm{d}z. \tag{14.87}$$

What are the physical dimensions for the mass transport $F^{(x)}$? There is a similar expression for mass crossing each of the other vertical faces of the column in the two horizontal directions $(\hat{\boldsymbol{x}}, \hat{\boldsymbol{y}})$. Using these expressions for the mass crossing the vertical side boundaries, 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

$$\frac{\partial}{\partial t}\left[\int_{\eta_{\mathrm{b}}}^{\eta} \rho\, \mathrm{d}z\right] = -\nabla \cdot \left[\int_{\eta_{\mathrm{b}}}^{\eta} \boldsymbol{u}\, \rho\, \mathrm{d}z\right] + Q_{\mathrm{m}}. \tag{14.88}$$

In this expression, Q_{m} is the mass transport entering the ocean through the surface, per horizontal area, as defined by equation (14.70c), so that

$$\int Q_{\mathrm{m}}\, \mathrm{d}x\, \mathrm{d}y = -\int \rho\, \Delta \boldsymbol{v} \cdot \hat{\boldsymbol{n}}\, \mathrm{d}\mathcal{S} \qquad \text{at } z = \eta. \tag{14.89}$$

In words, the mass budget in equation (14.88) 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 (14.88)?

EXERCISE 14.3: 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

$$M = \int_{\mathcal{R}} \rho\, \mathrm{d}V. \tag{14.90}$$

Show that the time change in the linear momentum for this ocean basin is given by

$$\frac{\mathrm{d}}{\mathrm{d}t}\left[\int_{\mathcal{R}} \rho\, \boldsymbol{v}\, \mathrm{d}V\right] = \int_{\mathcal{R}} \frac{\mathrm{D}\boldsymbol{v}}{\mathrm{D}t}\, \rho\, \mathrm{d}V. \tag{14.91}$$

Figure 14.8: *Cross-section of the integration region for Exercise 14.3, with the region extending from the ocean bottom at $z = \eta_\mathrm{b}(x, y)$ and the free surface at $z = \eta(x, y, t)$. An infinitesimal column is shown with cross-sectional area $\mathrm{d}A$, 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 (14.92).*

This result is identical to that derived in Exercise 14.1. Rather than just repeating the solution method used there, make explicit use of Leibniz's theorem, the kinematic boundary condition detailed in Section 14.4.2, and mass conservation.

Hint: Refer to Figure 14.8 for a schematic of the integration where we have expanded the volume integral into the form

$$\frac{\mathrm{d}}{\mathrm{d}t}\left[\int_{\mathcal{R}} \rho\,\boldsymbol{v}\,\mathrm{d}V\right] = \frac{\mathrm{d}}{\mathrm{d}t}\left[\int \mathrm{d}A \int_{\eta_\mathrm{b}}^{\eta} \rho\,\boldsymbol{v}\,\mathrm{d}z\right], \tag{14.92}$$

where the horizontal integral extends over the horizontal area of the basin, $\mathrm{d}A = \mathrm{d}x\,\mathrm{d}y$ is the time independent horizontal area element, $z = \eta_\mathrm{b}(x, y)$ 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 14.4.3 (in particular equation (14.69)). Unlike the formulation in Exercise 14.1, 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.

15

Tracer conservation

As in Chapter 14, we assume that mass is neither created nor destroyed anywhere within the fluid domain, so that mass within the fluid changes only via fluxes crossing the domain boundary. We here extend that discussion to the case of a non-homogeneous fluid comprised of multiple trace matter constituents, thus developing budgets for tracers within the fluid, both for fluid elements and for finite regions.

Our formulation is based on conceiving the non-homogeneous fluid as interpenetrating continua defined by the distinct matter constituents (e.g., salt and freshwater for the ocean or water vapor and dry air for the atmosphere). This conceptual framework allows us to formulate separate and uncoupled mass conservation statements for each constituent. Summing over all constituents renders a total mass budget. The mathematical form of the total mass budget is identical to that for a homogeneous fluid when making use of the *barycentric* (center of mass) velocity. The resulting constituent mass budgets have "diffusive" fluxes arising from constituent velocities being distinct from the barycentric velocity.

READER'S GUIDE TO THIS CHAPTER

The conceptual formulation pursued in this chapter is standard in the chemical physics community (e.g., Chapter 11 of *Aris* (1962) and Chapter II of *DeGroot and Mazur* (1984)), where chemical reactions are also included in addition to the transport processes arising from a multi-component fluid. For this chapter, we assume an understanding of the Eulerian and Lagrangian kinematic descriptions detailed in Chapter 12 and the mass conservation analysis in Chapter 14. Much of the material from this chapter is used for those subjects involving scalar fields such as potential enthalpy and material tracers. Furthermore, the Leibniz-Reynolds transport theorem of Section 15.3.4 is a kinematic expression that is central to all finite volume budgets developed in this book.

15.1 Fluid elements and the tracer equation **252**
15.1.1 Constituent parcels and regions 252
15.1.2 Total mass conservation 253
15.1.3 Revisiting the fluid element 254
15.1.4 The tracer equation 254
15.1.5 Compatibility between total mass and tracer mass 255
15.1.6 Comments 256
15.1.7 Further study 256

15.2 Passive tracers **256**
15.3 Budgets for arbitrary fluid regions **257**
15.3.1 Extensive and intensive fluid properties 257
15.3.2 General form of the finite domain integral 258
15.3.3 Eulerian (static) domain 258
15.3.4 Leibniz-Reynolds Transport Theorem 259
15.3.5 Revisiting Reynolds transport theorem 261
15.3.6 Summary of time derivatives acting on integrals 263
15.4 Boundary conditions for the tracer budget **263**
15.5 Brute force illustration of Leibniz-Reynolds **265**
15.5.1 Leibniz's rule plus kinematic boundary conditions 266
15.5.2 Summarizing the result 266
15.6 Comments on local conservation laws **267**

15.1 Fluid elements and the tracer equation

As defined in Section 12.2, 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 trace matter (i.e., tracers) with adjacent 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 Reynold's Transport Theorem for integration over a constant mass fluid region (Section 15.3.4). 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*.

15.1.1 Constituent parcels and regions

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 region of the fluid, $\mathcal{R}$, with volume V and total mass M. Inside of $\mathcal{R}$, 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, $\boldsymbol{v}^{(n)}$, as well as the mass density,

$$\rho^{(n)} = V^{-1} M^{(n)}, \tag{15.1}$$

with these constituent properties defined over a constituent material region, $\mathcal{R}^{(n)}$. In the continuum limit where the volume and mass in $\mathcal{R}$ get tiny yet the mass density remains finite, then the constituent velocity and mass density are continuous fields whose values are available at each point within the continuum fluid.

Strong formulation

The constituent material regions, $\mathcal{R}^{(n)}$, are subsets of the region $\mathcal{R}$, and they each move according to the corresponding constituent velocity, $\boldsymbol{v}^{(n)}$. That is, the material regions $\mathcal{R}^{(n)}$ maintain constant

constituent mass $M^{(n)}$. In the infinitesimal limit, mass conservation for $\mathcal{R}^{(n)}$ leads the constituent mass continuity equation[1]

$$\frac{\mathrm{D}^{(n)}\rho^{(n)}}{\mathrm{D}t} = -\rho^{(n)}\,\nabla\cdot\boldsymbol{v}^{(n)} \qquad \text{for each of the } n = 1, N \text{ constituents,} \tag{15.2}$$

where the constituent material time derivative is given by

$$\frac{\mathrm{D}^{(n)}}{\mathrm{D}t} = \frac{\partial}{\partial t} + \boldsymbol{v}^{(n)}\cdot\nabla. \tag{15.3}$$

We thus have N statements of mass conservation corresponding to each constituent material fluid parcel moving according to the velocity $\boldsymbol{v}^{(n)}$. Finally, note that the mass continuity equation (15.2) takes on its Eulerian form

$$\frac{\partial\rho^{(n)}}{\partial t} + \nabla\cdot(\rho^{(n)}\,\boldsymbol{v}^{(n)}) = 0. \tag{15.4}$$

Weak formulation

The previous formulation made use of infinitesimals and thus formulated the constituent mass budget in terms of partial differential equations. Here we offer a complementary weak formulation[2] whereby we consider an arbitrary region, $\mathcal{R}$, assumed to be fixed in space (i.e., an Eulerian volume). The mass of component n within $\mathcal{R}$ changes in time according to the finite volume budget

$$\frac{\mathrm{d}}{\mathrm{d}t}\int_{\mathcal{R}}\rho^{(n)}\,\mathrm{d}V = -\int_{\partial\mathcal{R}}\rho^{(n)}\,\boldsymbol{v}^{(n)}\cdot\hat{\boldsymbol{n}}\,\mathrm{d}\mathcal{S}, \tag{15.5}$$

which is a constituent form of the finite volume mass budget given for a homogeneous fluid by equation (14.17). Since the volume is fixed in space, we can move the time derivative across the integral to reveal

$$\int_{\mathcal{R}}\left[\frac{\partial\rho^{(n)}}{\partial t} + \nabla\cdot(\rho^{(n)}\,\boldsymbol{v}^{(n)})\right]\mathrm{d}V = 0, \tag{15.6}$$

where we also used the divergence theorm to convert the surface integral to a volume integral. Arbitrariness of the region, $\mathcal{R}$, means that this integral expression must be satisfied at each point of the conintuum, thus leading again to the Eulerian form of the constituent mass continuity equation (15.4).

15.1.2 Total mass conservation

Summing the Eulerian mass continuity equation (15.4) over all constituents leads to the continuity equation for the total mass

$$\frac{\partial\rho}{\partial t} + \nabla\cdot(\rho\,\boldsymbol{v}) = 0, \tag{15.7}$$

where the total mass density and *barycentric velocity* are given by

$$\rho = \sum_{n=1}^{N}\rho^{(n)} \qquad \text{and} \qquad \boldsymbol{v} = \rho^{-1}\sum_{n=1}^{N}\rho^{(n)}\,\boldsymbol{v}^{(n)}. \tag{15.8}$$

[1] There is no implied summation in equation (15.2).

[2] Recall our discussion of weak and strong formulation in Section 12.1.

Introducing the material time derivative following the barycentric velocity, $\mathrm{D}/\mathrm{D}t = \partial/\partial t + \boldsymbol{v}\cdot\nabla$, leads to the equivalent material form for the mass conservation equation

$$\frac{\mathrm{D}\rho}{\mathrm{D}t} = -\rho\,\nabla\cdot\boldsymbol{v}. \tag{15.9}$$

Note that the *barycenter* of a distribution of matter is the center of mass. We choose the term *barycentric velocity* for $\boldsymbol{v}$ to distinguish $\boldsymbol{v}$ from the molecular center of mass velocity, $\boldsymbol{v}^{(n)}$, of each constituent. The barycentric velocity plays a central role the conservation laws for multi-component fluids.

15.1.3 Revisiting the fluid element

The mass continuity equation (15.9) motivates us to define a fluid element as an infinitesimal fluid parcel that moves with barycentric velocity $\boldsymbol{v}$ and maintains a constant total mass

$$\delta M = \sum_{n=1}^{N} \delta M^{(n)}. \tag{15.10}$$

The fluid element does not maintain a constant mass for each constituent, since the fluid element moves at the barycentric velocity, $\boldsymbol{v}$, which generally differs from the constituent velocities $\boldsymbol{v}^{(n)}$. 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.

15.1.4 The tracer equation

Rather than keep track of each constituent velocity, $\boldsymbol{v}^{(n)}$, 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 (15.4)

$$\left[\frac{\partial}{\partial t} + \boldsymbol{v}^{(n)}\cdot\nabla\right]\rho^{(n)} = -\rho^{(n)}\nabla\cdot\boldsymbol{v}^{(n)} \tag{15.11}$$

and insert the barycentric velocity to both sides by adding $0 = \boldsymbol{v} - \boldsymbol{v}$

$$\left[\frac{\partial}{\partial t} + (\boldsymbol{v} - \boldsymbol{v} + \boldsymbol{v}^{(n)})\cdot\nabla\right]\rho^{(n)} = -\rho^{(n)}\nabla\cdot[\boldsymbol{v} - \boldsymbol{v} + \boldsymbol{v}^{(n)}]. \tag{15.12}$$

Rearrangment leads to

$$\left[\frac{\partial}{\partial t} + \boldsymbol{v}\cdot\nabla\right]\rho^{(n)} = -\rho^{(n)}\nabla\cdot\boldsymbol{v} - \nabla\cdot[\rho^{(n)}\,(\boldsymbol{v}^{(n)} - \boldsymbol{v})], \tag{15.13}$$

which can be written

$$\frac{\mathrm{D}\rho^{(n)}}{\mathrm{D}t} = -\rho^{(n)}\,\nabla\cdot\boldsymbol{v} - \nabla\cdot\boldsymbol{J}^{(n)}, \tag{15.14}$$

where we defined the constituent mass flux

$$\boldsymbol{J}^{(n)} = \rho^{(n)}\,(\boldsymbol{v}^{(n)} - \boldsymbol{v}). \tag{15.15}$$

Figure 15.1: *A finite sized cube as in Figure 14.1, 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, $\boldsymbol{v}$, there is a diffusive flux, $\boldsymbol{J}$, that contributes to the transfer of tracer across the cell face.*

The dimensions of $\boldsymbol{J}^{(n)}$ are mass of constituent n per time per area.

The material mass conservation equation (15.14) takes on the Eulerian form

$$\frac{\partial \rho^{(n)}}{\partial t} + \nabla \cdot (\boldsymbol{v}\,\rho^{(n)}) = -\nabla \cdot \boldsymbol{J}^{(n)}. \tag{15.16}$$

Introducing the tracer concentration $C^{(n)}$ according to

$$C^{(n)} = \frac{\rho^{(n)}}{\rho} = \frac{\delta M^{(n)}}{\delta M} = \frac{\text{mass of constituent } n \text{ in fluid element}}{\text{mass of fluid element}}, \tag{15.17}$$

leads to the tracer flux

$$\boldsymbol{J}^{(n)} = \rho\,C^{(n)}\,(\boldsymbol{v}^{(n)} - \boldsymbol{v}), \tag{15.18}$$

the Eulerian flux-form tracer budget

$$\frac{\partial(\rho\,C^{(n)})}{\partial t} + \nabla \cdot [\boldsymbol{v}\,\rho\,C^{(n)} + \boldsymbol{J}^{(n)}] = 0, \tag{15.19}$$

and the corresponding material form of the tracer equation

$$\rho\,\frac{\mathrm{D}C^{(n)}}{\mathrm{D}t} = -\nabla \cdot \boldsymbol{J}^{(n)}. \tag{15.20}$$

In Figure 15.1 we illustrate the contributions to the tracer evolution as viewed in the Eulerian flux-form tracer equation (15.19).

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

$$\rho\,C^{(n)}\,\boldsymbol{v}^{(n)} = \rho\,C^{(n)}\,(\boldsymbol{v}^{(n)} - \boldsymbol{v} + \boldsymbol{v}) = \boldsymbol{J}^{(n)} + \rho\,C^{(n)}\,\boldsymbol{v}. \tag{15.21}$$

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.

15.1.5 Compatibility between total mass and tracer mass

By construction, the Eulerian flux-form of the tracer equation (15.19) is compatible with the flux-form continuity equation

$$\frac{\partial(\rho\,C^{(n)})}{\partial t} + \nabla \cdot [\rho\,\boldsymbol{v}\,C^{(n)} + \boldsymbol{J}^{(n)}] = 0 \Longleftrightarrow \frac{\partial \rho}{\partial t} + \nabla \cdot (\rho\,\boldsymbol{v}) = 0. \tag{15.22}$$

Compatibility is manifest by summing the tracer equation over all constituents and using the identities

$$\sum_{n=1}^{N} C^{(n)} = 1 \qquad \text{and} \qquad \sum_{n=1}^{N} \boldsymbol{J}^{(n)} = 0. \tag{15.23}$$

Furthermore, through use of the barycentric velocity (15.8), 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 $\boldsymbol{J}^{(n)}$ since $\sum_{n=1}^{N} \boldsymbol{J}^{(n)} = 0$.

15.1.6 Comments

The tracer equation expresses the balance of mass for each trace constituent in the fluid. Again, a nonzero tracer flux, $\boldsymbol{J}^{(n)} = \rho\, C^{(n)} \left(\boldsymbol{v}^{(n)} - \boldsymbol{v}\right)$, arises when the barycentric velocity, $\boldsymbol{v}$, differs from the constituent velocity, $\boldsymbol{v}^{(n)}$. 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 47.2). Correspondingly, the mass of trace matter in a fluid element is altered in the presence of trace matter concentration gradients.

15.1.7 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 bioegeochemistry. 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).

15.2 Passive tracers

As defined in equation (15.17), 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 (15.8) 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 12.2 whose trajectories define the Lagrangian reference frame. We make use of passive tracers to probe the advective and diffusive features of the flow without in turn 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 13 and Section 14.4.2, we discussed the notion of a material fluid object, which is an object comprised of fluid particles that follow the velocity, $\boldsymbol{v}$. 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 $\boldsymbol{v}^{(n)} \neq \boldsymbol{v}$. 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.

15.3 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 14 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.

15.3.1 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, linear momentum. An intensive property generally does not change when removing some of the sample. Examples are number density (number of particles per volume), mass density (mass of substance per unit 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. We have more to say about intensive and extensive properties when considering thermodynamics in Chapter 22.

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 the 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, or is it closed? Is the region static (Eulerian) or do boundaries move? If the boundaries move, then do they move with fluid elements (Lagrangian) or are they moving in some other manner?

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 for the element

$$\Pi = \text{intensive fluid property} \tag{15.24a}$$

$$\Pi\,\rho\,\delta V = \text{extensive fluid property.} \tag{15.24b}$$

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. If Π is the Conservative Temperature, Θ, of a fluid element, then the corresponding extensive property, $\Theta\,C_p\,\rho\,\delta V$, is the potential enthalpy (C_p is the specific heat capacity; see Section 23.5).

We furthermore assume that Π satisfies the scalar conservation equation, written here in both its material form and Eulerian flux-form

$$\rho\,\frac{\mathrm{D}\Pi}{\mathrm{D}t} = -\nabla\cdot\boldsymbol{J} \qquad \Longleftrightarrow \qquad \frac{\partial(\rho\,\Pi)}{\partial t} + \nabla\cdot(\rho\,\Pi\,\boldsymbol{v} + \boldsymbol{J}) = 0, \tag{15.25}$$

where $\boldsymbol{J}$ is a flux such as that associated with the tracer equation derived in Section 15.1.4. Depending on the context, the budget equation (15.25) 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, in the absence of a flux convergence, $-\nabla \cdot \boldsymbol{J} = 0$, the scalar field is constant following a material fluid particle

$$-\nabla \cdot \boldsymbol{J} = 0 \Longrightarrow \frac{\mathrm{D}\Pi}{\mathrm{D}t} = 0. \tag{15.26}$$

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 12.3). 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 12.6).

15.3.2 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

$$\mathcal{I}[\mathcal{R}(t), t] = \int_{\mathcal{R}(t)} \Pi \, \rho \, \mathrm{d}V \equiv \int_{\mathcal{R}(t)} \varphi \, \mathrm{d}V, \tag{15.27}$$

where we introduced the shorthand

$$\varphi = \rho \, \Pi. \tag{15.28}$$

The integrand in equation (15.27) is a function of space and time, $\varphi = \varphi(\boldsymbol{x}, t)$, 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 14.3) or a constant mass fluid region open to the exchange of matter with the surroundings (Section 15.1). In both of these cases the region was denoted by $\mathcal{R}(\boldsymbol{v})$ since it moved with the fluid flow. Here we make no *a priori* assumption about how the region moves, or whether it moves or is static.

The total time derivative of $\mathcal{I}$ can be written as

$$\frac{\mathrm{d}\mathcal{I}}{\mathrm{d}t} = \left[\frac{\partial \mathcal{I}}{\partial t}\right]_{\mathcal{R}} + \frac{\mathrm{d}\mathcal{R}}{\mathrm{d}t} \left[\frac{\partial \mathcal{I}}{\partial \mathcal{R}}\right]_t. \tag{15.29}$$

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 (15.29) is directly analogous to the total time derivative of a field given by equation (12.26).

15.3.3 Eulerian (static) domain

We first consider an Eulerian domain, which is fixed in space and thus static. In this case we have

$$\frac{\mathrm{d}\mathcal{I}}{\mathrm{d}t} = \left[\frac{\partial \mathcal{I}}{\partial t}\right]_{\mathcal{R}} = \frac{\partial}{\partial t}\left[\int_{\mathcal{R}} \Pi \, \rho \, \mathrm{d}V\right] = \int_{\mathcal{R}} \left[\frac{\partial(\rho \, \Pi)}{\partial t}\right] \mathrm{d}V. \tag{15.30}$$

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 (15.29) vanishes. Furthermore,

since the domain is static, the volume element, $\mathrm{d}V$, provides a static partition of the total domain volume so that $\mathrm{d}V$ does not appear inside the time derivative. This case corresponds to the Eulerian budgets depicted in Figures 14.1, 14.2, and 15.1.

15.3.4 Leibniz-Reynolds Transport Theorem

We now allow the domain boundaries to be time dependent so that both terms in the total time derivative (15.29) 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 14.4). 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

$$\frac{\mathrm{d}}{\mathrm{d}t}\left[\int_{x_1(t)}^{x_2(t)} \varphi(x,t)\,\mathrm{d}x\right] = \int_{x_1(t)}^{x_2(t)} \frac{\partial \varphi}{\partial t}\,\mathrm{d}x + \frac{\mathrm{d}}{\mathrm{d}t}\left[\int_{x_1(t)}^{x_2(t)}\right]\varphi(x,t)\,\mathrm{d}x \tag{15.31a}$$

$$= \int_{x_1(t)}^{x_2(t)} \frac{\partial \varphi}{\partial t}\,\mathrm{d}x + \frac{\mathrm{d}x_2(t)}{\mathrm{d}t}\varphi(x_2,t) - \frac{\mathrm{d}x_1(t)}{\mathrm{d}t}\varphi(x_1,t) \tag{15.31b}$$

$$= \int_{x_1(t)}^{x_2(t)} \left[\frac{\partial \varphi}{\partial t} + \frac{\partial}{\partial x}\left(\varphi\,\frac{\mathrm{d}x}{\mathrm{d}t}\right)\right]\mathrm{d}x. \tag{15.31c}$$

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

$$\frac{\mathrm{d}}{\mathrm{d}t}\left[\int_{\mathcal{R}} \varphi\,\mathrm{d}V\right] = \int_{\mathcal{R}} \left[\frac{\partial \varphi}{\partial t} + \nabla\cdot(\varphi\,\boldsymbol{v}^{(b)})\right]\mathrm{d}V = \int_{\mathcal{R}} \frac{\partial \varphi}{\partial t}\,\mathrm{d}V + \oint_{\partial\mathcal{R}} \varphi\,\boldsymbol{v}^{(b)}\cdot\hat{\boldsymbol{n}}\,\mathrm{d}\mathcal{S}, \tag{15.32}$$

where the second equality made use of Gauss's 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

$$\boldsymbol{v}^{(b)} = \frac{\mathrm{d}\boldsymbol{x}}{\mathrm{d}t}. \tag{15.33}$$

Comments on the Leibniz-Reynolds Transport Theorem

The identity (15.32) 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 (15.32). 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 (15.32) offers a coordinate invariant measure of how an extensive property evolves within a region. Hence, by the rules of tensor analysis from Chapter 6, 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 51, where budgets are developed over arbitrary regions.

The nature of the boundary velocity, $\boldsymbol{v}^{(b)}$, warrants some comment. As defined by equation (15.33), 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, $\boldsymbol{v}^{(b)} \cdot \hat{\boldsymbol{n}}$. 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.

Transport theorem for region volume

Setting $\varphi = 1$ in the transport theorem (15.32) yields the time change for the region volume

$$\frac{\mathrm{d}}{\mathrm{d}t}\left[\int_{\mathcal{R}} \mathrm{d}V\right] = \oint_{\partial\mathcal{R}} \boldsymbol{v}^{(b)} \cdot \hat{\boldsymbol{n}}\, \mathrm{d}\mathcal{S}. \tag{15.34}$$

This result is quite sensible, as it 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 (14.22). The expressions are the same since $\boldsymbol{v}^{(b)} \cdot \hat{\boldsymbol{n}} = \boldsymbol{v}^{(b)} \cdot \hat{\boldsymbol{n}}$ for a material region.

The volume budget (15.34) holds for both compressible and incompressible fluids, with specialization to the incompressible case in Section 16.6.2. Finally, we further emphasize that the boundary term in equations (15.32) and (15.34) projects out just the normal component to the boundary velocity; we never make use of information about its tangential component. Indeed, information concerning the tangential component is not available to us without making dynamical assumptions that go beyond the kinematics considered here.

Transport theorem for a scalar field

We here derive a corollary to the transport theorem (15.32) that proves useful for budget analyses over moving regions, with generalization to the vector linear momentum provided in Section 19.5. For this purpose, make use of the Eulerian flux-form of the scalar conservation equation (15.25) so that the transport theorem is written

$$\frac{\mathrm{d}}{\mathrm{d}t}\left[\int_{\mathcal{R}} \rho\, \Pi\, \mathrm{d}V\right] = -\oint_{\partial\mathcal{R}} \left[\rho\, \Pi\, (\boldsymbol{v} - \boldsymbol{v}^{(b)}) + \boldsymbol{J}\right] \cdot \hat{\boldsymbol{n}}\, \mathrm{d}\mathcal{S}. \tag{15.35}$$

Setting $\Pi = 1$ gives an expression for the change in mass for the region

$$\frac{\mathrm{d}}{\mathrm{d}t}\left[\int_{\mathcal{R}} \rho\, \mathrm{d}V\right] = -\oint_{\partial\mathcal{R}} \rho\, (\boldsymbol{v} - \boldsymbol{v}^{(b)}) \cdot \hat{\boldsymbol{n}}\, \mathrm{d}\mathcal{S}. \tag{15.36}$$

The transport theorem (15.35) 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

Figure 15.2: *This figure depicts the contributions to the Leibnitz-Reynolds transport theoream (15.35). 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 8, 17, and 38. In particular, the vertical cell faces are assumed to have fixed positions, so that* $(\boldsymbol{v} - \boldsymbol{v}^{(b)}) \cdot \hat{\boldsymbol{n}} = \boldsymbol{v} \cdot \hat{\boldsymbol{n}}$ *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.*

integral for the transport of Π-stuff through the boundary of the region. The first transport 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 transport terms are projected onto the outward normal at the boundary. Hence, the budget is not affected by transport tangential to the boundary. Finally, for the mass budget (15.36), the diffusive flux vanishes since the mass of a fluid element moves according to the barycentric velocity of Section (15.1.2).

In Figure 15.2 we illustrate the transport theorem (15.35) 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. We offer further discussion of the kinematics of such *generalized vertical coordinate* models in Chapter 17 and their dynamics in Chapter 38.

15.3.5 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 14.3.4. The following results are special cases of the general expression (15.35).

Reynolds Transport Theorem

Let us apply the result (15.32) to a region that follows the fluid flow as defined by the barycentric velocity, $\boldsymbol{v}$. For this moving region, the time derivative of the region boundaries in equation (15.32)

is given by the fluid velocity thus leading to

$$\frac{\mathrm{d}}{\mathrm{d}t}\left[\int_{\mathcal{R}(\boldsymbol{v})} \varphi \,\mathrm{d}V\right] = \int_{\mathcal{R}(\boldsymbol{v})} \left[\frac{\partial \varphi}{\partial t} + \nabla\cdot(\boldsymbol{v}\,\varphi)\right]\mathrm{d}V = \int_{\mathcal{R}(\boldsymbol{v})} \left[\frac{\mathrm{D}\varphi}{\mathrm{D}t} + \varphi\,\nabla\cdot\boldsymbol{v}\right]\mathrm{d}V. \tag{15.37}$$

This result is the Reynolds transport theorem. The derivation given here is more general than that in Section 14.3.4 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. We did not assume the region boundaries are material. We can thus make use of Reynolds Transport Theorem (15.37) 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 made permeable via the diffusion of tracers.

Alternative form of Reynolds Transport Theorem

We can put the Reynolds Transport Theorem (15.37) into another useful form by reintroducing $\varphi = \rho\,\Pi$ and making use of mass continuity

$$\frac{1}{\rho}\frac{\mathrm{D}\rho}{\mathrm{D}t} = -\nabla\cdot\boldsymbol{v}. \tag{15.38}$$

Doing so yields the rather tidy result

$$\frac{\mathrm{d}}{\mathrm{d}t}\left[\int_{\mathcal{R}(\boldsymbol{v})} \Pi\,\rho\,\mathrm{d}V\right] = \int_{\mathcal{R}(\boldsymbol{v})} \left[\frac{\mathrm{D}\varphi}{\mathrm{D}t} + \varphi\,\nabla\cdot\boldsymbol{v}\right]\mathrm{d}V \tag{15.39a}$$

$$= \int_{\mathcal{R}(\boldsymbol{v})} \left[\frac{\mathrm{D}(\rho\,\Pi)}{\mathrm{D}t} + \rho\,\Pi\,\nabla\cdot\boldsymbol{v}\right]\mathrm{d}V \tag{15.39b}$$

$$= \int_{\mathcal{R}(\boldsymbol{v})} \left[\Pi\left(\frac{\mathrm{D}\rho}{\mathrm{D}t} + \rho\,\nabla\cdot\boldsymbol{v}\right) + \rho\,\frac{\mathrm{D}\Pi}{\mathrm{D}t}\right]\mathrm{d}V \tag{15.39c}$$

$$= \int_{\mathcal{R}(\boldsymbol{v})} \frac{\mathrm{D}\Pi}{\mathrm{D}t}\,\rho\,\mathrm{d}V. \tag{15.39d}$$

Heuristically, this result follows since $\rho\,\mathrm{d}V$ 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 (15.39d) one more step by inserting the material form of the scalar conservation equation (15.25) so that

$$\frac{\mathrm{d}}{\mathrm{d}t}\left[\int_{\mathcal{R}(\boldsymbol{v})} \Pi\,\rho\,\mathrm{d}V\right] = -\oint_{\partial\mathcal{R}(\boldsymbol{v})} \boldsymbol{J}\cdot\hat{\boldsymbol{n}}\,\mathrm{d}\mathcal{S}, \tag{15.40}$$

which is a special case of the general transport theorem (15.35) found by setting $(\boldsymbol{v} - \boldsymbol{v}^{(b)})\cdot\hat{\boldsymbol{n}} = 0$ 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 15.1.4. We extend this result to linear momentum in Section 19.5.

15.3.6 Summary of time derivatives acting on integrals

We here summarize the variety of time derivatives acting on integrals of scalar fields

$$\frac{\mathrm{d}}{\mathrm{d}t}\int_{\mathcal{R}} \rho\,\Pi\,\mathrm{d}V = \begin{cases} \int_{\mathcal{R}} \frac{\partial(\rho\,\Pi)}{\partial t}\,\mathrm{d}V = -\oint_{\partial\mathcal{R}} (\rho\,\boldsymbol{v}\,\Pi + \boldsymbol{J})\cdot\hat{\boldsymbol{n}}\,\mathrm{d}\mathcal{S} & \text{Eulerian (static) domain} \\ \int_{\mathcal{R}(\boldsymbol{v})} \rho\,\frac{\mathrm{D}\Pi}{\mathrm{D}t}\,\mathrm{d}V = -\oint_{\partial\mathcal{R}(\boldsymbol{v})} \boldsymbol{J}\cdot\hat{\boldsymbol{n}}\,\mathrm{d}\mathcal{S} & \text{Lagrangian (fluid flow) domain} \\ -\oint_{\partial\mathcal{R}} \left[\rho\,\Pi\,(\boldsymbol{v}-\boldsymbol{v}^{(b)}) + \boldsymbol{J}\right]\cdot\hat{\boldsymbol{n}}\,\mathrm{d}\mathcal{S} & \text{arbitrary domain,} \end{cases} \tag{15.41}$$

with the scalar fields assumed to satisfy the partial differential equation

$$\rho\,\frac{\mathrm{D}\Pi}{\mathrm{D}t} = -\nabla\cdot\boldsymbol{J} \Longleftrightarrow \frac{\partial(\rho\,\Pi)}{\partial t} + \nabla\cdot(\rho\,\boldsymbol{v}\,\Pi + \boldsymbol{J}) = 0. \tag{15.42}$$

As discussed in Section 12.1, the partial differential equation (15.42) is referred to as the *strong formulation* of the scalar budget, whereas the integral expressions in equation (15.41) provide various forms of the *weak formulations*.

15.4 Boundary conditions for the tracer budget

Here we extend the mass budget considerations from Section 14.5 to develop the budget for tracer mass (or potential enthalpy) within a fluid layer such as shown in Figure 15.3. Notably, we are interested in 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 isolines of generalized vertical coordinates whose layers are monotonically stacked in the vertical according to the discussion from Sections 8.8.1 and 17.2. 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 51. 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.

Figure 15.3: *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.*

The Leibniz-Reynolds transport theorem (15.35) provides our starting point

$$\frac{\mathrm{d}}{\mathrm{d}t}\left[\int_{\mathcal{R}} \rho\,C\,\mathrm{d}V\right] = -\int_{\partial\mathcal{R}} \left[\rho\,C\,(\boldsymbol{v}-\boldsymbol{v}^{(b)}) + \boldsymbol{J}\right]\cdot\hat{\boldsymbol{n}}\,\mathrm{d}\mathcal{S}, \tag{15.43}$$

where C is the concentration of a tracer that satisfies the advection-diffusion equation

$$\frac{\partial(\rho\, C)}{\partial t} + \nabla \cdot (\rho\, C\, \boldsymbol{v} + \boldsymbol{J}) = 0. \tag{15.44}$$

The left hand side of equation (15.43) is the time tendency for the mass of tracer within the region, such as the region $\mathcal{R}$ shown in Figure 15.3. This tendency is affected by transport across the layer boundaries, with three boundaries considered here.

Interior layer boundaries

The boundary transport across interior layer interfaces,

$$\text{interior boundary transport} = [\rho\, C\, (\boldsymbol{v} - \boldsymbol{v}^{(b)}) + \boldsymbol{J}] \cdot \hat{\boldsymbol{n}}\, \mathrm{d}\mathcal{S}, \tag{15.45}$$

measures the tracer mass transport due to advective fluxes across the moving layer boundaries (first term) and sugbrid scale fluxes (second term). The advective flux is sometimes known as the *dia-surface transport*, with a discussion of its kinematics given in Section 17.3.

Bottom boundary

At the bottom boundary, the no-normal flow condition means that

$$\boldsymbol{v} \cdot \hat{\boldsymbol{n}} = 0. \tag{15.46}$$

Consider the velocity of a point attached to the layer interface, $\boldsymbol{v}^{(b)}$, that also 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

$$\boldsymbol{v}^{(b)} \cdot \hat{\boldsymbol{n}} = 0. \tag{15.47}$$

Hence, the only contribution to the tracer budget at the bottom boundary comes through the subgrid scale flux $\boldsymbol{J}$

$$\text{bottom boundary transport} = \boldsymbol{J} \cdot \hat{\boldsymbol{n}}\, \mathrm{d}\mathcal{S}. \tag{15.48}$$

Geothermal heating is a common example of bottom boundary transport in the ocean. In the atmosphere, planetary boundary layer processes transfer tracer content from the land-atmosphere and sea-atmosphere boundary layer into the troposphere above.

Let us consider the geothermal heating with a bit more detail. In this case heating can be incorporated into the ocean as a boundary flux via

$$\text{geothermal heating} = c_p\, \boldsymbol{J}(\Theta) \cdot \hat{\boldsymbol{n}}\, \mathrm{d}\mathcal{S}, \tag{15.49}$$

where c_p is the heat capacity and Θ is the Conservative Temperature (see Chapter 50). If we furthermore assume the Θ flux is downgradient diffusion as per Fourier's Law with diffusivity $\kappa_\Theta > 0$ (Section 47.2.4), then the boundary condition takes the form

$$\text{geothermal heating} = -c_p\, \rho\, \kappa_\Theta\, \nabla\Theta \cdot \hat{\boldsymbol{n}}\, \mathrm{d}\mathcal{S}. \tag{15.50}$$

Geothermal heating thus supports a non-zero projection of the temperature gradient onto the normal direction. For those cases where the geothermal heating vanishes, or more generally for

tracers that have zero boundary flux, then the tracer must satisfy the following no-normal derivative boundary condition

$$\text{no flux boundary} = \nabla C \cdot \hat{\boldsymbol{n}} = 0. \tag{15.51}$$

In this case, isosurfaces of constant tracer are oriented normal to the boundary as depicted in Figure 15.4. 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 15.4: *In the absence of a boundary tracer flux, isosurfaces of tracer, C, intersect solid boundaries normal to the boundary as per equation (15.51): $\nabla C \cdot \hat{\boldsymbol{n}} = 0$, where $\hat{\boldsymbol{n}}$ is the outward boundary normal direction.*

Tracer boundary at the ocean surface

We make use of the kinematic boundary condition derived in Section 14.4.3 for the permeable surface ocean boundary, where the boundary condition (14.60) is given by

$$\rho\, \hat{\boldsymbol{n}} \cdot (\boldsymbol{v} - \boldsymbol{v}^{(b)}) = -\mathcal{Q}_{\mathrm{m}} \qquad \text{surface ocean boundary} \tag{15.52}$$

with $\mathcal{Q}_{\mathrm{m}}$ the mass per time per surface area crossing the boundary. We are thus led to the ocean surface boundary condition for trace matter

$$[\rho\, C\, (\boldsymbol{v} - \boldsymbol{v}^{(b)}) + \boldsymbol{J}] \cdot \hat{\boldsymbol{n}}\, \mathrm{d}\mathcal{S} = (C\, \mathcal{Q}_{\mathrm{m}} + \boldsymbol{J}) \cdot \hat{\boldsymbol{n}}\, \mathrm{d}\mathcal{S}. \tag{15.53}$$

The first term accounts for the transport of matter within the mass crossing the boundary. The second term accounts for transport through subgrid scale processes, such as turbulent air-sea boundary fluxes. In Section 50.3 we discuss the variety of salt and heat transports crossing the ocean boundary that contribute to changes in ocean buoyancy, most of which appear as boundary contributions to $\boldsymbol{J}$.

15.5 Brute force illustration of Leibniz-Reynolds

The Leibniz-Reynolds transport theorem

$$\frac{\mathrm{d}}{\mathrm{d}t}\left[\int_{\mathcal{R}} \rho\, C\, \mathrm{d}V\right] = -\int_{\partial\mathcal{R}} \left[\rho\, C\, (\boldsymbol{v} - \boldsymbol{v}^{(b)}) + \boldsymbol{J}\right] \cdot \hat{\boldsymbol{n}}\, \mathrm{d}\mathcal{S}, \tag{15.54}$$

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 15.3. Rather than make direct use of Leibniz-Reynolds, we use a brute force approach by expanding the volume integral according to

$$\frac{\mathrm{d}}{\mathrm{d}t}\left[\int_{\mathcal{R}} \rho\, C\, \mathrm{d}V\right] = \frac{\mathrm{d}}{\mathrm{d}t}\left[\int_{A(t)} \mathrm{d}A \int_{\eta_{\mathrm{b}}}^{\eta} \rho\, C\, \mathrm{d}z\right]. \tag{15.55}$$

In this equation, $\int_{A(t)} \mathrm{d}A$ is an integral over the horizontal area of the domain, with the lateral boundaries of the domain generally a function of time.

15.5.1 Leibniz's rule plus kinematic boundary conditions

Performing the time derivative in equation (15.55) and using Leibniz's rule yields

$$\frac{\mathrm{d}}{\mathrm{d}t}\left[\int_{\mathcal{R}} \rho\, C\, \mathrm{d}V\right] = \frac{\mathrm{d}A}{\mathrm{d}t}\left[\int_{-H}^{\eta} \rho\, C\, \mathrm{d}z\right]_{\text{bounds}} + \int_{A(t)} [\partial_t \eta\, (\rho\, C)_{z=\eta}]\, \mathrm{d}A + \int_{A(t)} \mathrm{d}A \int_{\eta_{\mathrm{b}}}^{\eta} \frac{\partial(\rho\, C)}{\partial t}\, \mathrm{d}z. \quad (15.56)$$

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 box of seawater or a periodic channel, then $\mathrm{d}A/\mathrm{d}t = 0$. The more general case has a 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, $\eta - \eta_{\mathrm{b}} = 0$, thus again revealing that the first term on the right hand side drops from the budget to render

$$\frac{\mathrm{d}}{\mathrm{d}t}\left[\int_{\mathcal{R}} \rho\, C\, \mathrm{d}V\right] = \int_{A(t)} [\partial_t \eta\, (\rho\, C)_{z=\eta}]\, \mathrm{d}A + \int_{A(t)} \mathrm{d}A \int_{\eta_{\mathrm{b}}}^{\eta} \frac{\partial(\rho\, C)}{\partial t}\, \mathrm{d}z. \quad (15.57)$$

For the second term on the right hand side of equation (15.57) we make use of the tracer equation (15.44) and Leibniz's rule to write

$$\int_{\eta_{\mathrm{b}}}^{\eta} \frac{\partial(\rho\, C)}{\partial t}\, \mathrm{d}z = -\int_{\eta_{\mathrm{b}}}^{\eta} \nabla_z \cdot (\rho\, C\, \boldsymbol{u} + \boldsymbol{J}_h)\, \mathrm{d}z - \int_{\eta_{\mathrm{b}}}^{\eta} \frac{\partial(\rho\, C\, w + J^z)}{\partial z}\, \mathrm{d}z \quad (15.58a)$$

$$= -\nabla_z \cdot \int_{-H}^{\eta} (\rho\, C\, \boldsymbol{u} + \boldsymbol{J}_h)\, \mathrm{d}z + \nabla(\eta - z) \cdot (\rho\, C\, \boldsymbol{v} + \boldsymbol{J})_{z=\eta} + \nabla(z - \eta_{\mathrm{b}}) \cdot (\rho\, C\, \boldsymbol{v} + \boldsymbol{J})_{z=\eta_{\mathrm{b}}} \quad (15.58b)$$

where we wrote $\boldsymbol{J} = \boldsymbol{J}_h + \hat{\boldsymbol{z}}\, J^z$. The surface terms ($z = \eta$) combine with the $\partial_t \eta$ term appearing in equation (15.57) to yield

$$\rho\, C \left[\frac{\partial \eta}{\partial t} + \boldsymbol{u} \cdot \nabla \eta - w\right] = C\, Q_{\mathrm{m}}, \quad (15.59)$$

where we used the surface kinematic boundary condition (14.73) to introduce the surface boundary mass flux Q_{m}. The bottom kinematic boundary condition eliminates the advective contribution at the bottom, $z = \eta_{\mathrm{b}}$, via the no normal flow condition (14.38)

$$\nabla(z - \eta_{\mathrm{b}}) \cdot \boldsymbol{v} = 0. \quad (15.60)$$

Finally, when integrated over the horizontal extent of the domain, the horizontal convergence term from equation (15.58b) 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.

15.5.2 Summarizing the result

Bringing the results together yields the budget equation

$$\frac{\mathrm{d}}{\mathrm{d}t}\left[\int_{\mathcal{R}} \rho\, C\, \mathrm{d}V\right] = \int_{z=\eta} (C\, Q_{\mathrm{m}} + \nabla(\eta - z) \cdot \boldsymbol{J})\, \mathrm{d}A + \int_{z=-H} \nabla(z - \eta_{\mathrm{b}}) \cdot \boldsymbol{J}\, \mathrm{d}A. \quad (15.61)$$

We now use the identity (14.65e) between horizontal area element, $\mathrm{d}A = \mathrm{d}x\,\mathrm{d}y$, and area element on the surface

$$\hat{\boldsymbol{n}}\,\mathrm{d}\mathcal{S} = -\nabla(\eta - z)\,\mathrm{d}A \qquad \text{at } z = \eta \tag{15.62a}$$

$$\hat{\boldsymbol{n}}\,\mathrm{d}\mathcal{S} = -\nabla(z - x\eta_{\mathrm{b}})\,\mathrm{d}A \qquad \text{at } z = \eta_{\mathrm{b}}, \tag{15.62b}$$

to write

$$\frac{\mathrm{d}}{\mathrm{d}t}\left[\int_{\mathcal{R}} \rho\, C\,\mathrm{d}V\right] = -\int_{z=\eta} (-C\,\mathcal{Q}_{\mathrm{m}} + \hat{\boldsymbol{n}}\cdot\boldsymbol{J})\mathrm{d}\mathcal{S} - \int_{z=\eta_{\mathrm{b}}} \hat{\boldsymbol{n}}\cdot\boldsymbol{J}\,\mathrm{d}\mathcal{S}, \tag{15.63}$$

where

$$Q_{\mathrm{m}}\,\mathrm{d}A = \mathcal{Q}_{\mathrm{m}}\,\mathrm{d}\mathcal{S} = -\rho\,(\boldsymbol{v} - \boldsymbol{v}^{(\eta)})\cdot\hat{\boldsymbol{n}}\,\mathrm{d}\mathcal{S} \tag{15.64}$$

according to equation (14.70c). The budget for fluid mass is realized by setting C to a constant and thus dropping the subgrid flux

$$\frac{\mathrm{d}}{\mathrm{d}t}\left[\int_{\mathcal{R}} \rho\,\mathrm{d}V\right] = \int_{z=\eta} \mathcal{Q}_{\mathrm{m}}\,\mathrm{d}\mathcal{S} = -\int_{z=\eta} \rho\,(\boldsymbol{v} - \boldsymbol{v}^{(\eta)})\cdot\hat{\boldsymbol{n}}\,\mathrm{d}\mathcal{S}. \tag{15.65}$$

The manipulations in this section have succeeded in bringing the tracer and mass budgets into the form of the Leibniz-Reynolds transport theorem (15.54). The process of doing so required far more tedium as compared to the elegance of merely starting from equation (15.54). Even so, our efforts provide a useful means to ground the formalism by unpacking the many steps summarized by Leibniz-Reynolds.

15.6 Comments on local conservation laws

We close this chapter by offering some general comments on conservation laws for field quantities, such as tracers, that satisfy Eulerian flux-form conservations laws in the form

$$\frac{\partial(\rho\, C)}{\partial t} = -\nabla\cdot(\rho\,\boldsymbol{v}\,C + \boldsymbol{J}). \tag{15.66}$$

In this chapter we saw how this differential equation leads to finite volume conservation properties for the mass, $\int_{\mathcal{R}} C\,\rho\,\mathrm{d}V$, of C-matter within the region $\mathcal{R}$. This result is a consequence of the local conservation of matter within the fluid. That is, the amount of matter changes at a point only through the local convergence of matter fluxes onto that point. There is no matter evolution arising from a non-local transport or a source/sink term that cannot be expressed as the convergence of a flux. The presence of such terms might be relevant for chemical or biogeochemical reactions, whereby matter is converted to other forms. But such processes cannot be represented as the convergence of a flux, and as such they are not contained in the conservation law (15.66). We sometimes refer to such non-flux convergent processes as a *non-conservative process*.

Conservation laws of the form (15.66) recur in our discussion of other properties in this book, such as mechanical energy (Chapter 18), absolute vorticity (Section 37.1.3) and potential vorticity (Chapter 36). These laws are consistent with basic notions of causality and locality that appear throughout physics. A lucid discussion of such conservation laws is offered in Section 27-1 of the Feynman Lectures Volume II.

16

Incompressible flow

In this chapter, we specialize the general kinematics from Chapters 12–14 to the case of an incompressible fluid. The velocity field for an incompressible fluid has zero divergence so that it can be written as the curl of a vector streamfunction. The streamfunction plays a central role for incompressible fluid kinematics.

READER'S GUIDE TO THIS CHAPTER

We presume an understanding of the kinematics of mass conservation from Chapter 14. This chapter introduces many concepts and tools of use in the remainder of the book.

16.1 Introduction to incompressible flow **269**
16.2 Kinematic boundary conditions **270**
16.3 Kinematic free surface equation **270**
16.4 Streamfunction for two-dimensional flow **271**
16.4.1 Streamfunction isolines are streamlines 272
16.4.2 Streamfunction is constant on material boundaries 272
16.4.3 The transport streamfunction 272
16.4.4 Gauge symmetry 273
16.4.5 Exact differential formulation 273
16.4.6 Concerning the Helmholz decomposition 274
16.5 Vector streamfunction for three-dimensional flow **275**
16.5.1 The transport streamfunction 275
16.5.2 Gauge symmetry 276
16.5.3 Transport streamfunction in the vertical gauge 276
16.5.4 Concerning a harmonic velocity potential 277
16.6 Evolution of finite volume and area **277**
16.6.1 Material volumes and areas 277
16.6.2 Arbitrary volume and area 278
16.7 Meridional-depth overturning circulation **278**
16.8 Exercises **280**

16.1 Introduction to incompressible flow

For many applications in geophysical fluid mechanics, we can make a simplifying assumption regarding the fluid kinematics. For the ocean, the Boussinesq approximation is well maintained (see

Chapter 25), whereby the volume of a fluid element is constant. Recalling the expression

$$\frac{1}{\delta V}\frac{\mathrm{D}(\delta V)}{\mathrm{D}t} = \nabla\cdot\boldsymbol{v} \tag{16.1}$$

from Section 13.4.1, we see that a constant volume for a fluid element constrains the velocity field to be non-divergent

$$\frac{1}{\delta V}\frac{\mathrm{D}(\delta V)}{\mathrm{D}t} = 0 \Longrightarrow \nabla\cdot\boldsymbol{v} = 0 \qquad \text{incompressible.} \tag{16.2}$$

A slightly less onerous constraint arises from the anelastic approximation, whereby

$$\nabla\cdot(\rho\,\boldsymbol{v}) = 0. \tag{16.3}$$

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.

The non-divergence constraint (16.2) 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.

16.2 Kinematic boundary conditions

For incompressible flow, there are slight modifications to the compressible boundary conditions detailed in Section 14.4. 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 (14.60) to a volume transport condition

$$\rho_0\,(\boldsymbol{v}-\boldsymbol{v}^{(s)})\cdot\hat{\boldsymbol{n}}\,\mathrm{d}\mathcal{S} = -\mathcal{Q}_{\mathrm{m}}\,\mathrm{d}\mathcal{S} \qquad \text{moving non-material boundary condition.} \tag{16.4}$$

Correspondingly, the kinematic boundary condition (14.73) applied at the ocean free surface takes on the form

$$\rho_0\,\frac{\mathrm{D}(z-\eta)}{\mathrm{D}t} = -Q_{\mathrm{m}} \Rightarrow w + \frac{Q_{\mathrm{m}}}{\rho_0} = \frac{\partial\eta}{\partial t} + \boldsymbol{u}\cdot\nabla\eta. \tag{16.5}$$

16.3 Kinematic free surface equation

We now derive an equation for the volume budget over a column of fluid. This equation provides a kinematic expression for the free surface evolution in an incompressible fluid. For this purpose, we vertically integrate the incompressibility constraint, $\nabla\cdot\boldsymbol{v}=0$, over the depth of an ocean column, from $z=\eta_{\mathrm{b}}(x,y)$ 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

$$0 = \int_{\eta_{\mathrm{b}}}^{\eta} \nabla\cdot\boldsymbol{v}\,\mathrm{d}z \tag{16.6a}$$

$$= w(\eta) - w(\eta_{\mathrm{b}}) + \int_{\eta_{\mathrm{b}}}^{\eta} \nabla\cdot\boldsymbol{u}\,\mathrm{d}z \tag{16.6b}$$

$$= w(\eta) - w(\eta_{\mathrm{b}}) + \nabla\cdot\left[\int_{\eta_{\mathrm{b}}}^{\eta}\boldsymbol{u}\,\mathrm{d}z\right] - \boldsymbol{u}(\eta)\cdot\nabla\eta + \boldsymbol{u}(\eta_{\mathrm{b}})\cdot\nabla\eta_{\mathrm{b}} \tag{16.6c}$$

$$= [w(\eta) - \boldsymbol{u}(\eta)\cdot\nabla\eta] - [w(\eta_{\mathrm{b}}) - \boldsymbol{u}(\eta_{\mathrm{b}})\cdot\nabla\eta_{\mathrm{b}}] + \nabla\cdot\left[\int_{\eta_{\mathrm{b}}}^{\eta}\boldsymbol{u}\,\mathrm{d}z\right], \tag{16.6d}$$

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 (16.5) and the bottom no-flow condition

$$w(\eta) - \boldsymbol{u} \cdot \nabla \eta = -\frac{Q_{\mathrm{m}}}{\rho_0} + \frac{\partial \eta}{\partial t} \qquad z = \eta \tag{16.7a}$$

$$w = -\boldsymbol{u} \cdot \nabla H \qquad z = \eta_{\mathrm{b}} \tag{16.7b}$$

to render the free surface equation for an incompressible fluid

$$\frac{\partial \eta}{\partial t} = \frac{Q_{\mathrm{m}}}{\rho_0} - \nabla \cdot \boldsymbol{U}, \tag{16.8}$$

where

$$\boldsymbol{U} = \int_{\eta_{\mathrm{b}}}^{\eta} \boldsymbol{u} \, \mathrm{d}z \tag{16.9}$$

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

$$\nabla \cdot \boldsymbol{U} = 0 \qquad \text{if } Q_{\mathrm{m}} = 0 \text{ and } \partial \eta / \partial t = 0. \tag{16.10}$$

Comparing the incompressible free surface quation (16.8) to the compressible free surface equation (14.78) 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 an incompressible 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 (incompressible) ocean models.

16.4 Streamfunction for two-dimensional flow

Vertical stratification of buoyancy and 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

$$\boldsymbol{u} = \hat{\boldsymbol{z}} \wedge \nabla \psi = -\hat{\boldsymbol{x}} \, \frac{\partial \psi}{\partial y} + \hat{\boldsymbol{y}} \, \frac{\partial \psi}{\partial x}, \tag{16.11}$$

where

$$\nabla_z = \hat{\boldsymbol{x}} \, \partial_x + \hat{\boldsymbol{y}} \, \partial_y \tag{16.12}$$

is the horizontal gradient operator. The constraint $\nabla_z \cdot \boldsymbol{u} = 0$ is satisfied since the partial derivative operators commute

$$\frac{\partial^2 \psi}{\partial x \partial y} = \frac{\partial^2 \psi}{\partial y \partial x}. \tag{16.13}$$

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 16.5.

16.4.1 Streamfunction isolines are streamlines

At any fixed time, the total differential of the streamfunction is

$$\mathrm{d}\psi = \frac{\partial \psi}{\partial x}\,\mathrm{d}x + \frac{\partial \psi}{\partial y}\,\mathrm{d}y = v\,\mathrm{d}x - u\,\mathrm{d}y, \tag{16.14}$$

where the second equality follows from equation (16.11). Instantaneous lines along which ψ is a constant satisfy

$$\mathrm{d}\psi = 0 \Longrightarrow \frac{\mathrm{d}x}{u} = \frac{\mathrm{d}y}{v}. \tag{16.15}$$

Furthermore, the normal direction to constant ψ lines

$$\hat{\boldsymbol{n}} = \frac{\nabla_z \psi}{|\nabla_z \psi|} = \frac{v\,\hat{\boldsymbol{x}} - u\,\hat{\boldsymbol{y}}}{|\boldsymbol{u}|} \tag{16.16}$$

is normal to the velocity

$$\boldsymbol{u}\cdot\nabla_z\psi = u\,v - v\,u = 0. \tag{16.17}$$

Consequently, at each time instance, lines of constant ψ are streamlines (see Section 12.8.2 for discussion of streamlines). This property in turn motivates the name *streamfunction*. Furthermore, through each point of a two-dimensional non-divergent fluid and at any particular instance in time, there is one and only one streamline passing through that point.

16.4.2 Streamfunction is constant on material boundaries

As a corollary to the results from Section 16.4.1, we know that the streamfunction is a spatial constant when evaluated along material boundaries where $\boldsymbol{u}\cdot\hat{\boldsymbol{n}} = 0$. This property follows from equation (16.17). We can also see it from

$$0 = \boldsymbol{u}\cdot\hat{\boldsymbol{n}} = (\hat{\boldsymbol{z}}\wedge\nabla_z\psi)\cdot\hat{\boldsymbol{n}} = (\hat{\boldsymbol{n}}\wedge\hat{\boldsymbol{z}})\cdot\nabla_z\psi = \hat{\boldsymbol{t}}\cdot\nabla_z\psi, \tag{16.18}$$

where $\hat{\boldsymbol{t}}$ a unit vector pointing tangent to the boundary. The operator $\hat{\boldsymbol{t}}\cdot\nabla_z\psi$ is the derivative of ψ computed along the boundary tangent at any given point along the boundary. Hence, $\hat{\boldsymbol{t}}\cdot\nabla_z\psi = 0$ means that ψ is a spatial constant along the boundary. Even though spatially constant, ψ along the boundary is generally a function of time.

16.4.3 The transport streamfunction

Consider an arbitrary curve in the fluid with endpoints $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ as depicted in Figure 16.1. At any particular time instance, the difference in streamfunction between these two points is given by

$$\psi(\boldsymbol{x}_1) - \psi(\boldsymbol{x}_2) = \int_{\boldsymbol{x}_1}^{\boldsymbol{x}_2} \mathrm{d}\psi = \int_{\boldsymbol{x}_1}^{\boldsymbol{x}_2} \left[\mathrm{d}x\,\frac{\partial\psi}{\partial x} + \mathrm{d}y\,\frac{\partial\psi}{\partial y}\right] = \int_{\boldsymbol{x}_1}^{\boldsymbol{x}_2} \nabla_z\psi\cdot\mathrm{d}\boldsymbol{x} = \int_{\boldsymbol{x}_1}^{\boldsymbol{x}_2} \nabla_z\psi\cdot\hat{\boldsymbol{t}}\,\mathrm{d}s. \tag{16.19}$$

For the final equality we wrote

$$\mathrm{d}\boldsymbol{x} = \hat{\boldsymbol{t}}\,\mathrm{d}s, \tag{16.20}$$

where

$$\mathrm{d}s = |\mathrm{d}\boldsymbol{x}| \tag{16.21}$$

is the element of arc-length along the curve, and $\hat{\boldsymbol{t}}$ is a unit tangent vector that points in the direction along the curve from $\boldsymbol{x}_1$ to $\boldsymbol{x}_2$. Now introduce the normal vector along the curve according to

$$\hat{\boldsymbol{t}} = \hat{\boldsymbol{n}}\wedge\hat{\boldsymbol{z}}, \tag{16.22}$$

Figure 16.1: *Depicting the transport between two points in a two-dimensional fluid. The transport is the line integral, $\int_{\boldsymbol{x}_1}^{\boldsymbol{x}_2} \boldsymbol{u} \cdot \hat{\boldsymbol{n}} \, \mathrm{d}s$, from point $\boldsymbol{x}_1$ to $\boldsymbol{x}_2$, with the normal direction $\hat{\boldsymbol{n}}$ pointing to the left when facing in the direction of the local tangent vector, $\hat{\boldsymbol{t}}$. By construction, $\hat{\boldsymbol{t}} \wedge \hat{\boldsymbol{n}} = \hat{\boldsymbol{z}}$, where $\hat{\boldsymbol{z}}$ points vertically out of the page. For a two-dimensionally incompressible flow, $\nabla_z \cdot \boldsymbol{u} = 0$, the transport between any two points is the difference in the transport streamfunction computed at these points, $\int_{\boldsymbol{x}_1}^{\boldsymbol{x}_2} \boldsymbol{u} \cdot \hat{\boldsymbol{n}} \, \mathrm{d}s = \psi(\boldsymbol{x}_1) - \psi(\boldsymbol{x}_2)$. 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.*

which points to the left when facting in the $\hat{\boldsymbol{t}}$ direction. We thus have

$$\psi(\boldsymbol{x}_1) - \psi(\boldsymbol{x}_2) = \int_{\boldsymbol{x}_1}^{\boldsymbol{x}_2} \nabla_z \psi \cdot (\hat{\boldsymbol{n}} \wedge \hat{\boldsymbol{z}}) \, \mathrm{d}s = \int_{\boldsymbol{x}_1}^{\boldsymbol{x}_2} (\hat{\boldsymbol{z}} \wedge \nabla_z \psi) \cdot \hat{\boldsymbol{n}} \, \mathrm{d}s = \int_{\boldsymbol{x}_1}^{\boldsymbol{x}_2} \boldsymbol{u} \cdot \hat{\boldsymbol{n}} \, \mathrm{d}s, \tag{16.23}$$

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

$$|\boldsymbol{u}| = |\nabla_z \psi|. \tag{16.24}$$

Given this connection between ψ and the integrated transport, we sometimes refer to ψ as the *transport streamfunction.*

16.4.4 Gauge symmetry

For a two-dimensional non-divergent flow, the constraint $\nabla_z \cdot \boldsymbol{u} = 0$ 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

$$\psi' = \psi + k \Rightarrow \boldsymbol{u}' = \boldsymbol{u}. \tag{16.25}$$

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 16.5.2.

16.4.5 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

$$\boldsymbol{A} \cdot \mathrm{d}\boldsymbol{x} \equiv (\boldsymbol{u} \wedge \hat{\boldsymbol{z}}) \cdot \mathrm{d}\boldsymbol{x} = v \, \mathrm{d}x - u \, \mathrm{d}y. \tag{16.26}$$

By construction

$$\hat{\boldsymbol{z}} \cdot (\nabla_z \wedge \boldsymbol{A}) = 0 \qquad \text{since} \qquad \nabla_z \cdot \boldsymbol{u} = 0, \tag{16.27}$$

which means that $\boldsymbol{A} \cdot \mathrm{d}\boldsymbol{x}$ is an exact differential (Section 2.8.1). Consequently, we can write

$$\boldsymbol{A} \cdot \mathrm{d}\boldsymbol{x} = (\boldsymbol{u} \wedge \hat{\boldsymbol{z}}) \cdot \mathrm{d}\boldsymbol{x} = \nabla_z \psi \cdot \mathrm{d}\boldsymbol{x} = \mathrm{d}\psi, \tag{16.28}$$

which then leads to the results derived earlier in this section where ψ is the transport streamfunction.

16.4.6 Concerning the Helmholz decomposition

We close our discussion in this section by tidying up some mathematical niceties concerning the Helmholz decomposition introduced in Section 2.9. For two-dimensional flows the decomposition takes the form

$$\boldsymbol{u} = \hat{\boldsymbol{z}} \wedge \nabla_z \Gamma + \nabla_z \Phi, \tag{16.29}$$

for some functions Γ and Φ. For non-divergent flows, Φ is constrained to be harmonic[1]

$$\nabla \cdot \boldsymbol{u} = 0 \Longrightarrow \nabla_z^2 \Phi = \nabla_z \cdot \nabla_z \Phi = 0. \tag{16.30}$$

As summarized in Table 16.1, it is sufficient to make use of just a transport 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 Helmholz decomposition for non-divergent two-dimensional flows.

NON-DIVERGENT VORTICAL FLOW	NON-DIVERGENT IRROTATIONAL FLOW
$\nabla_z \cdot \boldsymbol{u} = 0$	$\nabla_z \cdot \boldsymbol{u} = 0$
$\nabla_z \wedge \boldsymbol{u} \neq 0$	$\nabla_z \wedge \boldsymbol{u} = 0$
$\boldsymbol{u} = \hat{\boldsymbol{z}} \wedge \nabla_z \psi$	$\boldsymbol{u} = \nabla_z \phi$
$\hat{\boldsymbol{z}} \cdot (\nabla_z \wedge \boldsymbol{u}) = \nabla_z^2 \psi$	$\nabla_z^2 \phi = 0.$

Table 16.1: *Summarizing some mathematical properties of non-divergent two-dimensional velocity fields, $\nabla \cdot \boldsymbol{u} = 0$. The streamfunction is ψ and ϕ is the harmonic velocity potential.*

Non-divergent vortical flow

Return to the exact differential formulation from Section 16.4.5. In that formulation we noted that $\nabla_z \cdot \boldsymbol{u} = 0$ means that the differential $\boldsymbol{A} \cdot \mathrm{d}\boldsymbol{x} = (\boldsymbol{u} \wedge \hat{\boldsymbol{z}}) \cdot \mathrm{d}\boldsymbol{x}$ is exact. Making use of the Helmholz decomposition (16.29) renders

$$\boldsymbol{A} \cdot \mathrm{d}\boldsymbol{x} = (\boldsymbol{u} \wedge \hat{\boldsymbol{z}}) \cdot \mathrm{d}\boldsymbol{x} \tag{16.31a}$$

$$= [(\hat{\boldsymbol{z}} \wedge \nabla_z \Gamma) \wedge \hat{\boldsymbol{z}} + \nabla_z \Phi \wedge \hat{\boldsymbol{z}}] \cdot \mathrm{d}\boldsymbol{x} \tag{16.31b}$$

$$= [\nabla_z \Gamma + \nabla_z \Phi \wedge \hat{\boldsymbol{z}}] \cdot \mathrm{d}\boldsymbol{x}. \tag{16.31c}$$

To reveal the exactness of the right hand side requires the harmonic property of Φ so that we can write

$$\hat{\boldsymbol{z}} \cdot [\nabla_z \wedge (\nabla_z \Phi \wedge \hat{\boldsymbol{z}})] = -\nabla_z^2 \Phi = 0 \Longrightarrow \nabla_z \Phi \wedge \hat{\boldsymbol{z}} = \nabla_z \Psi, \tag{16.32}$$

in which case

$$\boldsymbol{A} \cdot \mathrm{d}\boldsymbol{x} \equiv v \, \mathrm{d}x - u \, \mathrm{d}y = \mathrm{d}(\Gamma + \Psi) \equiv \mathrm{d}\psi. \tag{16.33}$$

We conclude that for non-divergent vortical flow, we lose no generality by working just with the transport streamfunction, ψ, of Section 16.4.3. There is no need to also include a harmonic function.

[1] Recall our discussion of harmonic functions in Sections 2.2.2 and 3.4.1.

[2] The vorticity, $\nabla_z \wedge \boldsymbol{u}$, is a measure of the spin in the fluid and is the focus of Part VI of this book.

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

$$\nabla_z \wedge (\hat{\boldsymbol{z}} \wedge \nabla_z \Gamma) = \nabla_z^2 \Gamma = 0. \tag{16.34}$$

Consequently, we can write

$$\hat{\boldsymbol{z}} \wedge \nabla_z \Gamma = \nabla_z \gamma, \tag{16.35}$$

in which case

$$\boldsymbol{u} = \hat{\boldsymbol{z}} \wedge \nabla_z \Gamma + \nabla_z \Phi = \nabla_z (\gamma + \Phi) \equiv \nabla_z \phi. \tag{16.36}$$

Hence, for non-divergent irrotational flow, it is sufficient to work just with the harmonic velocity potential, ϕ.

16.5 Vector streamfunction for three-dimensional flow

A three-dimensional non-divergent velocity, $\nabla \cdot \boldsymbol{v} = 0$, can be specified by a vector streamfunction

$$\boldsymbol{v} = \nabla \wedge \boldsymbol{\Psi}. \tag{16.37}$$

The constraint $\nabla \cdot \boldsymbol{v} = 0$ is trivially satisfied since the divergence of the curl vanishes

$$\nabla \cdot (\nabla \wedge \boldsymbol{\Psi}) = 0. \tag{16.38}$$

16.5.1 The transport streamfunction

In a three-dimensional fluid, the volume transport (volume per time) of fluid crossing a surface is defined by the area integral

$$\mathcal{T}(\mathcal{S}) = \int_{\mathcal{S}} \boldsymbol{v} \cdot \hat{\boldsymbol{n}} \, \mathrm{d}\mathcal{S}, \tag{16.39}$$

where $\hat{\boldsymbol{n}}$ 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

$$\mathcal{T}(\mathcal{S}) = \int_{\mathcal{S}} \boldsymbol{v} \cdot \hat{\boldsymbol{n}} \, \mathrm{d}\mathcal{S} = \int_{\mathcal{S}} (\nabla \wedge \boldsymbol{\Psi}) \cdot \hat{\boldsymbol{n}} \, \mathrm{d}\mathcal{S} = \oint_{\partial \mathcal{S}} \boldsymbol{\Psi} \cdot \hat{\boldsymbol{t}} \, \mathrm{d}s, \tag{16.40}$$

where $\hat{\boldsymbol{t}} \, \mathrm{d}s$ 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 transport 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, $\boldsymbol{\Psi} = \nabla \chi$, since

$$\oint_{\partial \mathcal{S}} \boldsymbol{\Psi} \cdot \hat{\boldsymbol{t}} \, \mathrm{d}s = \oint_{\partial \mathcal{S}} \nabla \chi \cdot \hat{\boldsymbol{t}} \, \mathrm{d}s = \oint_{\partial \mathcal{S}} \nabla \chi \cdot \mathrm{d}\boldsymbol{x} = \oint_{\partial \mathcal{S}} \mathrm{d}\chi = 0. \tag{16.41}$$

Because $\boldsymbol{\Psi}$ has a connection to fluid transport, we refer to it as the *transport streamfunction*, just as we did for the streamfunction ψ in two-dimensional non-divergent flows (Section 16.4.3).

16.5.2 Gauge symmetry

For three-dimensional non-divergent flow, the constraint $\nabla \cdot \boldsymbol{v} = 0$ reduces the three functional degrees of freedom down to the two available from the vector streamfunction $\boldsymbol{v} = \nabla \wedge \boldsymbol{\Psi}$. Gauge symmetry manifests through the ability to add the gradient of an arbitrary function to $\boldsymbol{\Psi}$ without altering $\boldsymbol{v}$:

$$\boldsymbol{\Psi}' = \boldsymbol{\Psi} + \nabla\lambda \Rightarrow \boldsymbol{v}' = \boldsymbol{v}, \tag{16.42}$$

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 16.5.1. One particularly useful example of a transport streamfunction is that obtained by making the vertical gauge as discussed next.

16.5.3 Transport 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

$$u = \frac{\partial \Psi_3}{\partial y} - \frac{\partial \Psi_2}{\partial z} \qquad \text{and} \qquad v = \frac{\partial \Psi_1}{\partial z} - \frac{\partial \Psi_3}{\partial x} \qquad \text{and} \qquad w = \frac{\partial \Psi_2}{\partial x} - \frac{\partial \Psi_1}{\partial y}, \tag{16.43}$$

where the Cartesian components to the vector streamfunction are

$$\boldsymbol{\Psi} = \hat{\boldsymbol{x}}\, \Psi_1 + \hat{\boldsymbol{y}}\, \Psi_2 + \hat{\boldsymbol{z}}\, \Psi_3. \tag{16.44}$$

We follow studies of ocean mesoscale eddy parameterizations (e.g., see Sections 47.6.1 and 49.1) by choosing the *vertical gauge* whereby

$$\Psi_3 = 0 \Longleftrightarrow \text{vertical gauge} \tag{16.45}$$

so that

$$u = -\frac{\partial \Psi_2}{\partial z} \qquad \text{and} \qquad v = \frac{\partial \Psi_1}{\partial z} \qquad \text{and} \qquad w = \frac{\partial \Psi_2}{\partial x} - \frac{\partial \Psi_1}{\partial y}. \tag{16.46}$$

Vertically integrating the u, v equations from the bottom at $z = \eta_{\mathrm{b}}(x, y)$ up to an arbitrary geopotential leads to

$$\boldsymbol{\Psi} = \int_{\eta_{\mathrm{b}}}^{z} \boldsymbol{u}\, \mathrm{d}z' \wedge \hat{\boldsymbol{z}} \equiv \boldsymbol{U}(z) \wedge \hat{\boldsymbol{z}}, \tag{16.47}$$

where $\boldsymbol{U}(z) = \int_{\eta_{\mathrm{b}}}^{z} \boldsymbol{u}\, \mathrm{d}z'$ is the horizontal transport of fluid beneath a chosen geopotential. 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

$$\frac{\partial \Psi_2}{\partial x} = -\int_{\eta_{\mathrm{b}}}^{z} \partial_x u\, \mathrm{d}z' + u(\eta_{\mathrm{b}})\, \partial_x \eta_{\mathrm{b}} \qquad \text{and} \qquad \frac{\partial \Psi_1}{\partial y} = \int_{\eta_{\mathrm{b}}}^{z} \partial_y v\, \mathrm{d}z' - v(\eta_{\mathrm{b}})\, \partial_y \eta_{\mathrm{b}}, \tag{16.48}$$

so that

$$\frac{\partial \Psi_2}{\partial x} - \frac{\partial \Psi_1}{\partial y} = -\int_{\eta_{\mathrm{b}}}^{z} \left[\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y}\right] \mathrm{d}z' + \boldsymbol{u}(\eta_{\mathrm{b}}) \cdot \nabla\eta_{\mathrm{b}} = \int_{\eta_{\mathrm{b}}}^{z} \frac{\partial w}{\partial z'}\, \mathrm{d}z' + w(\eta_{\mathrm{b}}) = w(z), \tag{16.49}$$

where we used the bottom kinematic boundary condition (14.38) to write $w - \boldsymbol{u} \cdot \nabla\eta_{\mathrm{b}} = 0$ at $z = \eta_{\mathrm{b}}(x, y)$. Hence, knowledge of the vector streamfunction (16.47) 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 $z = \eta_{\mathrm{b}}(x, y)$ vanishes according to equation (16.40). We can trivially verify this result for the vertical gauge since

$$\boldsymbol{\Psi}(z = \eta_{\mathrm{b}}) = 0, \tag{16.50}$$

so that $\oint_{\partial\mathcal{S}} \boldsymbol{\Psi} \cdot \hat{\boldsymbol{t}}\, \mathrm{d}s = 0$ on the bottom.

16.5.4 Concerning a harmonic velocity potential

As for the two-dimensional case discussed in Section 16.4.6, we consider the relevance of an arbitrary harmonic velocity potential, χ, so that the velocity takes the form

$$\boldsymbol{v} = \nabla \wedge \boldsymbol{\Gamma} + \nabla\chi \qquad \text{with} \qquad \nabla^2\chi = 0. \tag{16.51}$$

Since $\nabla \cdot \nabla\chi = 0$ we can write $\nabla\chi$ as the curl of another vector

$$\nabla \cdot \nabla\chi = 0 \Longrightarrow \nabla\chi = \nabla \wedge \boldsymbol{\Lambda}, \tag{16.52}$$

in which case the velocity takes the form

$$\boldsymbol{v} = \nabla \wedge \boldsymbol{\Gamma} + \nabla\phi = \nabla \wedge (\boldsymbol{\Gamma} + \boldsymbol{\Lambda}) \equiv \nabla \wedge \boldsymbol{\Psi}. \tag{16.53}$$

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.

16.6 Evolution of finite volume and area

In this section we develop kinematic equations for the evolution of volume and area within an incompressible fluid. We start by considering a material region and then study an arbitrary region.

16.6.1 Material volumes and areas

As shown by equation (16.2), the volume of a material parcel remains materially constant in an incompressible flow. Correspondingly, a material fluid region maintains a constant volume

$$\frac{\mathrm{d}}{\mathrm{d}t}\int_{\mathcal{R}(\boldsymbol{v})} \mathrm{d}V = \int_{\mathcal{R}(\boldsymbol{v})} \frac{\mathrm{D}(\delta V)}{\mathrm{D}t} = \int_{\mathcal{R}(\boldsymbol{v})} (\nabla \cdot \boldsymbol{v})\, \mathrm{d}V = 0. \tag{16.54}$$

Note the appearance of a material time derivative on the inside of the integral arises since the integral is following material fluid particles (see Section 15.3.6). Likewise, following from the material area element equation (13.43), the area of a material region in a two-dimensional incompressible flow remains materially constant

$$\frac{\mathrm{d}}{\mathrm{d}t}\int_{\mathcal{S}(\boldsymbol{v})} \mathrm{d}\mathcal{S} = \int_{\mathcal{S}(\boldsymbol{v})} \frac{\mathrm{D}(\delta \mathcal{S})}{\mathrm{D}t} = \int_{\mathcal{S}(\boldsymbol{v})} (\nabla \cdot \boldsymbol{u})\, \mathrm{d}\mathcal{S} = 0. \tag{16.55}$$

This area preservation property is illustrated in Figure 16.2, in which a two-dimensional flow is seen to deform a black/white grid, yet to retain a fixed area.

Figure 16.2: *This figure is taken from Figure 2 of Welander (1955). It shows the deformation and rotation of a material area in two-dimensional non-divergent flow (left panel) after 6, 12, 24, and 36 hours from a model simulation. As discussed in Section 16.6, the area of a material region remains fixed in two-dimensional non-divergent flow.*

16.6.2 Arbitrary volume and area

We make use of the Leibniz-Reynolds transport theorem from Section 15.3.4 to develop the evolution equation for the volume of an arbitrary region. In particular equation (15.34) gives

$$\frac{\mathrm{d}}{\mathrm{d}t}\left[\int_{\mathcal{R}} \mathrm{d}V\right] = \oint_{\partial\mathcal{R}} \boldsymbol{v}^{(b)} \cdot \hat{\boldsymbol{n}}\, \mathrm{d}\mathcal{S}. \tag{16.56}$$

This result holds for both compressible and incompressible flows. But for incompressible flows we can go one step further by noting that

$$0 = \int_{\mathcal{R}} \nabla \cdot \boldsymbol{v}\, \mathrm{d}V = \oint_{\partial\mathcal{R}} \boldsymbol{v} \cdot \hat{\boldsymbol{n}}\, \mathrm{d}\mathcal{S}. \tag{16.57}$$

Importantly, this result holds only when integrating around the boundary of the full closed volume, $\mathcal{R}$. It does not necessarily mean that $\boldsymbol{v} \cdot \hat{\boldsymbol{n}} = 0$ holds at every point along the boundary. Indeed, when the boundary is time dependent $\boldsymbol{v} \cdot \hat{\boldsymbol{n}} \neq 0$ generally holds everywhere along the boundary.

Making use of equation (16.57) allows us to write

$$\frac{\mathrm{d}}{\mathrm{d}t}\left[\int_{\mathcal{R}} \mathrm{d}V\right] = \oint_{\partial\mathcal{R}} \boldsymbol{v}^{(b)} \cdot \hat{\boldsymbol{n}}\, \mathrm{d}\mathcal{S} = -\oint_{\partial\mathcal{R}} (\boldsymbol{v} - \boldsymbol{v}^{(b)}) \cdot \hat{\boldsymbol{n}}\, \mathrm{d}\mathcal{S}. \tag{16.58}$$

This result is identical to the mass budget equation (15.36) for the special case of a constant reference density appropriate for an incompressible fluid. The dia-surface transport, $(\boldsymbol{v}-\boldsymbol{v}^{(b)})\cdot\hat{\boldsymbol{n}}\, \mathrm{d}\mathcal{S}$, 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 (16.4).

16.7 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

$$V^\rho = \int_{x_1}^{x_2} \rho\, v \,\mathrm{d}x \qquad \text{and} \qquad W^\rho = \int_{x_1}^{x_2} \rho\, w \,\mathrm{d}x. \tag{16.59}$$

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.

To see how to create an overturning streamfunction, consider the zonal integrated area transport for an incompressible fluid[3]

$$V = \int_{x_1}^{x_2} v \,\mathrm{d}x \qquad \text{and} \qquad W = \int_{x_1}^{x_2} w \,\mathrm{d}x. \tag{16.60}$$

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

$$\frac{\partial V}{\partial y} = \frac{\partial}{\partial y}\left[\int_{x_1}^{x_2} v \,\mathrm{d}x\right] \tag{16.61a}$$

$$= \int_{x_1}^{x_2} \frac{\partial v}{\partial y} \,\mathrm{d}x \tag{16.61b}$$

$$= -\int_{x_1}^{x_2} \left[\frac{\partial u}{\partial x} + \frac{\partial w}{\partial z}\right] \mathrm{d}x \tag{16.61c}$$

$$= -\int_{x_1}^{x_2} \frac{\partial w}{\partial z} \,\mathrm{d}x \tag{16.61d}$$

$$= -\frac{\partial}{\partial z}\left[\int_{x_1}^{x_2} w \,\mathrm{d}x\right] \tag{16.61e}$$

$$= -\frac{\partial W}{\partial z}. \tag{16.61f}$$

These results show that the two-dimensional zonally integrated transport is non-divergent

$$\frac{\partial V}{\partial y} + \frac{\partial W}{\partial z} = 0. \tag{16.62}$$

Consequently, we can introduce a meridional-depth streamfunction

$$\Psi(y, z, t) = -\int_{-H_{\max}}^{z} V \,\mathrm{d}z' \tag{16.63}$$

to specify the flow. The lower limit of $z = -H_{\max}$ is a constant specified by the maximum depth in the domain, with zero transport for any regions below the fluid. To verify Ψ is a streamfunction, we compute

$$\frac{\partial \Psi}{\partial z} = -V \tag{16.64}$$

and

$$\frac{\partial \Psi}{\partial y} = -\frac{\partial}{\partial y}\left[\int_{-H_{\max}}^{z} V \,\mathrm{d}z'\right] = -\int_{-H_{\max}}^{z} \frac{\partial V}{\partial y} \,\mathrm{d}z' = \int_{-H_{\max}}^{z} \frac{\partial W}{\partial z} \,\mathrm{d}z' = W. \tag{16.65}$$

[3]The case for a steady compressible fluid follows analogously.

Figure 16.3: *An idealized rendition of the meridional-depth overturning circulation found in both the atmosphere and ocean. Shown here are streamlines 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 (16.62). 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.*

To reach this result, we made use of the non-divergent condition (16.62), and set

$$W(z = -H_{\max}) = 0. \tag{16.66}$$

Also, we are able to move the derivative across the lower limit of the integral since $H_{\max}$ is a spatial constant. An idealized version of the meridional-depth circulation is shown in Figure 16.3.

16.8 Exercises

EXERCISE 16.1: STREAMLINES FOR CELLULAR FLOW
Sketch the velocity field for this streamfunction

$$\psi(x, y) = A \sin(k\,x) \sin(l\,y), \tag{16.67}$$

where (k, l) are the zonal and meridional wavenumbers.

EXERCISE 16.2: TRANSPORT AND CIRCULATION
Consider a two dimensional velocity field, $\boldsymbol{u} = u\,\hat{\boldsymbol{x}} + v\,\hat{\boldsymbol{y}}$, that has both a zero divergence and a zero curl

$$\nabla \cdot \boldsymbol{u} = \nabla \wedge \boldsymbol{u} = 0. \tag{16.68}$$

(a) Show that the circulation of this velocity field through an arbitrary closed loop vanishes.

(b) Show that the transport of fluid crossing the same closed loop also vanishes.

EXERCISE 16.3: ZERO NET AREA TRANSPORT THROUGH STATIC CLOSED CURVE
For a two-dimensional non-divergent flow, show that there is zero net transport of area crossing an arbitrary static and simply connected closed curve.

EXERCISE 16.4: ZERO NET VOLUME TRANSPORT THROUGH STATIC CLOSED SURFACE
For a three-dimensional non-divergent flow, show that there is zero net transport of volume crossing an arbitrary static and simply connected closed surface within the fluid interior.

Figure 16.4: *Schematic for exercise 16.5, 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 fluid.*

EXERCISE 16.5: NET VOLUME TRANSPORT ACROSS AN ARBITRARY SURFACE

Consider a non-divergent 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 16.4. Integrate the volume transport over the surface, $\int_{\mathcal{S}} \boldsymbol{v} \cdot \hat{\boldsymbol{n}} \, \mathrm{d}\mathcal{S}$. Show that this transport vanishes. That is, the net transport across the surface is zero. Specialize this result to a horizontal surface so that we see there is zero area integrated vertical transport across the surface, $\int_{\mathcal{S}} w \, \mathrm{d}x \, \mathrm{d}y = 0$. Discuss these results. Note: see Section 17.3.8 for the more general case of a non-static surface.

EXERCISE 16.6: SOLID BODY ROTATION

Consider a velocity field corresponding to a time-independent solid-body rotation on a plane

$$\boldsymbol{u} = \Omega \, \hat{\boldsymbol{z}} \wedge \boldsymbol{x} = \Omega \, (-y \, \hat{\boldsymbol{x}} + x \, \hat{\boldsymbol{y}}), \tag{16.69}$$

where Ω is a constant rotation rate.

(a) Compute the relative vorticity $\nabla \wedge \boldsymbol{u}$.

(b) Compute the streamfunction $\boldsymbol{u} = \nabla \wedge (\hat{\boldsymbol{z}} \, \psi)$. Draw streamfunction contours (i.e., lines of constant streamfunction).

(c) Describe the geometry of material lines. Hint: since the velocity field is time-independent, material parcel trajectories are coincident with streamlines.

EXERCISE 16.7: ALTERNATIVE FORM OF MERIDIONAL-DEPTH STREAMFUNCTION

In equation (16.63), we introduced the meridional-depth overturning streamfunction

$$\Psi(y, z, t) = - \int_{-H_{\max}}^{z} V \, \mathrm{d}z'. \tag{16.70}$$

Show that an alternative streamfunction is given by

$$\Gamma(y, z, t) = \int_{y_s}^{y} W \, \mathrm{d}y', \tag{16.71}$$

where y_s is a constant latitude just south of the southern-most latitude in the domain.

Figure 16.5: *Schematic for exercise 16.9 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 16.8: VOLUME TRANSPORT THROUGH STREAMTUBE ENDS

Recall our discussion of streamtubes in Section 12.8.2 (see in particular Figure 12.5). For a steady non-divergent three-dimensional flow, show that the volume transport (volume per time) through the two streamtube ends balances

$$\int_{\mathcal{S}_1} \boldsymbol{v} \cdot \hat{\boldsymbol{n}}_1 \, \mathrm{d}\mathcal{S} + \int_{\mathcal{S}_2} \boldsymbol{v} \cdot \hat{\boldsymbol{n}}_2 \, \mathrm{d}\mathcal{S} = 0, \tag{16.72}$$

where $\hat{\boldsymbol{n}}_1$ and $\hat{\boldsymbol{n}}_2$ 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 (16.72) 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.

EXERCISE 16.9: AREA AVERAGE OF FREE SURFACE TIME TENDENCY

Consider an incompressible fluid bounded by a free upper surface and a solid bottom. Let $z = -H(x, y)$ 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 $h = H + \eta$ (see Figure 16.5). The horizontal extent of the layer is a function of time, and is defined by a vanishing thickness $h = H + \eta = 0$ (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.

17

General vertical coordinate 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 14.4 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 8. We make use of the dia-surface transport formulation to express the material time derivative operator using GVCs. This form for the material time operator allows us to decompose of the vertical velocity into motion relative to a moving GVC surface. In turn, we are afforded a means to reinterpret the velocity vector and particle trajectories. GVC kinematics also provides a means to express the subduction of fluid into the ocean interior beneath the mixed layer depth. 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 8, including the calculus using these non-orthogonal coordinates. It is essential to have a working knowledge of that material to understand material in the present chapter. Later in Chapter 38 we detail the dynamical equations using GVCs, with material in that chapter relying on the kinematics presented here. Following the treatment in Chapter 8, we here use the symbol σ to denote a generalized vertical coordinate.

17.1 Example generalized vertical coordinates **284**
17.1.1 Ocean free surface 284
17.1.2 Ocean bottom 284
17.1.3 Ocean mixed layer base 285
17.1.4 Interior generalized vertical coordinate surfaces 285
17.2 Specific thickness **285**
17.3 The dia-surface transport **286**
17.3.1 Flow normal to the GVC surface 286
17.3.2 Accounting for movement of the surface 287
17.3.3 We only care about divergent surface motion 288
17.3.4 Cross GVC transport in terms of GVC material evolution 288

17.3.5 Defining the dia-surface transport 288
17.3.6 Expressions for the dia-surface velocity component 289
17.3.7 An alternative definition of dia-surface velocity component 290
17.3.8 Area integrated dia-surface transport for incompressible fluids 290
17.4 Material time derivative **292**
17.5 Vertical velocity and dia-surface velocity **293**
17.5.1 Decomposing the vertical velocity 293
17.5.2 Another form of the vertical velocity decomposition 293
17.6 The velocity vector and fluid particle trajectories **295**
17.7 Subduction across the mixed layer base **297**
17.8 Mass continuity **297**
17.8.1 Cartesian coordinates 297
17.8.2 Generalized vertical coordinates 298
17.9 Layer integrated mass continuity **299**
17.9.1 Compressible fluids 299
17.9.2 Mass continuity using pressure coordinates 300
17.9.3 Incompressible fluids 302
17.9.4 Rescaled geopotential coordinates 302
17.10 Layer integrated tracer equation **303**

17.1 Example generalized vertical coordinates

We here consider some generalized vertical coordinates that will prove of use for our discussion in this chapter.

17.1.1 Ocean free surface

The first surface is the ocean free surface, whose kinematic boundary conditions were derived in Section 14.4.3. 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

$$\sigma(x, y, z, t) = z - \eta(x, y, t) = 0 \qquad \text{ocean free surface.} \tag{17.1}$$

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.

17.1.2 Ocean bottom

We may describe the solid Earth lower boundary mathematically by using the time independent expression

$$\sigma(x, y, z) = z + H(x, y) = 0 \qquad \text{ocean bottom.} \tag{17.2}$$

As detailed in Section 14.4.1, 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.

17.1.3 Ocean mixed layer base

Let

$$\sigma = z - \eta^{(\mathrm{mld})}(x, y, t) = 0 \tag{17.3}$$

represent the vertical position of the ocean mixed layer base. The corresponding normal vector is given by

$$\hat{\boldsymbol{n}}^{(\mathrm{mld})} = \frac{\nabla\,(z - \eta^{(\mathrm{mld})})}{|\nabla\,(z - \eta^{(\mathrm{mld})})|}. \tag{17.4}$$

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 17.7).

17.1.4 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.

17.2 Specific thickness

As mentioned in Section 8.8.1, 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

$$\frac{\partial z}{\partial \sigma} = z_\sigma, \tag{17.5}$$

which must then be single signed for useful 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 $z = \eta(x, y, t)$. 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-Helmholz billows and gravitational convection. Note that for both the ocean bottom and free surface

$$\frac{\partial z}{\partial \sigma} = 1 \qquad \text{ocean free surface and fluid/solid interface.} \tag{17.6}$$

Furthermore, this relation also holds, trivially, for geopotential coordinates in which $\sigma = z$.

We refer to the Jacobian z_σ as the *specific thickness* and sometimes find it useful to write it as

$$\mathbb{h} = z_\sigma = \frac{\partial z}{\partial \sigma}. \tag{17.7}$$

This name is motivated by noting that the vertical thickness of an infinitesimal layer of coordinate thickness $\delta\sigma$ is given by

$$\delta z = \frac{\partial z}{\partial \sigma}\,\delta\sigma = \mathbb{h}\,\delta\sigma. \tag{17.8}$$

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

$$\delta z = \frac{\partial z}{\partial \sigma}\,\delta b = N^{-2}\,\delta b, \tag{17.9}$$

with

$$N^2 = \frac{\partial b}{\partial z} \tag{17.10}$$

the squared buoyancy frequency (Section 21.4.4) in a Boussinesq fluid (Chapter 25). For a hydrostatic fluid using pressure as the vertical coordinate, the thickness of a pressure layer is

$$\delta z = \frac{\partial z}{\partial p}\,\delta p = -\frac{1}{\rho\, g}\,\delta p \tag{17.11}$$

where we used the hydrostatic relation (Section 24.3)

$$\frac{\partial p}{\partial z} = -\rho\, g \tag{17.12}$$

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.

17.3 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.

17.3.1 Flow normal to the GVC surface

At an arbitrary point on a surface of constant generalized vertical coordinate (see Figure 17.1), the rate at which fluid moves in the direction normal to the surface is given by

$$\text{RATE OF FLUID FLOW IN DIRECTION } \hat{\boldsymbol{n}} = \boldsymbol{v}\cdot\hat{\boldsymbol{n}}, \tag{17.13}$$

where

$$\hat{\boldsymbol{n}} = \frac{\nabla\sigma}{|\nabla\sigma|}, \tag{17.14}$$

is the surface unit normal. Two examples are useful to ground this expression in common experience. For the ocean free surface, $\sigma = z - \eta$, the unit normal takes the form

$$\hat{\boldsymbol{n}} = \frac{\nabla\,(z-\eta)}{|\nabla\,(z-\eta)|} = \frac{\hat{\boldsymbol{z}} - \nabla\eta}{\sqrt{1+|\nabla\eta|^2}} \qquad \text{ocean free surface,} \tag{17.15}$$

whereas at the solid Earth bottom, $\sigma = z + H$,

$$\hat{\boldsymbol{n}} = -\frac{\nabla\,(z+H)}{|\nabla\,(z+H)|} = -\frac{\hat{\boldsymbol{z}} + \nabla H}{\sqrt{1+|\nabla H|^2}} \qquad \text{ocean bottom.} \tag{17.16}$$

Introducing the material time derivative

$$\frac{\mathrm{D}\sigma}{\mathrm{D}t} = \frac{\partial\sigma}{\partial t} + \boldsymbol{v}\cdot\nabla\sigma \tag{17.17}$$

to equation (17.13) leads to the identity

$$\boldsymbol{v}\cdot\hat{\boldsymbol{n}} = \frac{1}{|\nabla\sigma|}\left[\frac{\mathrm{D}\sigma}{\mathrm{D}t} - \frac{\partial\sigma}{\partial t}\right]. \tag{17.18}$$

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 17.1: *A surface of constant generalized vertical coordinate, σ = constant, within a fluid. The normal direction, $\hat{\boldsymbol{n}} = \nabla\sigma/|\nabla\sigma|$, points in the direction of increasing σ. We show an example velocity vector for a fluid particle, $\boldsymbol{v}$, as well as the velocity, $\boldsymbol{v}^{(\sigma)}$, of a point living on the surface. Note that kinematics is only concerned with the normal component to the surface velocity, $\boldsymbol{v}^{(\sigma)} \cdot \hat{\boldsymbol{n}}$, as per equation (17.24). We require dynamical information to obtain information about the tangential component of $\boldsymbol{v}^{(\sigma)}$, but such information is not required for this chapter. Following equation (17.29), the horizontal projection of the surface area element is given by $\mathrm{d}A = |\cos\theta|\,\mathrm{d}\mathcal{S}$, where θ is the angle between the surface and the horizontal and $\mathrm{d}A = \mathrm{d}x\,\mathrm{d}y$.*

17.3.2 Accounting for movement of the surface

A GVC surface is generally moving. So to diagnose the net transport of fluid penetrating the surface requires us to subtract the velocity of the surface, $\boldsymbol{v}^{(\sigma)}$, from the velocity of a fluid particle. We are thus led to

$$\text{RATE THAT FLUID CROSSES A MOVING GVC SURFACE} = \hat{\boldsymbol{n}} \cdot (\boldsymbol{v} - \boldsymbol{v}^{(\sigma)}). \tag{17.19}$$

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

$$\delta\sigma = \delta\boldsymbol{x} \cdot \nabla\sigma + \delta t\, \partial_t \sigma. \tag{17.20}$$

Now restrict attention to a point fixed on a constant σ surface, in which

$$\delta\sigma = \delta\boldsymbol{x}^{(\sigma)} \cdot \nabla\sigma + \delta t\, \partial_t \sigma = 0, \tag{17.21}$$

where $\delta\boldsymbol{x}^{(\sigma)}$ is a differential increment following the moving surface. We define the velocity of that point as

$$\boldsymbol{v}^{(\sigma)} = \frac{\delta\boldsymbol{x}^{(\sigma)}}{\delta t}, \tag{17.22}$$

in which case equation (17.21) implies that at each point within the fluid,

$$\frac{\partial\sigma}{\partial t} + \boldsymbol{v}^{(\sigma)} \cdot \nabla\sigma = 0. \tag{17.23}$$

We can likewise write this equation as one for the normal component of the surface velocity

$$\boldsymbol{v}^{(\sigma)} \cdot \hat{\boldsymbol{n}} = -\frac{1}{|\nabla\sigma|}\frac{\partial\sigma}{\partial t}. \tag{17.24}$$

Hence, we reach the sensible result that the normal component to the velocity of the surface vanishes when the surface is static.

17.3.3 We only care about divergent surface motion

For the kinematics of fluid motion relative to arbitrary generalized vertical coordinates, we are only concerned with the normal component to the surface velocity, $\boldsymbol{v}^{(\sigma)} \cdot \hat{\boldsymbol{n}}$. That is, we are only concerned with divergent motion of the surface, defined as motion parallel to the surface normal direction, $\hat{\boldsymbol{n}}$. We have no concern for rotational or tangential motion, which is motion perpendicular to $\hat{\boldsymbol{n}}$. 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 the tangential motion of the surface.

17.3.4 Cross GVC transport in terms of GVC material evolution

Using expression (17.24) in equation (17.19) leads to the net flux of fluid crossing the GVC surface

$$\hat{\boldsymbol{n}} \cdot (\boldsymbol{v} - \boldsymbol{v}^{(\sigma)}) = \frac{1}{|\nabla\sigma|} \frac{\mathrm{D}\sigma}{\mathrm{D}t}. \tag{17.25}$$

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, $\boldsymbol{v}$, of Section 15.1.2. For multi-component fluids, $\dot{\sigma} = 0$ can still, in principle, be associed with trace matter exchange across the surface via diffusion so long as the net matter crossing the surface is zero. However, for the ocean, matter diffusion also occurs with heat diffusion, in which case $\dot{\sigma} = 0$ only occurs in the absence of both matter and heat diffusion.

17.3.5 Defining the dia-surface transport

The area normalizing the volume flux in equation (17.25) is the area $\mathrm{d}\mathcal{S}$ of an infinitesimal patch on the surface of constant generalized vertical coordinate with outward unit normal $\hat{\boldsymbol{n}}$. We now follow the trigonometry discussed in Section 8.16 to introduce the horizontal projection of this area, $\mathrm{d}A$, 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

$$\frac{\mathrm{d}\mathcal{S}}{|\nabla\sigma|} = \frac{\mathrm{d}\mathcal{S}}{\sqrt{(\partial\sigma/\partial x)^2 + (\partial\sigma/\partial y)^2 + (\partial\sigma/\partial z)^2}} \tag{17.26a}$$

$$= \frac{\mathrm{d}\mathcal{S}}{|\partial\sigma/\partial z| \sqrt{[(\partial\sigma/\partial x)/(\partial\sigma/\partial z)]^2 + [(\partial\sigma/\partial y)/(\partial\sigma/\partial z)]^2 + 1}} \tag{17.26b}$$

$$= \frac{\mathrm{d}\mathcal{S}}{|\partial\sigma/\partial z| \sqrt{1 + \tan^2\theta}} \tag{17.26c}$$

$$= \left|\frac{\partial z}{\partial \sigma}\right| |\cos\theta| \, \mathrm{d}\mathcal{S} \tag{17.26d}$$

$$= \left|\frac{\partial z}{\partial \sigma}\right| \mathrm{d}A. \tag{17.26e}$$

The equality (17.26c) introduced the angle, θ, between the boundary surface and the horizontal plane. The squared slope of this surface given by (see Section 8.11)

$$\tan^2\theta = \frac{\nabla_z\sigma \cdot \nabla_z\sigma}{(\partial\sigma/\partial z)^2} = \nabla_\sigma z \cdot \nabla_\sigma z. \tag{17.27}$$

The equality (17.26d) made use of a trigonometric identity so that

$$|\cos\theta|^{-1} = |z_\sigma \, \nabla\sigma|. \tag{17.28}$$

Furthermore, the equality (17.26e) introduced the horizontal projection of the area,

$$\mathrm{d}A = |\cos\theta| \, \mathrm{d}\mathcal{S}. \tag{17.29}$$

We now introduce the *dia-surface velocity component* for the GVC coordinate

$$w^{(\dot\sigma)} = \frac{\partial z}{\partial \sigma} \frac{\mathrm{D}\sigma}{\mathrm{D}t} = z_\sigma \, \dot\sigma, \tag{17.30}$$

which measures the volume of fluid passing through the surface, per unit horizontal area, per unit time

$$w^{(\dot\sigma)} \equiv \hat{\boldsymbol{n}} \cdot (\boldsymbol{v} - \boldsymbol{v}^{(\sigma)}) \, \frac{\mathrm{d}\mathcal{S}}{\mathrm{d}A} \tag{17.31}$$

$$= \frac{(\text{VOLUME/TIME}) \text{ FLUID THROUGH SURFACE}}{\text{HORIZONTAL AREA OF SURFACE}}, \tag{17.32}$$

so that

$$w^{(\dot\sigma)} \, \mathrm{d}A \equiv \hat{\boldsymbol{n}} \cdot (\boldsymbol{v} - \boldsymbol{v}^{(\sigma)}) \, \mathrm{d}\mathcal{S}. \tag{17.33}$$

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.

17.3.6 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 8.11, allows us to write the dia-surface velocity component in the following equivalent forms

$$w^{(\dot\sigma)} = \frac{\partial z}{\partial \sigma} \frac{\mathrm{D}\sigma}{\mathrm{D}t} \tag{17.34a}$$

$$= \frac{\partial z}{\partial \sigma} \, |\nabla\sigma| \, \hat{\boldsymbol{n}} \cdot (\boldsymbol{v} - \boldsymbol{v}^{(\sigma)}) \tag{17.34b}$$

$$= \frac{\partial z}{\partial \sigma} \, \nabla\sigma \cdot \boldsymbol{v} - \frac{\partial z}{\partial \sigma} \, |\nabla\sigma| \, \hat{\boldsymbol{n}} \cdot (\boldsymbol{v} - \boldsymbol{v}^{(\sigma)}) \tag{17.34c}$$

$$= (\hat{\boldsymbol{z}} - \nabla_\sigma z) \cdot \boldsymbol{v} + \frac{\partial z}{\partial \sigma} \frac{\partial \sigma}{\partial t} \tag{17.34d}$$

$$= (\hat{\boldsymbol{z}} - \nabla_\sigma z) \cdot \boldsymbol{v} - \frac{\partial z}{\partial t} \tag{17.34e}$$

$$= w - (\partial_t + \boldsymbol{u} \cdot \nabla_\sigma) z, \tag{17.34f}$$

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 (8.22b) and (8.22c))

$$\nabla_\sigma z = -z_\sigma \, \nabla_z \sigma \tag{17.35}$$

to express the slope of the σ surface as projected onto the horizontal direction plane, as well as the corresponding identity (8.22a) for the time derivative

$$\left[\frac{\partial z}{\partial t}\right]_\sigma = -\frac{[\partial\sigma/\partial t]_z}{[\partial\sigma/\partial z]} \tag{17.36}$$

The form given by equation (17.34f) directly relates the vertical component to the fluid particle velocity to the dia-surface velocity component

$$w = \frac{\mathrm{D}z}{\mathrm{D}t} \longleftrightarrow w^{(\dot\sigma)} = \frac{\partial z}{\partial\sigma}\frac{\mathrm{D}\sigma}{\mathrm{D}t} = w - (\partial_t + \boldsymbol{u}\cdot\nabla_\sigma)z. \tag{17.37}$$

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

$$w^{(\dot\sigma)} = w - \boldsymbol{u}\cdot\nabla_\sigma z \qquad \text{static surface,} \tag{17.38}$$

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

$$w^{(\dot\sigma)} = w = \frac{\mathrm{D}z}{\mathrm{D}t} \qquad \text{GVC surface static and flat,} \tag{17.39}$$

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 (17.34f) thus offers a useful means to distinguish w from $w^{(\dot\sigma)}$.

17.3.7 An alternative definition of dia-surface velocity component

In some literature presentations, the dia-surface velocity component is taken to be

$$w^{\mathrm{dia}} = \hat{\boldsymbol{n}}\cdot(\boldsymbol{v} - \boldsymbol{v}^{(\sigma)}) = \frac{1}{|\nabla\sigma|}\frac{\mathrm{D}\sigma}{\mathrm{D}t}. \tag{17.40}$$

For example, *Groeskamp et al.* (2019) prefer this definition for watermass analysis. As seen in Chapter 51, the reason to prefer expression (17.40) for watermass transformation 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.

17.3.8 Area integrated dia-surface transport for incompressible fluids

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 fluid whereby $\nabla\cdot\boldsymbol{v} = 0$. Incompressibility means that for any closed surface within the fluid interior, the following identity holds via the divergence theorem

$$0 = \int_{\mathcal{R}} \nabla\cdot\boldsymbol{v}\,\mathrm{d}V = \oint_{\partial\mathcal{R}} \hat{\boldsymbol{n}}\cdot\boldsymbol{v}\,\mathrm{d}\mathcal{S}. \tag{17.41}$$

Figure 17.2: *A constant GVC surface, σ = constant, within an ocean basin that intersects the bottom. The region $\mathscr{R}$ is bounded above by the σ surface and below by the solid-earth. Along the constant σ surface an incompressible fluid satisfies $\int_{\sigma=const} \hat{\boldsymbol{n}} \cdot \boldsymbol{v} \, \mathrm{d}\mathcal{S} = 0$.*

Notably, only in the case of a static surface do we conclude there is no net flow across the surface (see Exercise 16.5). 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 17.2, 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 (17.41) means that the area integrated flow normal to the GVC vanishes

$$\int_{\sigma=\text{const}} \hat{\boldsymbol{n}} \cdot \boldsymbol{v} \, \mathrm{d}\mathcal{S} = 0. \tag{17.42}$$

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, $\int_{z=\text{const}} w \, \mathrm{d}A = 0$ (see Exercise 16.5). What about other GVCs?

To address this question consider the general result

$$\int_{\sigma=\text{const}} \hat{\boldsymbol{n}} \cdot (\boldsymbol{v} - \boldsymbol{v}^{(\sigma)}) \, \mathrm{d}\mathcal{S} = \int_{\sigma=\text{const}} w^{\text{dia}} \, \mathrm{d}\mathcal{S} = \int_{\sigma=\text{const}} w^{(\dot{\sigma})} \, \mathrm{d}A, \tag{17.43}$$

where again $\mathrm{d}A = \mathrm{d}x \, \mathrm{d}y$. Now make use of the property (17.42) for incompressible flows as well as the identity (17.24) to render

$$\int_{\sigma=\text{const}} w^{(\dot{\sigma})} \, \mathrm{d}A = 0 - \int_{\sigma=\text{const}} \hat{\boldsymbol{n}} \cdot \boldsymbol{v}^{(\sigma)} \, \mathrm{d}\mathcal{S} \tag{17.44a}$$

$$= \int_{\sigma=\text{const}} \frac{\partial \sigma / \partial t}{|\nabla \sigma|} \, \mathrm{d}\mathcal{S} \tag{17.44b}$$

$$= \int_{\sigma=\text{const}} \frac{\partial \sigma}{\partial t} \left| \frac{\partial z}{\partial \sigma} \right| \mathrm{d}A \tag{17.44c}$$

$$= - \int_{\sigma=\text{const}} \left[\frac{\partial z}{\partial t} \right]_{\sigma} \mathrm{d}A. \tag{17.44d}$$

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

$$\int_{\sigma=\text{const}} w^{(\dot{\sigma})} \, \mathrm{d}A = - \left[\frac{\partial}{\partial t} \right]_{\sigma} \int_{\sigma=\text{const}} z \, \mathrm{d}A. \tag{17.45}$$

This identity means that for an incompressible fluid, 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 $\int_{\sigma=\text{const}} w^{(\dot{\sigma})}\,\mathrm{d}A = 0$. However, this result does not hold for other coordinates, such as the rescaled vertical coordinate, $\sigma = z^*$ defined by equation (17.112). In this case

$$z^* = H\,\frac{z-\eta}{H+\eta} \tag{17.46a}$$

$$\frac{\partial z}{\partial z^*} = 1 + H/\eta > 0 \tag{17.46b}$$

$$\left[\frac{\partial z}{\partial t}\right]_{z^*} = \frac{\partial \eta}{\partial t}\,(1 + z^*/H), \tag{17.46c}$$

so that

$$\int_{z^*=\text{const}} w^{(\dot{z^*})}\,\mathrm{d}A = \int_{z^*=\text{const}} (\partial\eta/\partial t)\,(1+z^*/H)\,\mathrm{d}A, \tag{17.47}$$

which is generally nonzero. For example, consider a flat bottom so that

$$\int_{z^*=\text{const}} w^{(\dot{z^*})}\,\mathrm{d}A = (1+z^*/H)\int_{z^*=\text{const}} (\partial\eta/\partial t)\,\mathrm{d}A = (1+z^*/H)\int_{z^*=\text{const}} (Q_{\mathrm{m}}/\rho_0)\,\mathrm{d}A, \tag{17.48}$$

where Q_{m} is the surface mass flux and we made use of the incompressible free surface equation (16.8). In this case the area integrated dia-surface transport across a z^* surface is proportional to the area integrated surface mass flux.

17.4 Material time derivative

The expression (17.30) for $w^{(\dot{\sigma})}$ brings the material time derivative operator into the following equivalent forms

$$\frac{\mathrm{D}}{\mathrm{D}t} = \left[\frac{\partial}{\partial t}\right]_z + \boldsymbol{u}\cdot\nabla_z + w\,\frac{\partial}{\partial z} \tag{17.49a}$$

$$= \left[\frac{\partial}{\partial t}\right]_\sigma + \boldsymbol{u}\cdot\nabla_\sigma + \frac{\mathrm{D}\sigma}{\mathrm{D}t}\,\frac{\partial}{\partial \sigma} \tag{17.49b}$$

$$= \left[\frac{\partial}{\partial t}\right]_\sigma + \boldsymbol{u}\cdot\nabla_\sigma + w^{(\dot{\sigma})}\,\frac{\partial}{\partial z}. \tag{17.49c}$$

Note that the chain-rule means that

$$\frac{\partial}{\partial \sigma} = \frac{\partial z}{\partial \sigma}\,\frac{\partial}{\partial z}, \tag{17.50}$$

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 surface. This sitution 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 (17.49c) only has a horizontal two-dimensional

advective component $\boldsymbol{u} \cdot \nabla_b$. 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 (17.49a) 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.

17.5 Vertical velocity and dia-surface velocity

Making use of the material time derivative operator (17.49c) 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

$$w = \frac{\mathrm{D}z}{\mathrm{D}t} = \left[\frac{\partial z}{\partial t}\right]_\sigma + \boldsymbol{u} \cdot \nabla_\sigma z + w^{(\dot\sigma)} = \frac{\partial z}{\partial \sigma}\left[-\frac{\partial \sigma}{\partial t} - \boldsymbol{u} \cdot \nabla_z \sigma + \frac{\mathrm{D}\sigma}{\mathrm{D}t}\right], \tag{17.51}$$

and the corresponding expressions for the dia-surface velocity component are given by

$$w^{(\dot\sigma)} = \frac{\partial z}{\partial \sigma}\frac{\mathrm{D}\sigma}{\mathrm{D}t} = \frac{\partial z}{\partial \sigma}\left[\frac{\partial \sigma}{\partial t} + \boldsymbol{u} \cdot \nabla_z \sigma + w\,\frac{\partial \sigma}{\partial t}\right] = -\left[\frac{\partial z}{\partial t}\right]_\sigma - \boldsymbol{u} \cdot \nabla_\sigma z + w. \tag{17.52}$$

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 (17.32) 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, its appearance in the w equation does not necessarily mean that it 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.

17.5.1 Decomposing the vertical velocity

The expression

$$w = \left[\frac{\partial z}{\partial t}\right]_\sigma + \boldsymbol{u} \cdot \nabla_\sigma z + w^{(\dot\sigma)} \tag{17.53}$$

decomposes the vertical velocity of a fluid particle into (A) changes to the vertical position of the σ-surface at a particular horizontal point, (B) lateral particle motion projected onto a sloped σ-surface, (C) 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 $\nabla_\sigma z$.

17.5.2 Another form of the vertical velocity decomposition

Consider the velocity for the surface itself, $\boldsymbol{v}^{(\sigma)}$, which satisfies (Section 17.3.2)

$$\frac{\partial \sigma}{\partial t} + \boldsymbol{v}^{(\sigma)} \cdot \nabla \sigma = 0. \tag{17.54}$$

Making use of the triple product identities from Section 8.4

$$\frac{\partial z}{\partial \sigma}\nabla \sigma = -\nabla_\sigma z + \hat{\boldsymbol{z}} \qquad \text{and} \qquad \frac{\partial z}{\partial \sigma}\left[\frac{\partial \sigma}{\partial t}\right]_z = -\left[\frac{\partial z}{\partial t}\right]_\sigma \tag{17.55}$$

brings equation (17.54) into the form

$$\left[\frac{\partial z}{\partial t}\right]_\sigma = (\hat{\boldsymbol{z}} - \nabla_\sigma z)\cdot \boldsymbol{v}^{(\sigma)} \Longrightarrow \hat{\boldsymbol{z}}\cdot \boldsymbol{v}^{(\sigma)} = \left[\frac{\partial z}{\partial t}\right]_\sigma + \boldsymbol{u}^{(\sigma)}\cdot \nabla_\sigma z, \tag{17.56}$$

where $\boldsymbol{u}^{(\sigma)}$ is the horizontal component to the surface velocity $\boldsymbol{v}^{(\sigma)}$. 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 pass that point

$$\left[\frac{\partial z}{\partial t}\right]_\sigma = -\boldsymbol{u}^{(\sigma)}\cdot \nabla_\sigma z \qquad \text{if } \hat{\boldsymbol{z}}\cdot \boldsymbol{v}^{(\sigma)} = 0. \tag{17.57}$$

Returning to the general result (17.56) allows us to write

$$\left[\frac{\partial z}{\partial t}\right]_\sigma + \boldsymbol{u}\cdot \nabla_\sigma z = \hat{\boldsymbol{z}}\cdot \boldsymbol{v}^{(\sigma)} + (\boldsymbol{u} - \boldsymbol{u}^{(\sigma)})\cdot \nabla_\sigma z. \tag{17.58}$$

Furthermore, return to the fundamental definition of the dia-surface velocity component detailed in Section 17.3, in which we showed that

$$w^{(\dot\sigma)} = \frac{\partial z}{\partial \sigma}\frac{\mathrm{D}\sigma}{\mathrm{D}t} = \frac{\partial z}{\partial \sigma}\nabla\sigma\cdot(\boldsymbol{v} - \boldsymbol{v}^{(\sigma)}) = (-\nabla_\sigma z + \hat{\boldsymbol{z}})\cdot(\boldsymbol{v} - \boldsymbol{v}^{(\sigma)}). \tag{17.59}$$

This expression, along with equation (17.58), leads to the rather elaborate decomposition of the vertical velocity component according to motion of a generalized vertical coordinate surface

$$w = \underbrace{\left[\hat{\boldsymbol{z}}\cdot \boldsymbol{v}^{(\sigma)} + (\boldsymbol{u} - \boldsymbol{u}^{(\sigma)})\cdot \nabla_\sigma z\right]}_{(\partial_t + \boldsymbol{u}\cdot\nabla_\sigma)z} + \underbrace{\left[\hat{\boldsymbol{z}}\cdot \boldsymbol{v} - \hat{\boldsymbol{z}}\cdot \boldsymbol{v}^{(\sigma)} - (\boldsymbol{u} - \boldsymbol{u}^{(\sigma)})\cdot \nabla_\sigma z\right]}_{w^{(\dot\sigma)}}. \tag{17.60}$$

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 $\hat{\boldsymbol{z}}\cdot\boldsymbol{v} = w$, thus trivially revealing $w = w$. The decomposition of w is rather pedantic when viewed in the unwrapped form of equation (17.60). 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: $\boldsymbol{u} = \boldsymbol{u}^{(\sigma)}$. Alternatively, consider the case with flat σ-surfaces so that $\nabla_\sigma z = 0$. In either case the vertical velocity is given by

$$w = \underbrace{\left[\hat{\boldsymbol{z}}\cdot \boldsymbol{v}^{(\sigma)}\right]}_{(\partial_t + \boldsymbol{u}\cdot\nabla_\sigma)z} + \underbrace{\left[\hat{\boldsymbol{z}}\cdot(\boldsymbol{v} - \boldsymbol{v}^{(\sigma)})\right]}_{w^{(\dot\sigma)}}. \tag{17.61}$$

  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

$$w = 0 \tag{17.62a}$$

$$= \left[\frac{\partial z}{\partial t}\right]_\sigma + \boldsymbol{u} \cdot \nabla_\sigma z + w^{(\dot\sigma)} \tag{17.62b}$$

$$= \underbrace{\left[\hat{\boldsymbol{z}} \cdot \boldsymbol{v}^{(\sigma)} + (\boldsymbol{u} - \boldsymbol{u}^{(\sigma)}) \cdot \nabla_\sigma z\right]}_{(\partial_t + \boldsymbol{u}\cdot\nabla_\sigma) z} + \underbrace{\left[-\hat{\boldsymbol{z}} \cdot \boldsymbol{v}^{(\sigma)} - (\boldsymbol{u} - \boldsymbol{u}^{(\sigma)}) \cdot \nabla_\sigma z\right]}_{w^{(\dot\sigma)}}. \tag{17.62c}$$

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

$$-w^{(\dot\sigma)} = \left[\frac{\partial z}{\partial t}\right]_\sigma + \boldsymbol{u} \cdot \nabla_\sigma z \qquad \text{if } w = 0. \tag{17.63}$$

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 merely since $w^{(\dot\sigma)} \neq 0$ can arise from motion of a σ-surface alone.

17.6 The velocity vector and fluid particle trajectories

Recall from Section 17.5 the alternative forms for the vertical velocity component given by equation (17.51). We focus on the form

$$w = \left[\frac{\partial z}{\partial t}\right]_\sigma + \boldsymbol{u} \cdot \nabla_\sigma z + w^{(\dot\sigma)} \tag{17.64}$$

so that the velocity vector is written[1]

$$\boldsymbol{v} = u\,\hat{\boldsymbol{x}} + v\,\hat{\boldsymbol{y}} + w\,\hat{\boldsymbol{z}} \tag{17.65a}$$

$$= u\,\hat{\boldsymbol{x}} + v\,\hat{\boldsymbol{y}} + \left[(\partial z/\partial t)_\sigma + \boldsymbol{u} \cdot \nabla_\sigma z + w^{(\dot\sigma)}\right] \hat{\boldsymbol{z}} \tag{17.65b}$$

$$= u\left[\hat{\boldsymbol{x}} + \hat{\boldsymbol{z}}\,(\partial z/\partial x)_\sigma\right] + v\left[\hat{\boldsymbol{y}} + \hat{\boldsymbol{z}}\,(\partial z/\partial y)_\sigma\right] + \left[(\partial z/\partial t)_\sigma + w^{(\dot\sigma)}\right] \hat{\boldsymbol{z}}. \tag{17.65c}$$

To help further understand these velocity expressions we consider the following three cases, each of which are illustrated in Figure 17.3.

- STEADY AND MATERIAL σ-SURFACE: The velocity vector is aligned with the instantaneous σ-surface ($\boldsymbol{v} \cdot \nabla\sigma = 0$) when the σ-surface is steady ($\partial\sigma/\partial t = 0$) and material ($\mathrm{D}\sigma/\mathrm{D}t = 0$). Hence, we can diagnose the vertical velocity component in terms of the horizontal via

$$w\,\partial\sigma/\partial z = -\boldsymbol{u} \cdot \nabla_z \sigma \Longrightarrow w = \boldsymbol{u} \cdot \nabla_\sigma z, \tag{17.66}$$

[1] As discussed in Section 8.6, we can connect these expressions to the contravariant representation of the velocity vector using GVCs.

Figure 17.3: *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 (17.67). The slope of the σ-surface in the $\hat{\boldsymbol{x}}$-direction is given by $\tan\theta = (\partial z/\partial x)_\sigma$. The middle panel is for a non-steady material σ-surface whereby the velocity of a particle takes on the form (17.68), 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 (17.69). 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).*

where we used the triple product identities (8.22b) and (8.22c) for the final equality. The velocity vector thus takes on the form

$$\boldsymbol{v} = u\,[\hat{\boldsymbol{x}} + \hat{\boldsymbol{z}}\,(\partial z/\partial x)_\sigma] + v\,[\hat{\boldsymbol{y}} + \hat{\boldsymbol{z}}\,(\partial z/\partial y)_\sigma] \qquad \partial\sigma/\partial t = 0 \text{ and } \mathrm{D}\sigma/\mathrm{D}t = 0. \tag{17.67}$$

The velocity vector is determined only by the horizontal velocity plus the slope of the σ surface.

- NON-STEADY AND MATERIAL σ-SURFACE: Next consider material σ surfaces ($\mathrm{D}\sigma/\mathrm{D}t = 0$) that move ($\partial_t\sigma \neq 0$), in which case the velocity vector is

$$\boldsymbol{v} = u\,[\hat{\boldsymbol{x}} + \hat{\boldsymbol{z}}\,(\partial z/\partial x)_\sigma] + v\,[\hat{\boldsymbol{y}} + \hat{\boldsymbol{z}}\,(\partial z/\partial y)_\sigma] + (\partial z/\partial t)_\sigma\,\hat{\boldsymbol{z}} \qquad \mathrm{D}\sigma/\mathrm{D}t = 0. \tag{17.68}$$

To remain on the moving surface, the fluid particle must move vertically by the extra amount $(\partial z/\partial t)_\sigma\,\delta t\,\hat{\boldsymbol{z}}$ 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, $w^{(\dot\sigma)} = (\partial z/\partial\sigma)\,\mathrm{D}\sigma/\mathrm{D}t$:

$$\boldsymbol{v} = u\,[\hat{\boldsymbol{x}} + \hat{\boldsymbol{z}}\,(\partial z/\partial x)_\sigma] + v\,[\hat{\boldsymbol{y}} + \hat{\boldsymbol{z}}\,(\partial z/\partial y)_\sigma] + \left[(\partial z/\partial t)_\sigma + w^{(\dot\sigma)}\right]\hat{\boldsymbol{z}}. \tag{17.69}$$

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 17.5, 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 17.5.2 with a static particle and moving σ-surface.

17.7 Subduction across the mixed layer base

Consider the GVC (17.3) defined according to the mixed layer base. The dia-surface mass transport across this surface leads us to define the subduction

$$-\mathcal{S}^{(\text{subduction})} \equiv \rho \, \mathrm{d}A \left[\frac{\mathrm{D}(z - \eta^{(\text{mld})})}{\mathrm{D}t} \right] \qquad \text{at } z = \eta^{(\text{mld})}(x, y, t), \tag{17.70}$$

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 $\mathrm{d}A$ is the horizontal projection of the area on the mixed layer base. Expanding the material time derivative leads to

$$-\left[\frac{\mathcal{S}^{(\text{subduction})}}{\rho \, \mathrm{d}A} \right] = w - [\partial_t + \boldsymbol{u} \cdot \nabla] \, \eta^{(\text{mld})} \qquad \text{at } z = \eta^{(\text{mld})}(x, y, t), \tag{17.71}$$

where again we define

$$\mathcal{S}^{(\text{subduction})} > 0 \qquad \text{subduction} \tag{17.72}$$

$$\mathcal{S}^{(\text{subduction})} < 0 \qquad \text{obduction.} \tag{17.73}$$

This kinematic definition of subduction corresponds to that given by *Cushmin-Roisin* (1987).

17.8 Mass continuity

We here derive the Eulerian equation for mass continuity (14.9) using generalized vertical coordinates. We then specialize to incompressible fluids, in which mass conservation is converted to volume conservation. To start, recall that mass conservation for a fluid parcel states that

$$\rho \, \delta V = \rho \, \delta x \, \delta y \, \delta z = \rho \, \delta x \, \delta y \, z_\sigma \, \delta \sigma \tag{17.74}$$

is materially constant. To develop the Eulerian expressions we first consider the case of Cartesian coordinates.

17.8.1 Cartesian coordinates

Consider the expression

$$\frac{1}{\rho \, \delta V} \frac{\mathrm{D}\,(\rho \, \delta V)}{\mathrm{D}t} = \frac{1}{\rho} \frac{\mathrm{D}\rho}{\mathrm{D}t} + \frac{1}{\delta V} \frac{\mathrm{D}(\delta V)}{\mathrm{D}t}. \tag{17.75}$$

Now make use of Cartesian coordinates to write for the volume

$$\frac{1}{\delta V} \frac{\mathrm{D}\,(\delta V)}{\mathrm{D}t} = \frac{1}{\delta x \, \delta y \, \delta z} \frac{\mathrm{D}(\delta x \, \delta y \, \delta z)}{\mathrm{D}t} \tag{17.76a}$$

$$= \frac{1}{\delta x} \frac{\mathrm{D}(\delta x)}{\mathrm{D}t} + \frac{1}{\delta y} \frac{\mathrm{D}(\delta y)}{\mathrm{D}t} + \frac{1}{\delta z} \frac{\mathrm{D}(\delta z)}{\mathrm{D}t} \tag{17.76b}$$

$$= \frac{\delta u}{\delta x} + \frac{\delta v}{\delta y} + \frac{\delta w}{\delta z} \tag{17.76c}$$

$$= \nabla \cdot \boldsymbol{v}. \tag{17.76d}$$

Setting $\mathrm{D}(\rho \, \delta V)/\mathrm{D}t = 0$ leads to the familiar expression for the continuity equation

$$\frac{\mathrm{D}\rho}{\mathrm{D}t} = -\rho \, \nabla \cdot \boldsymbol{v}. \tag{17.77}$$

17.8.2 Generalized vertical coordinates

We follow the above procedure but now with generalized vertical coordinates so that

$$\frac{1}{\delta V}\frac{\mathrm{D}\,(\delta V)}{\mathrm{D}t} = \frac{1}{\delta x\,\delta y\,z_\sigma\,\delta\sigma}\frac{\mathrm{D}(\delta x\,\delta y\,z_\sigma\,\delta\sigma)}{\mathrm{D}t} \tag{17.78a}$$

$$= \frac{1}{\delta x}\frac{\mathrm{D}(\delta x)}{\mathrm{D}t} + \frac{1}{\delta y}\frac{\mathrm{D}(\delta y)}{\mathrm{D}t} + \frac{1}{z_\sigma}\frac{\mathrm{D}(z_\sigma)}{\mathrm{D}t} + \frac{1}{\delta\sigma}\frac{\mathrm{D}(\delta\sigma)}{\mathrm{D}t} \tag{17.78b}$$

$$= \frac{\delta u}{\delta x} + \frac{\delta v}{\delta y} + \frac{1}{z_\sigma}\frac{\mathrm{D}(z_\sigma)}{\mathrm{D}t} + \frac{\delta(\dot\sigma)}{\delta\sigma} \tag{17.78c}$$

$$= \nabla_\sigma\cdot\boldsymbol{u} + \frac{1}{z_\sigma}\frac{\mathrm{D}(z_\sigma)}{\mathrm{D}t} + \frac{\partial\dot\sigma}{\partial\sigma} \tag{17.78d}$$

where we introduced the shorthand

$$\dot\sigma = \frac{\mathrm{D}\sigma}{\mathrm{D}t}. \tag{17.79}$$

Note that we set

$$\frac{\delta u}{\delta x} + \frac{\delta v}{\delta y} = \nabla_\sigma\cdot\boldsymbol{u} \tag{17.80}$$

since we are working with generalized vertical coordinates so that we consider infinitesimal displacements occurring on constant σ surfaces. We are thus led to

$$\frac{1}{\rho\,\delta V}\frac{\mathrm{D}(\rho\,\delta V)}{\mathrm{D}t} = \nabla_\sigma\cdot\boldsymbol{u} + \frac{1}{z_\sigma}\frac{\mathrm{D}z_\sigma}{\mathrm{D}t} + \frac{\partial\dot\sigma}{\partial\sigma} + \frac{1}{\rho}\frac{\mathrm{D}\rho}{\mathrm{D}t} = 0. \tag{17.81}$$

Now use the material time derivative in the form (17.49b) to derive the Eulerian expression of mass conservation

$$\frac{\partial(\rho\,z_\sigma)}{\partial t} + \nabla_\sigma\cdot(\rho\,z_\sigma\boldsymbol{u}) + \frac{\partial(\rho\,z_\sigma\,\dot\sigma)}{\partial\sigma} = 0, \tag{17.82}$$

where the time derivative is computed holding σ fixed. We can furthermore introduce the dia-surface velocity component

$$w^{(\dot\sigma)} = z_\sigma\,\dot\sigma \tag{17.83}$$

so that mass continuity takes the form

$$\frac{\partial(\rho\,z_\sigma)}{\partial t} + \nabla_\sigma\cdot(\rho\,z_\sigma\boldsymbol{u}) + \frac{\partial(\rho\,w^{(\dot\sigma)})}{\partial\sigma} = 0. \tag{17.84}$$

Alternatively, we can reintroduce the material time derivative operator to write the mass continuity equation (17.82) in the form

$$\frac{1}{\rho\,z_\sigma}\frac{\mathrm{D}(\rho\,z_\sigma)}{\mathrm{D}t} = -(\nabla_\sigma\cdot\boldsymbol{u} + \partial\dot\sigma/\partial\sigma), \tag{17.85}$$

where we used equation (17.49b) to write

$$\frac{\mathrm{D}}{\mathrm{D}t} = \left[\frac{\partial}{\partial t}\right]_\sigma + \boldsymbol{u}\cdot\nabla_\sigma + \dot\sigma\,\frac{\partial}{\partial\sigma}. \tag{17.86}$$

17.9 Layer integrated mass continuity

The formulation thus far has been continuous, with the only assumption made that the specific thickness, $\hbar = \partial z/\partial\sigma$, is single signed. We here consider a discrete increment in the generalized vertical coordinate,

$$\sigma - \delta\sigma/2 \leq \sigma' \leq \sigma + \delta\sigma/2, \tag{17.87}$$

and formulate the mass budget over this layer whose thickness is given by

$$h \equiv \int_{z(\sigma-\delta\sigma/2)}^{z(\sigma+\delta\sigma/2)} \mathrm{d}z = \int_{\sigma-\delta\sigma/2}^{\sigma+\delta\sigma/2} \frac{\partial z}{\partial\sigma}\,\mathrm{d}\sigma, \tag{17.88}$$

and whose mass per horizontal area is

$$\delta m = \int_{z(\sigma-\delta\sigma/2)}^{z(\sigma+\delta\sigma/2)} \rho\,\mathrm{d}z = \int_{\sigma-\delta\sigma/2}^{\sigma+\delta\sigma/2} \rho\, z_\sigma\,\mathrm{d}\sigma = \overline{\rho}\, h, \tag{17.89}$$

where $\overline{\rho}$ is the layer averaged density. Note that for Boussinesq fluids the mass per area equals to the layer thickness times the reference density

$$\delta m = \rho_0\, h \qquad \text{Boussinesq.} \tag{17.90}$$

As defined by equation (17.88), the thickness of a layer is relatively large in regions where $\partial\sigma/\partial z$ is small; i.e., in weakly stratified regions. 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, so that $\partial p/\partial z = -\rho\, g$ (Section 17.9.2).

The formulation in this section, and its companion for tracers in Section 17.10, 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.

17.9.1 Compressible fluids

Performing a layer integral of the specific thickness equation (17.84) renders

$$\int_{\sigma-\delta\sigma/2}^{\sigma+\delta\sigma/2} \left[\frac{\partial(\rho\, z_\sigma)}{\partial t} + \nabla_\sigma \cdot (\rho\, z_\sigma \boldsymbol{u}) + \frac{\partial(\rho\, w^{(\dot\sigma)})}{\partial\sigma} \right] \mathrm{d}\sigma = 0. \tag{17.91}$$

The dia-surface term integrates to a finite difference across the layer

$$\int_{\sigma-\delta\sigma/2}^{\sigma+\delta\sigma/2} \left[\frac{\partial(\rho\, z_\sigma)}{\partial t} + \nabla_\sigma \cdot (\rho\, z_\sigma \boldsymbol{u}) \right] = -\Delta_\sigma(\rho\, w^{(\dot\sigma)}), \tag{17.92}$$

where we introduced the finite difference operator for properties defined at the layer interface

$$\Delta_\sigma(\Phi) = \Phi(\sigma + \delta\sigma/2) - \Phi(\sigma - \delta\sigma/2). \tag{17.93}$$

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

$$\left[\frac{\partial}{\partial t}\right]_\sigma \int_{\sigma-\delta\sigma/2}^{\sigma+\delta\sigma/2} \rho\, z_\sigma \,\mathrm{d}\sigma + \nabla_\sigma \cdot \int_{\sigma-\delta\sigma/2}^{\sigma+\delta\sigma/2} \rho\, \boldsymbol{u}\, z_\sigma \,\mathrm{d}\sigma = -\Delta_\sigma(\rho\, w^{(\dot\sigma)}). \tag{17.94}$$

The first term involves the layer averaged density times the layer thickness as per equation (17.89). The second term involves the layer averaged density-weighted velocity, which is the layer averaged horizontal mass flux

$$\int_{\sigma-\delta\sigma/2}^{\sigma+\delta\sigma/2} \rho\, \boldsymbol{u}\, z_\sigma \,\mathrm{d}\sigma = h\, \overline{\rho}\, \overline{\boldsymbol{u}}. \tag{17.95}$$

We are thus led to the layer integrated continuity equation

$$\left[\frac{\partial(h\,\overline{\rho})}{\partial t}\right]_\sigma + \nabla_\sigma \cdot (h\, \overline{\rho}\, \overline{\boldsymbol{u}}) + \Delta_\sigma(\rho\, w^{(\dot\sigma)}) = 0. \tag{17.96}$$

When evolving the fields in a discrete numerical model, we have information only about layer averged fields. So how do we estimate the depth average of the horizontal advective flux, $\overline{\rho\, \boldsymbol{u}}$, appearing in equation (17.96)? One method interprets all fields as their layer averaged values so that $\overline{\rho\, \boldsymbol{u}} = \overline{\rho}\, \overline{\boldsymbol{u}}$, thus considering uncomputed sub-layer correlations $\overline{\rho'\, \boldsymbol{u}'}$ 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

$$\delta m = \int_{\sigma-\delta\sigma/2}^{\sigma+\delta\sigma/2} \rho\, z_\sigma \,\mathrm{d}\sigma = \overline{\rho}\, h = -g^{-1}\, \delta p. \tag{17.97}$$

Correspondingly, the layer integrated horizontal mass flux equals to the mass increment times the pressure-layer averaged velocity

$$\int_{\sigma-\delta\sigma/2}^{\sigma+\delta\sigma/2} \rho\, \boldsymbol{u}\, z_\sigma \,\mathrm{d}\sigma = -g^{-1} \int_{p-\delta p/2}^{p+\delta p/2} \boldsymbol{u}\, \mathrm{d}p = -g^{-1}\, \overline{\boldsymbol{u}}\, \delta p = h\, \overline{\rho}\, \overline{\boldsymbol{u}}. \tag{17.98}$$

With either of the above two methods, we are led to the same layer integrated continuity equation, which we write in the generic form

$$\left[\frac{\partial(h\,\rho)}{\partial t}\right]_\sigma + \nabla_\sigma \cdot (h\, \rho\, \boldsymbol{u}) + \Delta_\sigma\, (\rho\, w^{(\dot\sigma)}) = 0. \tag{17.99}$$

We illustrate contributions to this layer mass budget in Figure 17.4.

17.9.2 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 (17.88)) takes on the following form

$$h = \int_{p-\delta p/2}^{p+\delta p/2} \frac{\partial z}{\partial p}\, \mathrm{d}p = -\int_{p-\delta p/2}^{p+\delta p/2} \frac{\mathrm{d}p}{\rho\, g}, \tag{17.100}$$

Figure 17.4: *Illustrating the terms contributing to changes in layer mass according to the layer integrated continuity equation (17.99). 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 $w^{(\dot\sigma)}$.*

so that its mass per unit area is

$$\int_{p-\delta p/2}^{p+\delta p/2} \rho \frac{\partial z}{\partial p}\, \mathrm{d}p = -\delta p/g. \tag{17.101}$$

The mass continuity equation (17.99) thus becomes

$$\frac{\partial(\delta p)}{\partial t} + \nabla_p \cdot (\boldsymbol{u}\, \delta p) + \Delta_p\, (\dot{p}) = 0. \tag{17.102}$$

The partial time derivative vanishes since it is computed by holding pressure fixed so that the pressure increment has a zero time tendency

$$\left[\frac{\partial(\delta p)}{\partial t}\right]_p = 0. \tag{17.103}$$

Likewise, $\nabla_p\, (\delta p) = 0$. Thus, we can divide by δp to render the continuity equation

$$\nabla_p \cdot \boldsymbol{u} + \frac{\partial \dot{p}}{\partial p} = 0 \qquad \text{compressible hydrostatic.} \tag{17.104}$$

This equation is isomorphic to the incompressible continuity equation (for both hydrostatic and non-hydrostatic fluids) written using geopotential coordinates (see Chapter 25)

$$\nabla_z \cdot \boldsymbol{u} + \frac{\partial \dot{z}}{\partial z} = 0 \qquad \text{incompressible,} \tag{17.105}$$

where $w = \dot{z}$ is the vertical component to the velocity vector. In particular, note that for both cases 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 17.8.2, which starts from

$$\frac{\mathrm{D}(\rho\, \delta V)}{\mathrm{D}t} = 0. \tag{17.106}$$

In pressure coordinates the volume of the fluid element takes the form

$$\delta V = \delta x\,\delta y\,\delta z = \delta x\,\delta y\,\left[\frac{\partial z}{\partial p}\right]\delta p = -(\rho\,g)^{-1}\,\delta x\,\delta y\,\delta p. \tag{17.107}$$

Consequently,

$$0 = \frac{\mathrm{D}(\rho\,\delta V)}{\mathrm{D}t} = g^{-1}\left(\frac{\mathrm{D}\,(\delta x\,\delta y\,\delta p)}{\mathrm{D}t}\right), \tag{17.108}$$

so that

$$0 = \frac{1}{\delta x\,\delta y\,\delta p}\left(\frac{\mathrm{D}\,(\delta x\,\delta y\,\delta p)}{\mathrm{D}t}\right) = \nabla_p\cdot\boldsymbol{u} + \frac{\partial\dot{p}}{\partial p}. \tag{17.109}$$

The second step made use of the isomorphism between this result and that for equation (13.48) that holds for a geopotential vertical coordinate.

17.9.3 Incompressible fluids

Specializing to an incompressible volume conserving fluid (see Chapter 25) yields the incompressible layer thickness equation

$$\frac{\partial h}{\partial t} + \nabla_\sigma\cdot(h\,\boldsymbol{u}) + \Delta_\sigma w^{(\dot{\sigma})} = 0. \tag{17.110}$$

Further specializing to the case of zero dia-surface transport leads to

$$\frac{\partial h}{\partial t} + \nabla_\sigma\cdot(h\,\boldsymbol{u}) = 0 \qquad \text{no dia-surface transport.} \tag{17.111}$$

This case is commonly studied for adiabatic fluids using isopycnal coordinates, in which isopycnal surfaces are material (Section 46.1).

17.9.4 Rescaled geopotential coordinates

The rescaled geopotential coordinate

$$z^* = H\,\frac{z-\eta}{H+\eta} \qquad \text{and} \qquad -H \le z^* \le 0, \tag{17.112}$$

is commonly used in Boussinesq ocean models, where $z = \eta(x,y,t)$ is the ocean free surface and $z = \eta_{\mathrm{b}}(x,y) = -H(x,y)$ is the ocean bottom. The thickness of a coordinate layer is given by

$$h = \mathrm{d}z = \frac{\partial z}{\partial z^*}\,\mathrm{d}z^* = (1+\eta/H)\,\mathrm{d}z^*. \tag{17.113}$$

The depth integrated column thickness and depth integrated coordinate thickness are given by

$$\int_{-H}^{\eta}\mathrm{d}z = H+\eta \qquad \text{and} \qquad \int_{z^*(-H)}^{z^*(\eta)}\mathrm{d}z^* = H. \tag{17.114}$$

Correspondingly, the depth integrated thickness equation is given by the depth integrated volume budget derived in Section 16.3

$$\frac{\partial\eta}{\partial t} + \nabla\cdot\boldsymbol{U} + [w^{(\dot{\sigma})}_{z^*=0} - w^{(\dot{\sigma})}_{z^*=-H}] = 0. \tag{17.115}$$

We assume no volume flow through the ocean bottom so that $w^{(\dot{\sigma})}_{z^*=-H} = 0$, whereas

$$-\rho_0\,w^{(\dot{\sigma})}_{z^*=0} = Q_{\mathrm{m}} \tag{17.116}$$

is the mass flux crossing the ocean free surface.

17.10 Layer integrated tracer equation

Figure 17.5: *Illustrating the terms contributing to changes in layer tracer content according to the layer integrated tracer equation (17.121). 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 15.1.4 is given by

$$\rho\, \frac{\mathrm{D}C}{\mathrm{D}t} = -\nabla \cdot \boldsymbol{J}, \tag{17.117}$$

where $\boldsymbol{J}$ is a subgrid scale flux. Now introduce the material time derivative operator in the form (17.49b) to have

$$\rho \left[\frac{\partial C}{\partial t} + \boldsymbol{u} \cdot \nabla_\sigma C + \dot\sigma\, \partial_\sigma C \right] = -\nabla \cdot \boldsymbol{J}, \tag{17.118}$$

Multiplying by the specific thickness and making use of the mass conservation equation (17.84) renders the flux-form Eulerian equation

$$\frac{\partial (z_\sigma\, \rho\, C)}{\partial t} + \nabla_\sigma \cdot (z_\sigma\, \rho\, C\, \boldsymbol{u}) + \frac{\partial (\rho\, C\, w^{(\dot\sigma)})}{\partial \sigma} = -\left[\nabla_\sigma \cdot (z_\sigma\, \boldsymbol{J}^h) + \frac{\partial (z_\sigma\, \nabla\sigma \cdot \boldsymbol{J})}{\partial \sigma} \right], \tag{17.119}$$

where we made use of expression (8.80) for the subgrid scale operator. Now perform a layer integral as detailed in Section 17.9 and use the layer mass continuity equation (17.99) to yield the layer integrated tracer equation

$$\frac{\partial (h\, \rho\, C)}{\partial t} + \nabla_\sigma \cdot (h\, \rho\, C\, \boldsymbol{u}) + \Delta_\sigma (\rho\, C\, w^{(\dot\sigma)}) = -\left[\nabla_\sigma \cdot (h\, \boldsymbol{J}^h) + \Delta_\sigma (z_\sigma\, \nabla\sigma \cdot \boldsymbol{J}) \right]. \tag{17.120}$$

Alternatively, we can bring all terms to the left hand side renders

$$\frac{\partial (h\, \rho\, C)}{\partial t} + \nabla_\sigma \cdot (h\, \rho\, C\, \boldsymbol{u} + h\, \boldsymbol{J}^h) + \Delta_\sigma (\rho\, C\, w^{(\dot\sigma)} + J^{(\sigma)}) = 0 \tag{17.121}$$

where we wrote

$$J^{(\sigma)} = z_\sigma\, \nabla\sigma \cdot \boldsymbol{J}. \tag{17.122}$$

We illustrate contributions to the layer tracer budget (17.121) in Figure 17.5. Note that we interpret these layer integrated fields and fluxes as per the discussion in Section 17.9.1.

Part IV

Geophysical fluid dynamics

Dynamics is the area of mechanics that studies the cause of motion. For a Newtonian system, such as a geophysical fluid, the cause of motion is understood when we dissect 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 allows us to establish general principles based on underlying assumptions and/or postulates, and then to see how these principles manifest in a variety of specific contexts.

Forces of concern here 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 Coriolis and planetary centrifugal accelerations in the same manner as encountered in our study of geophysical particle mechanics in Part II.

We start the development in this part by formulating the equations of motion (linear and angular momentum) in Chapter 18, and then diving into the details of the forces that appear within these equations. These forces arise from frictional and pressure stresses (Chapters 19 and 20) as well as buoyancy (Chapter 21). Buoyancy, frictional stress and pressure stress act within a continuous fluid but are absent from a point particle system (Part II), thus motivating us to examine their physical properties in some detail.

When studying buoyancy in Chapter 21, we make use of an equation of state that provides the mass density of a fluid element as a function of thermodynamical properties of the fluid. In Chapter 22 we review the equilibrium thermodynamic principles that find use in our studies. In Chapter 23 we extend the relations found in equilibrium thermodynamic to the linear irreversible thermodynamics of relevant for a moving geophysical fluid. This extension allows us in Chapter 23 to study how mechanical energy is exchanged with internal energy in the presence of pressure work and friction.

Chapters 24 and 25 introduce a variety of approximate equations that allow us to focus on selected dynamical regimes by filtering out certain phenomena. It is here that we encounter the hydrostatic approximation and Boussinesq approximation 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 geostrophic mechanics, which is a diagnostic balance appropriate for describing large-scale geophysical flows in which pressure is balanced by the Coriolis acceleration. Geostrophic balance is but one of the many balances considered in Chapter 27, which introduces a suite of flow regimes pertaining to horizontal motions. We close this part of the book in Chapter 28, which examines the Ekman boundary layer where the Coriolis acceleration balances vertical friction.

18

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 among the continuum of fluid elements.

READER'S GUIDE TO THIS CHAPTER

We make liberal use of results from classical point particle mechanics moving around a rotating sphere as detailed in Part II and from the fluid kinematics discussed in Part III. The material in this chapter is fundamental to nearly all the subsequent chapters in this book.

18.1 Linear momentum equation **307**
18.1.1 Body forces 308
18.1.2 Contact forces 308
18.1.3 Equation of motion 309
18.1.4 Further study 310
18.2 Special forms for the momentum equation **310**
18.2.1 Spherical coordinates 310
18.2.2 Vector-invariant velocity equation 311
18.3 Exact hydrostatic balance **312**
18.4 Axial angular momentum **313**
18.4.1 Axial angular momentum conserving motion of a ring of air 314
18.4.2 Sketching the atmospheric angular momentum budget 315
18.4.3 Further study 316

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 and for an infinitesimal fluid element.

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, $\boldsymbol{f}_{\text{body}}$, multiplied by the mass of the media

$$\boldsymbol{F}_{\text{body}} = \int_{\mathcal{R}} \boldsymbol{f}_{\text{body}} \, \rho \, \mathrm{d}V. \tag{18.1}$$

For example, the effecctive gravitational force (central gravity plus planetary centrifugal) acting on a volume of fluid is given by

$$\boldsymbol{F}_{\text{gravity}} = \int_{\mathcal{R}} \boldsymbol{g} \, \rho \, \mathrm{d}V, \tag{18.2}$$

where $\boldsymbol{g}$ is the acceleration of gravity. Likewise, the Coriolis force is given by

$$\boldsymbol{F}_{\text{Coriolis}} = -2 \int_{\mathcal{R}} (\boldsymbol{\Omega} \wedge \boldsymbol{v}) \, \rho \, \mathrm{d}V. \tag{18.3}$$

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 continuous media by affecting the boundary enclosing the region. The total contact force exerted on the region through its boundaries is given by

$$\boldsymbol{F}_{\text{contact}} = \oint_{\partial\mathcal{R}} \mathbb{T} \cdot \hat{\boldsymbol{n}} \, \mathrm{d}\mathcal{S}, \tag{18.4}$$

where $\hat{\boldsymbol{n}}$ is the outward normal direction orienting the domain boundary with $\mathrm{d}\mathcal{S}$ the associated area element, and $\mathbb{T}$ is the second order *stress tensor*. Contact forces are present in continuum matter but absent in point particle matter. Hence, they represent a fundamentally new element to the fluid dynamical equations relative to the equations of point particles detailed in Part II.

Stresses from friction and pressure

As discussed in Chapter 19, there are two general types of stress that concern us: diagonal stresses associated with reversible momentum exchange through pressure, and stresses associated with irreversible exchange of momentum through friction. We thus find it convenient to decompose the stress tensor components according to

$$\mathbb{T}^{ab} = \mathbb{t}^{ab} - p \, g^{ab}. \tag{18.5}$$

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 the metric tensor that summarizes the local geometry, and it equals to the Kronecker or unit tensor for Cartesian coordinates in Euclidean space (Section 6.1). The frictional stress tensor is written $\mathbb{t}^{ab}$. 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

$$\boldsymbol{F}_{\text{contact}} = \oint_{\partial\mathcal{R}} (\mathbb{T} \cdot \hat{\boldsymbol{n}} - p\,\hat{\boldsymbol{n}})\,\mathrm{d}\mathcal{S}, \tag{18.6}$$

where the integral is taken over the bounding surface of the domain whose outward normal is $\hat{\boldsymbol{n}}$. Given this expression for contact forces acting on the boundary of a fluid domain, it is seen that positive pressure acts in the direction opposite to the surface's outward normal. That is, 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. The divergence of the stress tensor then leads to a force acting on the fluid element boundaries. The forces arising from molecular viscosity provide an irreversible exchange of momentum that acts to reduce the kinetic energy of fluid elements. 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. Nonetheless, these exchanges are also commonly parameterized via a symmetric stress tensor.

18.1.3 Equation of motion

The linear momentum of a fluid region is given by

$$\boldsymbol{P} = \int_{\mathcal{R}} \boldsymbol{v}\,\rho\,\mathrm{d}V. \tag{18.7}$$

Applying Newton's law of motion to the continuum leads to the finite volume equation of motion

$$\frac{\mathrm{d}}{\mathrm{d}t}\int_{\mathcal{R}} \boldsymbol{v}\,\rho\,\mathrm{d}V = \int_{\mathcal{R}} \rho\,\boldsymbol{f}_{\text{body}}\,\mathrm{d}V + \oint_{\partial\mathcal{R}} \mathbb{T}\cdot\hat{\boldsymbol{n}}\,\mathrm{d}\mathcal{S}. \tag{18.8}$$

The time derivative can be either material, as for a constant mass fluid region moving with the barycentric velocity, or Eulerian, as for a fixed region in space (see Section 15.3). Appying Gauss's law (Section 2.7.2) to the area integral yields

$$\frac{\mathrm{d}}{\mathrm{d}t}\int_{\mathcal{R}} \rho\,\boldsymbol{v}\,\mathrm{d}V = \int_{\mathcal{R}} (\rho\,\boldsymbol{f}_{\text{body}} + \nabla\cdot\mathbb{T})\,\mathrm{d}V. \tag{18.9}$$

General form of the equation of motion for a fluid element

Since the volume under consideration is arbitrary, the integral relation (18.9) is satisfied for an arbitrary region. We apply the result to an infinitesimal fluid element moving with the flow so that

$$\frac{\mathrm{D}(\rho\,\boldsymbol{v}\,\delta V)}{\mathrm{D}t} = \delta V\,(\rho\,\boldsymbol{f}_{\text{body}} + \nabla\cdot\mathbb{T}). \tag{18.10}$$

Assuming the mass for the fluid element is constant then leads to the equation of motion

$$\rho\,\frac{\mathrm{D}\boldsymbol{v}}{\mathrm{D}t} = \rho\,\boldsymbol{f}_{\text{body}} + \nabla\cdot\mathbb{T}. \tag{18.11}$$

Momentum equation for a rotating fluid in a gravitational field

Now specialize the momentum equation (18.11) to suit our needs. We first write the stress tensor in terms of the deviatoric component from friction and a diagonal component from pressure (equation (18.6))

$$\rho \frac{\mathrm{D}\boldsymbol{v}}{\mathrm{D}t} = \rho \, \boldsymbol{f}_{\text{body}} - \nabla p + \nabla \cdot \mathbb{T}. \tag{18.12}$$

Next, move to a rotating terrestrial reference frame and thus expose the Coriolis acceleration and the effective gravitational force (Section 9.13)

$$\rho \frac{\mathrm{D}\boldsymbol{v}}{\mathrm{D}t} + 2\,\rho\,\boldsymbol{\Omega} \wedge \boldsymbol{v} = -\rho\,\nabla\Phi - \nabla p + \nabla \cdot \mathbb{T}. \tag{18.13}$$

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 9.11). 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 (18.13) for the equation of motion offers a suitable starting point for studies of geophysical fluid dynamics. It often goes by the name of *Navier-Stokes* equation. However, that name is more correctly applied to the non-rotating case with a specific form for the friction operator. We thus refer to equation (18.13) as Newton's law of motion for a rotating fluid.

18.1.4 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. Fundamentals of geopotential coordinates and sea level are summarized in *Gregory et al.* (2019).

18.2 Special forms for the momentum equation

We here display some special forms of the momentum equation that will prove of use in our study.

18.2.1 Spherical coordinates

Geophysical fluids move on a rotating planet with the planet commonly assumed to have a spherical geometry. We thus find it useful to display the spherical coordinate form for the equations of motion. For this purpose, make use of the acceleration as derived in Section 9.13.3 for the point particle. In particular, use the geopotential coordinate to measure radial distances from the center of the sphere (equation (23.91)), as well as the longitude and latitude angular coordinates defined by Figure 7.1. Notably, the point particle time derivative translates into a material time derivative for fluid elements. We are thus led to the spherical equations of motion in their full glory

$$\frac{\mathrm{D}u}{\mathrm{D}t} + \frac{u\,(w - v\tan\phi)}{r} + 2\,\Omega\,(w\cos\phi - v\sin\phi) = -\frac{1}{\rho\,r\cos\phi}\,\frac{\partial p}{\partial\lambda} + F^{\lambda} \tag{18.14}$$

$$\frac{\mathrm{D}v}{\mathrm{D}t} + \frac{v\,w + u^2\tan\phi}{r} + 2\,\Omega\,u\sin\phi = -\frac{1}{\rho\,r}\,\frac{\partial p}{\partial\phi} + F^{\phi} \tag{18.15}$$

$$\frac{\mathrm{D}w}{\mathrm{D}t} - \frac{u^2 + v^2}{r} - 2\,\Omega\,u\cos\phi = -g - \frac{1}{\rho}\,\frac{\partial p}{\partial r} + F^{r}, \tag{18.16}$$

where we introduced the spherical components to the friction acceleration

$$\boldsymbol{F} = F^{\lambda}\,\hat{\boldsymbol{\lambda}} + F^{\phi}\,\hat{\boldsymbol{\phi}} + F^{r}\,\hat{\boldsymbol{r}}. \tag{18.17}$$

We also note the spherical coordinate form for the gradient operator

$$\nabla = \frac{\hat{\boldsymbol{\lambda}}}{r\cos\phi}\,\frac{\partial}{\partial\lambda} + \frac{\hat{\boldsymbol{\phi}}}{r}\,\frac{\partial}{\partial\phi} + \hat{\boldsymbol{r}}\,\frac{\partial}{\partial r}, \tag{18.18}$$

as well as the material time derivative operator

$$\frac{\mathrm{D}}{\mathrm{D}t} = \frac{\partial}{\partial t} + \boldsymbol{v}\cdot\nabla = \frac{\partial}{\partial t} + \frac{u}{r\cos\phi}\,\frac{\partial}{\partial\lambda} + \frac{v}{r}\,\frac{\partial}{\partial\phi} + w\,\frac{\partial}{\partial r}. \tag{18.19}$$

We can write the spherical momentum equations in a bit more compact form by introducing the spherical coordinate velocity field (see equation (9.53))

$$\boldsymbol{v} = \boldsymbol{u} + \hat{\boldsymbol{r}}\,w = u\,\hat{\boldsymbol{\lambda}} + v\,\hat{\boldsymbol{\phi}} + w\,\hat{\boldsymbol{r}} \tag{18.20}$$

and the corresponding spherical coordinate acceleration

$$\boldsymbol{A}_{\text{sphere}} = \frac{\mathrm{D}u}{\mathrm{D}t}\,\hat{\boldsymbol{\lambda}} + \frac{\mathrm{D}v}{\mathrm{D}t}\,\hat{\boldsymbol{\phi}} + \frac{\mathrm{D}w}{\mathrm{D}t}\,\hat{\boldsymbol{r}}. \tag{18.21}$$

We also introduce the expression (9.83c) for the metric acceleration to render

$$\rho\frac{\mathrm{D}\boldsymbol{v}}{\mathrm{D}t} + 2\,\rho\,\boldsymbol{\Omega}\wedge\boldsymbol{v} = -\rho\,\nabla\Phi - \nabla p + \rho\,\boldsymbol{F}, \tag{18.22}$$

where we have the acceleration relative to the rotating frame

$$\frac{\mathrm{D}\boldsymbol{v}}{\mathrm{D}t} = \boldsymbol{A}_{\text{sphere}} + \frac{1}{r}\left[u\tan\phi\,(\hat{\boldsymbol{r}}\wedge\boldsymbol{v}) + w\,\boldsymbol{u} - \hat{\boldsymbol{r}}\,\boldsymbol{u}\cdot\boldsymbol{u}\right]. \tag{18.23}$$

For some purposes it is convenient to combine one piece of the metric acceleration to the Coriolis acceleration to yield

$$\boldsymbol{A}_{\text{sphere}} + \frac{1}{r}\left[w\,\boldsymbol{u} - \hat{\boldsymbol{r}}\,\boldsymbol{u}\cdot\boldsymbol{u}\right] + \left[2\,\boldsymbol{\Omega} + \frac{u\tan\phi\,\hat{\boldsymbol{r}}}{r}\right]\wedge\boldsymbol{v} = -\left[\frac{\rho\nabla\Phi + \nabla p}{\rho}\right] + \boldsymbol{F}. \tag{18.24}$$

18.2.2 Vector-invariant velocity equation

The metric terms appearing in the momentum equation (18.24) 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 (2.37)) for the nonlinear self-advection term

$$(\boldsymbol{v}\cdot\nabla)\,\boldsymbol{v} = \boldsymbol{\omega}\wedge\boldsymbol{v} + \nabla(\boldsymbol{v}\cdot\boldsymbol{v})/2, \tag{18.25}$$

where $\boldsymbol{\omega} = \nabla\wedge\boldsymbol{v}$ is the vorticity, which is studied in Part VI of this book. We derive the corresponding *vector-invariant* form of the velocity equation using Cartesian coordinates and then invoke general covariance (Section 5.2) to extend the result to arbitrary coordinates.[1] Making use of equation (18.25) thus leads to the material acceleration

$$\frac{\mathrm{D}\boldsymbol{v}}{\mathrm{D}t} = \frac{\partial\boldsymbol{v}}{\partial t} + \boldsymbol{\omega}\wedge\boldsymbol{v} + \nabla(\boldsymbol{v}\cdot\boldsymbol{v})/2 \tag{18.26}$$

[1] See Section 4.4.4 of *Griffies* (2004) for a detailed derivation using general coordinates.

so that the momentum equation (18.13) becomes the vector-invariant velocity equation

$$\frac{\partial \boldsymbol{v}}{\partial t} + (2\,\boldsymbol{\Omega} + \boldsymbol{\omega}) \wedge \boldsymbol{v} = -\nabla(\Phi + \boldsymbol{v}\cdot\boldsymbol{v}/2) + (1/\rho)\,(-\nabla p + \nabla\cdot\mathbb{T}). \tag{18.27}$$

This equation forms the basis for Bernoulli's theorem in Section 23.4.4 as well as the vorticity equation in Chapter 34.

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, $\boldsymbol{v} = 0$. The equation of motion (18.13) has an exact static solution so long as the pressure gradient force balances the gravitational force

$$\nabla p = -\rho\,\nabla\Phi, \tag{18.28}$$

and where the frictional stress tensor has zero divergence. Equation (18.28) constitutes the *hydrostatic balance*. As justified in Chapter 24, 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 purely static flow, in which case the hydrostatic balance (18.28) is an exact solution to the equation of motion.

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

  $$\nabla\rho \wedge \nabla\Phi = 0, \tag{18.29}$$

  which in turn means that density surfaces are parallel to geopotentials so that

  $$\rho = \rho(\Phi) \qquad \text{static fluid.} \tag{18.30}$$

  For the geopotential $\Phi = g\,z$, a static fluid is realized if the density has no horizontal gradients; i.e., it is everywhere flat so that

  $$\rho = \rho(z) \qquad \text{static fluid with } \Phi = g\,z. \tag{18.31}$$

  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 (18.28) onto an infinitesimal space increment, $\mathrm{d}\boldsymbol{x}$, renders

  $$\mathrm{d}\boldsymbol{x}\cdot\nabla p = -\rho\,\mathrm{d}\boldsymbol{x}\cdot\nabla\Phi \Longrightarrow \frac{\mathrm{d}p}{\mathrm{d}\Phi} = -\rho. \tag{18.32}$$

  Hence, the difference in hydrostatic pressure between any two geopotential is given by the integral

  $$p(\Phi_2) - p(\Phi_1) = -\int_{\Phi_1}^{\Phi_2} \rho(\Phi)\,\mathrm{d}\Phi. \tag{18.33}$$

Figure 18.1: *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 $\mathrm{d}z > 0$, density ρ, and horizontal cross-sectional area $\mathrm{d}A$. 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, $\boldsymbol{F}^{\mathrm{press}}(z + \mathrm{d}z) = -\hat{\boldsymbol{z}}\, p(z + \mathrm{d}z)\, \mathrm{d}A$, whereas the pressure force at the bottom of the layer acts upward, $\boldsymbol{F}^{\mathrm{press}}(z) = +\hat{\boldsymbol{z}}\, p(z)\, \mathrm{d}A$. 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, $p(z + \mathrm{d}z) - p(z) = -g\,\rho(z)\,\mathrm{d}z$, so that pressure at the top of the layer is less than that at the bottom.*

If $\Phi = g\,z$ then we recover the familiar expression

$$p(z_2) - p(z_1) = -g \int_{z_1}^{z_2} \rho(z)\, \mathrm{d}z, \tag{18.34}$$

so that the difference in hydrostatic presure 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 18.1.

18.4 Axial angular momentum

As in our discussion of a point particle in Section 10.6, the axial angular momentum of a fluid element is given by

$$L^z = (\rho\, \delta V)\, r_\perp\, (u + r_\perp\, \Omega) \equiv (\rho\, \delta V)\, l^z \tag{18.35}$$

where

$$l^z = r_\perp\, (u + r_\perp\, \Omega) \tag{18.36}$$

is the axial angular momentum per unit mass, and the distance to the polar rotation axis,

$$r_\perp = r \cos\phi \tag{18.37}$$

is the moment-arm for determining the torques acting on a fluid element. Making use of the zonal momentum equation (18.14) leads to the material time change

$$\frac{\mathrm{D}l^z}{\mathrm{D}t} = (u + 2\,\Omega\, r_\perp)\, \frac{\mathrm{D}r_\perp}{\mathrm{D}t} + r_\perp\, \frac{\mathrm{D}u}{\mathrm{D}t} \tag{18.38a}$$

$$= (u + 2\,\Omega\, r_\perp)\, \frac{\mathrm{D}r_\perp}{\mathrm{D}t} + (u + 2\,\Omega\, r_\perp)\, (v \sin\phi - w \cos\phi) - \frac{1}{\rho}\, \frac{\partial p}{\partial \lambda} \tag{18.38b}$$

$$= (u + 2\,\Omega\, r_\perp)\, \left[\frac{\mathrm{D}r_\perp}{\mathrm{D}t} + v \sin\phi - w \cos\phi\right] - \frac{1}{\rho}\, \frac{\partial p}{\partial \lambda} \tag{18.38c}$$

$$= -\frac{1}{\rho}\, \frac{\partial p}{\partial \lambda}, \tag{18.38d}$$

Figure 18.2: *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.*

so that

$$\rho \frac{\mathrm{D}l^z}{\mathrm{D}t} = -\frac{\partial p}{\partial \lambda} \Longrightarrow \frac{\partial(\rho\, l^z)}{\partial t} + \nabla \cdot (\rho\, l^z\, \boldsymbol{v}) = -\frac{\partial p}{\partial \lambda}. \tag{18.39}$$

Comparing to the result for a point particle in Section 10.6 reveals that the axial angular momentum of a fluid element is affected by the zonal pressure gradient. Hence, in the absence of a zonal pressure gradient the axial angular momentum is materially invariant just like for the point particle. In particular, the physical results described in Section 10.7 hold whereby 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 along the trajectory of a fluid element. However, on a smooth spherical planet without meridional boundaries there is a zero zonally integrated zonal pressure gradient

$$\int_{\text{ring}} \frac{\partial p}{\partial \lambda}\, \mathrm{d}\lambda = 0. \tag{18.40}$$

Hence, a constant mass ring of fluid circling the planet (Figure 18.2) will preserve its axial angular momentum in the absence of friction

$$\frac{\mathrm{d}}{\mathrm{d}t} \oint_{\text{ring}} \rho\, l^z\, \mathrm{d}V = \oint_{\text{ring}} \rho \frac{\mathrm{D}l^z}{\mathrm{D}t}\, \mathrm{d}V = 0. \tag{18.41}$$

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 $l^z = \Omega\, R_\mathrm{e}^2$. Moving the ring vertically to $r \neq R_\mathrm{e}$ while maintaining constant axial angular momentum induces a zonal velocity given by

$$u = \frac{\Omega\, (R_\mathrm{e}^2 - r^2)}{r}. \tag{18.42}$$

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

$$u = \frac{\Omega\, R_e \sin^2\phi}{\cos\phi}. \tag{18.43}$$

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, and consider a ring of radius R_e that starts with zero relative velocity at the equator. Equation (18.43) says that the westerly winds induced by axial angular momentum conserving motion have the following speeds at a selection of latitudes

$$u(10^\circ) = 14 \text{ m s}^{-1} \qquad u(20^\circ) = 58 \text{ m s}^{-1} \qquad u(30^\circ) = 134 \text{ m s}^{-1}. \tag{18.44}$$

The values at higher latitudes is unbounded since $\cos\phi \to 0$ 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: frictional dissipation between the atmosphere and land and 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.[2] Rather than an inviscid fluid, we consider the dissipation due to boundary layer 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 18.3, 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 49.2.4). 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.

[2] In 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 18.3: *A sketch of the Hadley circulation. At the low latitudes, low level winds are predominently 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 momentun, 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.*

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).

19

Stress in fluids

A fluid element experiences two kinds of forces. *External* or *body* forces act throughout the fluid element. The accumulated effects from a body force results from integrating the body force over each point within the 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. The exchange of dynamical properties is mediated by mechanical interactions, with this exchange mathematically formulated as a stress (force per unit area) acting on the boundary of the fluid element. Contact forces satisfy Newton's third law, also known as the action/reaction law. Hence, the net contact force acting on a finite region arise just from the contact forces acting at the region boundary. Contact forces communicated through stresses embody the fundamental distinction between the dynamics of a fluid element and the dynamics of a point particle.

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 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 for arbitrary coordinates through the general tensor analysis detailed in Chapter 6.

19.1 Cauchy's stress principle and Newton's laws **318**
19.1.1 Cauchy's stress principle 318

19.1.2 Newton's third law and local equilibrium of contact stresses 319
19.1.3 Comments on the local equilibrium relation 320
19.2 The stress tensor **321**
19.3 Angular momentum and stress tensor symmetry **322**
19.4 Flux-form Eulerian momentum equation **324**
19.5 Linear momentum for arbitrary regions **325**
19.6 Relating stress to pressure and strain **325**
19.6.1 Thermodynamic pressure and mechanical pressure 326
19.6.2 Friction tensor for a Newtonian fluid 326
19.6.3 General properties of the friction tensor 326
19.6.4 Friction tensor related to the rate of strain tensor 327
19.6.5 Frictional stress for a vertically sheared zonal flow 327
19.6.6 Comments and further study 328
19.7 The Reynolds number and turbulence **328**
19.8 Stress on an interface **329**
19.8.1 General formulation 330
19.8.2 Solid material boundary 330
19.8.3 No-slip boundary condition 331
19.8.4 Lagrangian interface 332
19.8.5 Permeable interface 332
19.8.6 Comments 333
19.9 Exercises **333**

19.1 Cauchy's stress principle and Newton's laws

We here develop some general properties of contact forces and the associated contact stresses. For that purpose, consider an arbitrary smooth closed region, $\mathcal{R}$, of fluid with volume $V = \int_{\mathcal{R}} \mathrm{d}V$ and mass $M = \int_{\mathcal{R}} \rho \, \mathrm{d}V$ (Figure 19.1). Furthermore, let $\partial\mathcal{R}$ be the bounding surface for the region, and let $\hat{\boldsymbol{n}}$ 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 contact stress vector (force per unit area) acting at a point on $\partial\mathcal{R}$. Cauchy's stress principle asserts that the contact stress vector is a function of the position, time, and boundary normal

$$\boldsymbol{\tau} = \boldsymbol{\tau}(\boldsymbol{x}, t, \hat{\boldsymbol{n}}). \tag{19.1}$$

For example, a static fluid in hydrostatic balance has the contact stress vector equal to just the pressure. Furthermore, the stress from pressure is oriented along the inward normal to reflect the compressive nature of the stress

$$\boldsymbol{\tau} = -p(\boldsymbol{x}, t)\, \hat{\boldsymbol{n}} \qquad \text{static fluid in hydrostatic balance.} \tag{19.2}$$

We return to this form of the stress vector in Section 19.6.1.

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 Chapter 44, 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 dynamical equations in this book.

19.1.2 Newton's third law and local equilibrium of contact stresses

Newton's second law says that unbalanced forces acting on a physical system, such as a region of fluid, affect a time change to the system's linear momentum. Consider a region, $\mathcal{R}(\boldsymbol{v})$, 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

$$\frac{\mathrm{d}}{\mathrm{d}t}\int_{\mathcal{R}(\boldsymbol{v})} \boldsymbol{v}\,\rho\,\mathrm{d}V = \int_{\mathcal{R}(\boldsymbol{v})} \boldsymbol{f}\,\rho\,\mathrm{d}V + \oint_{\partial\mathcal{R}(\boldsymbol{v})} \boldsymbol{\tau}\,\mathrm{d}\mathcal{S}, \tag{19.3}$$

where $\int_{\mathcal{R}(\boldsymbol{v})} \boldsymbol{f}\,\rho\,\mathrm{d}V$ 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 in the volume integrals are well behaved 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 contact 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

$$\lim_{L\to 0}\frac{1}{L^2}\oint_{\partial\mathcal{R}(\boldsymbol{v})} \boldsymbol{\tau}\,\mathrm{d}\mathcal{S} = 0. \tag{19.4}$$

This behavior means that contact forces at a point in the fluid must be in local equilibrium.

A direct implication of the local equilibrium statement is that contact stress vectors that respect Cauchy's principle (19.1) satisfy

$$\boldsymbol{\tau}(\boldsymbol{x}, t, \hat{\boldsymbol{n}}) = -\boldsymbol{\tau}(\boldsymbol{x}, t, -\hat{\boldsymbol{n}}). \tag{19.5}$$

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. It is of fundamental importance throughout our study of contact forces and their associated stresses acting within the fluid and at fluid 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, the simplest stress we consider in this chapter is that from pressure, which acts solely in a compressive manner and takes the form

$$\boldsymbol{\tau}^{\text{press}}(\boldsymbol{x}, t, \hat{\boldsymbol{n}}) = -p\,\hat{\boldsymbol{n}}. \tag{19.6}$$

This stress trivially satisfues the Newton's third law relation (19.5) since

$$\boldsymbol{\tau}^{\text{press}}(\boldsymbol{x}, t, \hat{\boldsymbol{n}}) = -\boldsymbol{\tau}^{\text{press}}(\boldsymbol{x}, t, -\hat{\boldsymbol{n}}). \tag{19.7}$$

Figure 19.1: *Schematic of the net body force, $\boldsymbol{F}_{\text{body}}$, acting on a finite region of fluid, plus the accumulation of contact stress vectors, $\boldsymbol{\tau}$, acting on the region boundaries. The net body force is determined by a volume integral of the body force at each point within the volume. In contrast, since the contact stresses are in local equilibrium, the volume integral of the contact stress divergence reduces to an area integral of the contact stress over the region boundary. 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.*

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 contact forces, such as pressure, cannot lead to motion. However, that suggestion is incorrect since contact 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 contact forces 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 contact stresses over finite regions within the ocean fluid lead to accelerations; e.g., ocean circulation. However, when integrated over the full ocean domain, all contact stresses cancel pointwise. Consequently, the net contact forces acting on the full ocean domain reduce to those 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 contact forces acting over its boundaries plus the volume integrated body forces from gravity and Coriolis sum to zero.

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

$$\boldsymbol{F}^{\text{pressure}} = -\oint_{\partial\mathscr{R}} p\,\hat{\boldsymbol{n}}\,\mathrm{d}\mathcal{S} = -\int_{\mathscr{R}} \nabla p\,\mathrm{d}V, \tag{19.8}$$

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 accelerations from the contact pressure stress is referred to as *form stress*. This name arises since the stress depends on the form, or shape, of the interface on which pressure acts.

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.*

19.2 The stress tensor

Cauchy's stress principle reduces the mathematical complexity of describing contact stress vectors. To further pursue that description, 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, $\hat{\boldsymbol{n}} = (\hat{n}_1, \hat{n}_2, \hat{n}_3)$, 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 coordintes, such as Cartesian used here, we can make use of general tensor analysis (Chapters 5 and 6) to move from specific coordinates to general coordinates.

In the limit that the tetrahedron size goes to zero, local equilibrium of the contact forces means that

$$-\sum_{m=1}^{3} \boldsymbol{\tau}_{(m)}\, \mathrm{d}A_m + \boldsymbol{\tau}_{\hat{\boldsymbol{n}}}\, \mathrm{d}A = 0, \tag{19.9}$$

where we use the shorthand expression for the outward normal directed stress vector

$$\boldsymbol{\tau}(\boldsymbol{x}, t, \hat{\boldsymbol{n}}) = \boldsymbol{\tau}_{\hat{\boldsymbol{n}}}. \tag{19.10}$$

In equation (19.9), $\boldsymbol{\tau}_{(m)}\, \mathrm{d}A_m$ (no implied summation) is the contact force vector acting on the face with outward normal parallel to the corresponding coordinate axis and $\boldsymbol{\tau}_{\hat{\boldsymbol{n}}}\, \mathrm{d}A$ is the contact force acting on the slanted face with outward normal $\hat{\boldsymbol{n}}$. 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

$$\mathrm{d}A_m = \hat{n}_m \, \mathrm{d}A, \tag{19.11}$$

so that the local equilibrium relation (19.9) becomes

$$\boldsymbol{\tau}_{\hat{\boldsymbol{n}}} - \sum_{m=1}^{3} \boldsymbol{\tau}_{(m)} \, \hat{n}_m = 0. \tag{19.12}$$

Figure 19.3: *Tetrahedron region of fluid with contact stresses acting on the four faces. Note that the stresses are not necessarily directed normal to the faces. Local equilibrium of contact stresses means that the accumulation of these four stresses around the region adds to zero as the region volume goes to zero.*

Now organize the stress components into a 3×3 matrix, $\mathbb{T}_{mn}$, measuring the $n'th$ component of the $m'th$ contact stress $\boldsymbol{\tau}_{(m)}$. That is, let $\mathbb{T}_{mn}$ measure the force per area in the $n-$direction along a surface whose outward normal points in the m-direction, as in Figure 19.4. Making use of $\mathbb{T}_{mn}$ in the expression (19.12) of local equilibrium leads to

$$\sum_{m=1}^{3} \hat{n}_m \, \mathbb{T}_{mn} = (\boldsymbol{\tau}_{\hat{\boldsymbol{n}}})_n. \tag{19.13}$$

This relation can be written as

$$\hat{\boldsymbol{n}} \cdot \mathbb{T} = \boldsymbol{\tau}_{\hat{\boldsymbol{n}}}, \tag{19.14}$$

so that the stress vector acting on a surface oriented according to a normal vector, $\hat{\boldsymbol{n}}$, equals to the projection of the stress tensor onto the normal vector.

How do we know that $\mathbb{T}_{mn}$ are components to a tensor? The normal direction, $\hat{\boldsymbol{n}}$, and the contact stress vector, $\boldsymbol{\tau}_{\hat{\boldsymbol{n}}}$, are components to first order tensors (i.e., vectors). The quotient rule from tensor analysis means that equation (19.13) yields stresses, $\mathbb{T}_{mn}$, that are components to a second order *stress tensor*. As components to a second order tensor, the $\mathbb{T}_{mn}$ transform under coordinate transformation according to the rules developed in Chapter 1 for Cartesian tensors and Chapter 6 for general tensors.

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 contact stress. We here derive a constraint placed on the stress

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 $n'th$ direction along the face with outward normal in the $m'th$ direction.*

tensor that is imposed by studying angular momentum. For this purpose, consider a constant mass fluid element that has a Cartesian position $\boldsymbol{x}$ relative to an arbitrary origin. The angular momentum of the fluid element with respect to the origin is

$$\boldsymbol{L} = \rho\,\delta V\,(\boldsymbol{x} \wedge \boldsymbol{v}), \tag{19.15}$$

and its material time evolution is

$$\frac{\mathrm{D}\boldsymbol{L}}{\mathrm{D}t} = \rho\,\delta V\,\boldsymbol{x} \wedge \frac{\mathrm{D}\boldsymbol{v}}{\mathrm{D}t}, \tag{19.16}$$

which follows since $\mathrm{D}(\rho\,\delta V)/\mathrm{D}t = 0$, $\mathrm{D}\boldsymbol{x}/\mathrm{D}t = \boldsymbol{v}$, and $\boldsymbol{v} \wedge \boldsymbol{v} = 0$. Making use of Cauchy's form for the equation of motion (18.11)

$$\rho\,\frac{\mathrm{D}\boldsymbol{v}}{\mathrm{D}t} = \rho\,\boldsymbol{f} + \nabla \cdot \mathbb{T} \tag{19.17}$$

allows us to write the angular momentum evolution as

$$\frac{\mathrm{D}\boldsymbol{L}}{\mathrm{D}t} = \delta V\,\boldsymbol{x}\,\wedge\,(\rho\boldsymbol{f} + \nabla \cdot \mathbb{T}). \tag{19.18}$$

The first term arises from body forces (e.g., central gravity, planetary centrifubal, and Coriolis) and the second term arises from contact stresses. Focusing on the contact stresses and expanding the divergence operation renders

$$\left[\frac{\mathrm{D}L_m}{\mathrm{D}t}\right]_{\text{contact stress}} = \delta V\,\epsilon_{mnp}\,x_n\,(\nabla \cdot \mathbb{T})_p \tag{19.19a}$$

$$= \delta V\,\epsilon_{mnp}\,x_n\,\partial_q \mathbb{T}_{qp} \tag{19.19b}$$

$$= \delta V\,\epsilon_{mnp}\,[\partial_q(x_n\,\mathbb{T}_{qp}) - (\partial_q x_n)\,\mathbb{T}_{qp}] \tag{19.19c}$$

$$= \delta V\,\epsilon_{mnp}\,[\partial_q(x_n\,\mathbb{T}_{qp}) - \mathbb{T}_{np}], \tag{19.19d}$$

where the final equality follows since $\partial_q x_n = \delta_{qn}$. The first term on the right hand side accounts for movement of torques within the fluid from one region to another. When integrating over a finite region, this term accounts for torques that arise from stresses acting on the region boundaries. For example, stresses at the boundary between the ocean and atmosphere create torques on both media that alter their respective angular momenta.

The second term in equation (19.19d) contributes to angular momentum throughout the fluid region. It is a source of torque that is *not* associated with body forces (gravity, Coriolis, planetary centrifugal). What are these torques? Given that they do not arise from body forces, they must be associated with internal molecular degrees of freedom that introduce *force couplets* rendering a torque at each point within the fluid. However, geophysical fluids (air and water) do not spontaneously generate angular momentum at each point within the fluid region as they have no internally generated torques. Rather, the only means for an element of water or air to modify its angular momentum is through boundary torques, transfer of torques between fluid elements, and through torques due to body forces. To remove internal torque sources requires us to consider only symmetric stress tensors

$$\mathbb{T}_{mn} = \mathbb{T}_{nm} \Longrightarrow \epsilon_{mnp}\, \mathbb{T}_{np} = 0. \tag{19.20}$$

This is a central property of the stresses acting on geophysical fluids.

19.4 Flux-form Eulerian momentum equation

We often find it useful to consider Cauchy's form of the momentum equation (19.17) 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 (14.9) so that

$$\rho \frac{\mathrm{D}\boldsymbol{v}}{\mathrm{D}t} = \rho\, [\partial_t \boldsymbol{v} + (\boldsymbol{v}\cdot\nabla)\, \boldsymbol{v}] \tag{19.21a}$$

$$= \rho\, [\partial_t \boldsymbol{v} + (\boldsymbol{v}\cdot\nabla)\, \boldsymbol{v}] + \boldsymbol{v}\, (\partial_t \rho + \nabla\cdot(\rho\, \boldsymbol{v})] \tag{19.21b}$$

$$= \partial_t(\rho\, \boldsymbol{v}) + \nabla\cdot[\rho\, \boldsymbol{v}\otimes\boldsymbol{v}], \tag{19.21c}$$

where $\boldsymbol{v}\otimes\boldsymbol{v}$ is the outer product of the velocity vector with components

$$(\boldsymbol{v}\otimes\boldsymbol{v})_{mn} = v_m\, v_n. \tag{19.22}$$

Consequently, the momentum equation (19.17) takes on the flux-form Eulerian expression

$$\partial_t(\rho\, \boldsymbol{v}) + \nabla\cdot[\rho\, \boldsymbol{v}\otimes\boldsymbol{v}] = \rho\, \boldsymbol{f} + \nabla\cdot\mathbb{T}. \tag{19.23}$$

Alternatively, we can move the advection of momentum term onto the right hand side so that

$$\partial_t(\rho\, \boldsymbol{v}) = \rho\, \boldsymbol{f} + \nabla\cdot[\mathbb{T} - \rho\, \boldsymbol{v}\otimes\boldsymbol{v}], \tag{19.24}$$

which takes on the component form

$$\partial_t(\rho\, v_m) = \rho\, f_m + \partial_n[\mathbb{T}_{mn} - \rho\, v_m\, v_n]. \tag{19.25}$$

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,

$$\mathbb{T}^{\text{kinetic}}_{mn} = -\rho\, (\boldsymbol{v}\otimes\boldsymbol{v})_{mn} = -\rho\, v_m\, v_n, \tag{19.26}$$

as the *mechanical stress* or *kinetic stress*, which arises from the mechanical interactions between moving fluid elements in addition to that from pressure and viscosity. The turbulent contribution to the mechanical stress is known as the *Reynolds stress*.

19.5 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 (15.32)

$$\frac{\mathrm{d}}{\mathrm{d}t}\left[\int_{\mathcal{R}} \varphi \,\mathrm{d}V\right] = \int_{\mathcal{R}} \frac{\partial \varphi}{\partial t}\,\mathrm{d}V + \oint_{\partial\mathcal{R}} \varphi\, \boldsymbol{v}^{(b)}\cdot\hat{\boldsymbol{n}}\,\mathrm{d}\mathcal{S}, \tag{19.27}$$

where $\boldsymbol{v}^{(b)}$ is the velocity of the region boundary, $\partial\mathcal{R}$, with $\hat{\boldsymbol{n}}$ 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 (19.25) leads to

$$\frac{\mathrm{d}}{\mathrm{d}t}\left[\int_{\mathcal{R}} \rho\, v_m \,\mathrm{d}V\right] = \int_{\mathcal{R}} \partial_t(\rho\, v_m)\,\mathrm{d}V + \oint_{\partial\mathcal{R}} (\rho\, v_m)\, \boldsymbol{v}^{(b)}\cdot\hat{\boldsymbol{n}}\,\mathrm{d}\mathcal{S} \tag{19.28a}$$

$$= \int_{\mathcal{R}} [\rho\, f_m + \partial_n(\mathbb{T}_{mn} - \rho\, v_m\, v_n)]\,\mathrm{d}V + \oint_{\partial\mathcal{R}} (\rho\, v_m)\, \boldsymbol{v}^{(b)}\cdot\hat{\boldsymbol{n}}\,\mathrm{d}\mathcal{S} \tag{19.28b}$$

$$= \int_{\mathcal{R}} \rho\, f_m\,\mathrm{d}V + \oint_{\partial\mathcal{R}} (\mathbb{T}_{mn} - \rho\, v_m\, v_n)\,\hat{n}_n\,\mathrm{d}\mathcal{S} + \oint_{\partial\mathcal{R}} (\rho\, v_m)\, \boldsymbol{v}^{(b)}\cdot\hat{\boldsymbol{n}}\,\mathrm{d}\mathcal{S} \tag{19.28c}$$

$$= \int_{\mathcal{R}} \rho\, f_m\,\mathrm{d}V + \oint_{\partial\mathcal{R}} [\mathbb{T}_{mn} + \rho\, v_m\,(v_n^{(b)} - v_n)]\,\hat{n}_n\,\mathrm{d}\mathcal{S}. \tag{19.28d}$$

We can write this relation in the more elegant geometric form as

$$\frac{\mathrm{d}}{\mathrm{d}t}\left[\int_{\mathcal{R}} \rho\, \boldsymbol{v}\,\mathrm{d}V\right] = \int_{\mathcal{R}} \rho\, \boldsymbol{f}\,\mathrm{d}V + \oint_{\partial\mathcal{R}} [\mathbb{T} + \rho\, \boldsymbol{v}\otimes(\boldsymbol{v}^{(b)} - \boldsymbol{v})]\cdot\hat{\boldsymbol{n}}\,\mathrm{d}\mathcal{S}. \tag{19.29}$$

As a relation between geometric objects (vectors and tensors), the momentum budget (19.29) 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 contact stresses acting on the region boundary. Notably, the contact 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 19.8 we specialize the budget (19.29) to an infinitesimally thin interface, with that analysis used to develop stress conditions for a surface within a single fluid media as well as for the boundary between two fluids.

We refer to a *Lagrangian region* as one that moves with the barycentric velocity so that $\boldsymbol{v}^{(b)} = \boldsymbol{v}$, in which case the mechanical stress is eliminated from the finite volume momentum budget (19.29). In fact, to eliminate the mechanical stress only requires the normal components of the boundary velocity to equal that of a fluid element, $(\boldsymbol{v}-\boldsymbol{v}^{(b)})\cdot\hat{\boldsymbol{n}} = 0$. In either case we distinguish a Lagrangian region by writing $\mathcal{R}(\boldsymbol{v})$ to emphasize that the region moves with the barycentric fluid velocity, $\boldsymbol{v}$. For this case the linear momentum is only affected by body forces as well as contact stresses contained in the stress tensor

$$(\boldsymbol{v} - \boldsymbol{v}^{(b)})\cdot\hat{\boldsymbol{n}} = 0 \Longrightarrow \frac{\mathrm{d}}{\mathrm{d}t}\left[\int_{\mathcal{R}(\boldsymbol{v})} \rho\, \boldsymbol{v}\,\mathrm{d}V\right] = \int_{\mathcal{R}(\boldsymbol{v})} \rho\, \boldsymbol{f}\,\mathrm{d}V + \oint_{\partial\mathcal{R}(\boldsymbol{v})} \mathbb{T}\cdot\hat{\boldsymbol{n}}\,\mathrm{d}\mathcal{S}. \tag{19.30}$$

This relation is Reynold's transport theorem (Section 15.3.5) as applied to linear momentum.

19.6 Relating stress to pressure and 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. In the present context, a constitutive relation provides a relation between stress and properties of the fluid as well as properties of the fluid flow (e.g., pressure, rate of strain).

19.6.1 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 said to be in *hydrostatic balance* and the corresponding stress tensor and stress vector are written

$$\mathbb{T}_{mn} = -p\,\delta_{mn} \Longleftrightarrow \mathbb{T}\cdot\hat{\boldsymbol{n}} = -p\,\hat{\boldsymbol{n}}, \tag{19.31}$$

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 thermodynamcis in Chapter 22. If we furthermore assume that local thermodynamic equilibrium is maintained for fluid elements within a moving fluid, then we can continue making this identification between mechanical pressure and thermodynamical pressure (see Section 4.5 of *Kundu et al.* (2012) or Section 1.10 of *Salmon* (1998)).

However, when assuming the fluid is incompressible (Chapter 16) as for the oceanic Boussinesq fluid (Chapter 25), we lose the equality between the mechanical pressure and thermodynamical pressure. Indeed, when working with an incompressible fluid we only concern ourselves with mechanical pressure, with mechanical pressure instantaneously conforming to the needs of incompressibility (see Exercise 25.2 and Section 39.4). In this manner, an incompressible fluid generally does not have access to the thermodynamical pressure. It is for this reason that the equation of state for *in situ* density in an incompressible fluid, energetic consistency requires us to make use of the geopotential rather than the thermodynamic pressure (see Section 25.5).

19.6.2 Friction tensor for a Newtonian fluid

A moving fluid has a more complex stress relation than a static fluid. To account for such complexity we introduce an additional contribution to the stress tensor

$$\mathbb{T}_{mn} = -p\,\delta_{mn} + \mathbb{\tau}_{mn} \Longleftrightarrow \mathbb{T} = -p\,\mathbb{I} + \mathbb{\tau}. \tag{19.32}$$

The pressure term remains isotropic. We introduced the additional stress tensor $\mathbb{\tau}$ referred to as the *friction tensor*. This stress captures the irreversible exchanges of momentum between moving fluid elements arising from viscosity. The friction tensor vanishes in a perfect fluid where viscosity is assumed to be zero. It also vanishes when the fluid is static since there are no strains between fluid elements that can be transferred, through viscosity, into a stress. When it has a zero trace, $\mathbb{\tau}_{mm} = 0$, then it is called the *deviatoric* stress tensor, which we further discuss below.

19.6.3 General properties of the friction tensor

To help develop some of the general properties of the friction tensor, consider a fluid in uniform motion. Boosting the reference frame allows us to move to a reference frame where the fluid is static. Through Galilean invariance (Section 12.6) we expect the dynamics to remain unchanged. Since the friction tensor vanishes when the fluid is static, Galilean invariance implies that the friction tensor vanishes when the fluid undergoes uniform motion in any direction. We can extend this result to finite regions of the fluid where a group of fluid elements is moving coherently in a

single direction. Moving to a local rest frame again renders a zero friction tensor when the fluid locally has uniform motion; i.e., zero velocity gradients in the flow region.

Galilean invariance thus leads us to consider a friction tensor that is a function of gradients in the velocity field, $\partial_m v_n$. Furthemore, 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 13.2.4. Such fluids are known as *Newtonian fluids*. The linear stress-rate of strain relation takes the same mathematical form as *Hooke's Law* used in the study of elastic materials.

19.6.4 Friction tensor related to the rate of strain tensor

The simplest form of this relation is when there is a single viscosity so that the relation between the stress and rate of strain takes the form

$$\mathbb{T}_{mn} = \rho\,\nu\,(\partial_m v_n + \partial_n v_m) = 2\,\rho\,\nu\,\mathbb{S}_{mn}, \tag{19.33}$$

with $\nu > 0$ a kinematic viscosity (dimensions of squared length per time) (see Section 47.2.5) and $\mathbb{S}_{mn}$ components to the rate of strain tensor introduced in Section 13.2.6. Note that with this form of the friction tensor the trace of the full stress tensor takes the form

$$\mathbb{T}_{mm} = -p + 2\,\rho\,\nu\,\nabla\cdot\boldsymbol{v}. \tag{19.34}$$

The trace deviates from pressure in the presence of velocity divergence. However, for an incompressible fluid in which $\nabla\cdot\boldsymbol{v} = 0$ (Chapter 16), the trace equals to minus the pressure, in which case $\mathbb{T}_{mn}$ is known as the *deviatoric* stress tensor.[1] Even for a compressible fluid, we generally assume the friction tensor has a zero trace.

The viscous force per volume is given by the divergence of the stress tensor

$$\rho\,F_m = 2\,\partial_n\,(\rho\,\nu\,\mathbb{S}_{mn}). \tag{19.35}$$

Assuming the dynamic viscosity, $\rho\,\nu$, is constant leads to

$$\rho\,\boldsymbol{F} = \rho\,\nu\,[\nabla(\nabla\cdot\boldsymbol{v}) + \nabla^2\boldsymbol{v}] \qquad \text{if } \rho\,\nu \text{ is constant.} \tag{19.36}$$

Furthermore, if the flow is incompressible then

$$\rho\,\boldsymbol{F} = \rho\,\nu\,\nabla^2\boldsymbol{v} \qquad \text{if } \rho\,\nu \text{ is constant and } \nabla\cdot\boldsymbol{v} = 0. \tag{19.37}$$

Again, even for a compressible fluid, we generally assume the incompressible form (19.37) for the friction force.

19.6.5 Frictional stress for a vertically sheared zonal flow

Consider the case of an incompressible flow field that only has a zonal component with a vertical shear (Figure 19.5)

$$\boldsymbol{v} = u(z)\,\hat{\boldsymbol{x}}. \tag{19.38}$$

In this case the only non-zero 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 $\hat{\boldsymbol{z}}$

[1] A 2nd order tensor in 3-dimensions, $\mathbb{D}$, has a *deviator* with components given by $\mathbb{D}'_{mn} = \mathbb{D}_{mn} - (1/3)\,\delta_{mn}\,\mathbb{D}_{mm}$.

direction. The frictional force acting on that area is given by the area integral of the frictional stress

$$\boldsymbol{F}_{\text{area}} = \int \mathbb{T} \cdot \hat{\boldsymbol{n}} \, \mathrm{d}\mathcal{S} = \int \mathbb{T} \cdot \hat{\boldsymbol{z}} \, \mathrm{d}x \, \mathrm{d}y = \frac{\hat{\boldsymbol{x}}}{2} \, \rho \, \nu \, A \, \frac{\partial u}{\partial z}, \tag{19.39}$$

where $A = \int \mathrm{d}x \, \mathrm{d}y$ is the horizontal area. 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 $\boldsymbol{v} = u(z) \, \hat{\boldsymbol{x}}$ we have

$$\left[\frac{\partial(\rho \, v_m)}{\partial t} \right]_{\text{viscous}} = \partial_n \mathbb{T}_{nm} \Longrightarrow \left[\frac{\partial(\rho \, u)}{\partial t} \right]_{\text{viscous}} = \rho \, \nu \, \frac{\partial^2 u}{\partial z^2}, \tag{19.40}$$

so that zonal momentum is preferentially deposited to or removed from regions with high vertical curvature in the zonal velocity.

Figure 19.5: *Sample profile of zonal velocity possessing a vertical shear: $\boldsymbol{v} = u(z) \, \hat{\boldsymbol{x}}$ and with a no-slip boundary condition at $z = 0$. The resulting zonal frictional stress arises from the nonzero vertical shear and viscosity.*

19.6.6 Comments and further study

There are more elaborate relations between the stress and rate of strain tensors than that from molecular viscosity in equation (19.33). The most general form for a Newtonian fluid introduces a fourth-order viscosity tensor as in Section 4.5 of *Kundu et al.* (2012) and Chapter 17 in *Griffies* (2004). Nonlinear relations can arise in Newtonian fluids when the viscous tensor is a function of the flow, such as with the Smagorinsky scheme commonly used for Large Eddy Simulations (LES) (*Smagorinsky*, 1993). Further nonlinearities arise in non-Newtonian fluids, which make use of products of the rate of strain tensor for computing stress.

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. *Anstey and Zanna* (2017) offer a compelling approach with a subgrid scale stress tensor containing a non-zero trace, thus resulting in a modification to pressure.

19.7 The Reynolds number and turbulence

How important is friction relative to other terms in the momentum equation? In particular, how does it compare to the material acceleration? When determining friction as per the Laplacian

operator in equation (19.37), the ratio of the scales for material (inertial) acceleration to frictional acceleration is called the Reynolds number

$$Re = \frac{\text{inertial accelerations}}{\text{frictional accelerations}} = \frac{U/T}{\nu U/L^2} = \frac{L^2/T}{\nu} = \frac{L\,U}{\nu}, \tag{19.41}$$

where U is a typical velocity scale, L is a typical length scale, and we assumed that $U = L/T$ so that the time scale is determined by advection.

The Reynolds number is dependent on the velocity and length scales. For large-scale oceanic flows, such as Gulf Stream rings (see Figure 26.1) with $L \approx 10^5$ m, $U \approx 0.1 - 1.0$ m s^{-1} and (page 75 of *Gill* (1982))

$$\nu_{\text{water}} = 10^{-6}\ \text{m}^2\ \text{s}^{-1}, \tag{19.42}$$

then

$$Re_{\text{Gulf Stream}} = 10^{10} - 10^{11}. \tag{19.43}$$

For the atmosphere, we take $L = 10^6$ m for a typical atmospheric weather system, $U = 10$ m s^{-1} for the speed, and

$$\nu_{\text{air}} = 1.4 \times 10^{-5}\ \text{m}^2\ \text{s}^{-1}, \tag{19.44}$$

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

$$Re_{\text{weather system}} = 10^{12}. \tag{19.45}$$

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.

The ocean and atmosphere exhibit a huge variety of turbulent regimes, from the macroturbulence of 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.

19.8 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 to 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 Chapter 44 since we are interested in length scales larger than the capillary-gravity waves where surface tension is important, with equation (44.16) indicating that these wavelengths are roughly $\lambda_{\text{cap-grav}} \approx 17$ cm.

19.8.1 General formulation

Formulation of the stress boundary conditions follows from applying the finite volume momentum equation (19.29) to a pillbox region straddling a moving interface such as that shown in Figure 19.6. The sides of the box have thickness h and the top and bottom have area $\delta\mathcal{S}$. In the limit that the pillbox thickness goes to zero, the volume integrals in equation (19.29) 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 pillbox boundary

$$\oint_{\partial\text{pillbox}} [\mathbb{T} + \rho\, \boldsymbol{v} \otimes (\boldsymbol{v}^{(b)} - \boldsymbol{v})] \cdot \hat{\boldsymbol{n}}\, \mathrm{d}\mathcal{S} = \oint_{\partial\text{pillbox}} [-p\, \mathbb{I} + \mathbb{T} + \rho\, \boldsymbol{v} \otimes (\boldsymbol{v}^{(b)} - \boldsymbol{v})] \cdot \hat{\boldsymbol{n}}\, \mathrm{d}\mathcal{S} = 0. \tag{19.46}$$

The end-caps on the pillbox vanish as $h \to 0$, in which case we have no constraint based on the stresses acting on the end-caps. Instead, the $h \to 0$ 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 contact stresses discussed in Section 19.1.2. For the stresses acting on the interface in Figure 19.6 we have

$$[-p_{\mathrm{A}}\, \mathbb{I} + \mathbb{T}_{\mathrm{A}} + \rho_{\mathrm{A}}\, \boldsymbol{v}_{\mathrm{A}} \otimes (\boldsymbol{v}^{(b)} - \boldsymbol{v}_{\mathrm{A}})] \cdot \hat{\boldsymbol{n}}_{\mathrm{A}} + [-p_{\mathrm{B}}\, \mathbb{I} + \mathbb{T}_{\mathrm{B}} + \rho_{\mathrm{B}}\, \boldsymbol{v}_{\mathrm{B}} \otimes (\boldsymbol{v}^{(b)} - \boldsymbol{v}_{\mathrm{B}})] \cdot \hat{\boldsymbol{n}}_{\mathrm{B}} = 0. \tag{19.47}$$

Setting $\hat{\boldsymbol{n}} = \hat{\boldsymbol{n}}_{\mathrm{B}} = -\hat{\boldsymbol{n}}_{\mathrm{A}}$ leads to

$$[-p_{\mathrm{A}}\, \mathbb{I} + \mathbb{T}_{\mathrm{A}} + \rho_{\mathrm{A}}\, \boldsymbol{v}_{\mathrm{A}} \otimes (\boldsymbol{v}^{(b)} - \boldsymbol{v}_{\mathrm{A}})] \cdot \hat{\boldsymbol{n}} = [-p_{\mathrm{B}}\, \mathbb{I} + \mathbb{T}_{\mathrm{B}} + \rho_{\mathrm{B}}\, \boldsymbol{v}_{\mathrm{B}} \otimes (\boldsymbol{v}^{(b)} - \boldsymbol{v}_{\mathrm{B}})] \cdot \hat{\boldsymbol{n}}, \tag{19.48}$$

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 Chapter 44). Hence, setting $p_{\mathrm{A}} = p_{\mathrm{B}}$ allows us to cancel pressure thus leaving an interface stress condition involving just the frictional stress and kinetic stress

$$[\mathbb{T}_{\mathrm{A}} + \rho_{\mathrm{A}}\, \boldsymbol{v}_{\mathrm{A}} \otimes (\boldsymbol{v}^{(b)} - \boldsymbol{v}_{\mathrm{A}})] \cdot \hat{\boldsymbol{n}} = [\mathbb{T}_{\mathrm{B}} + \rho_{\mathrm{B}}\, \boldsymbol{v}_{\mathrm{B}} \otimes (\boldsymbol{v}^{(b)} - \boldsymbol{v}_{\mathrm{B}})] \cdot \hat{\boldsymbol{n}}. \tag{19.49}$$

We now consider a variety of examples to illustrate this condition.

19.8.2 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 14.4.2)

$$(\boldsymbol{v}_{\mathrm{A}} - \boldsymbol{v}^{(b)}) \cdot \hat{\boldsymbol{n}} = (\boldsymbol{v}_{\mathrm{B}} - \boldsymbol{v}^{(b)}) \cdot \hat{\boldsymbol{n}} = 0. \tag{19.50}$$

A nonzero $\boldsymbol{v}^{(b)}$ 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 $\boldsymbol{v}^{(b)} = 0$ for solid boundaries. Since there is no contribution from the kinetic stress, the stress condition (19.49) reduces to

$$\mathbb{T}_{\mathrm{A}} \cdot \hat{\boldsymbol{n}} = \mathbb{T}_{\mathrm{B}} \cdot \hat{\boldsymbol{n}}. \tag{19.51}$$

This identity is consistent with

$$\mathbb{T}_{\mathrm{A}} \cdot \hat{\boldsymbol{n}} = \boldsymbol{\tau}^{\text{friction A}}(\boldsymbol{x}, t, \hat{\boldsymbol{n}}) = -\boldsymbol{\tau}^{\text{friction A}}(\boldsymbol{x}, t, -\hat{\boldsymbol{n}}) = \mathbb{T}_{\mathrm{B}} \cdot \hat{\boldsymbol{n}} = \boldsymbol{\tau}^{\text{friction B}}(\boldsymbol{x}, t, \hat{\boldsymbol{n}}), \tag{19.52}$$

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.

Figure 19.6: *An infinitesimal pillbox 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 $\boldsymbol{v}^{(b)}$. We orient the interface through the outward normals according to $\hat{\boldsymbol{n}} = \hat{\boldsymbol{n}}_B = -\hat{\boldsymbol{n}}_A$. 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, $\hat{\boldsymbol{n}} \cdot \hat{\boldsymbol{z}} \neq 0$, 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 $\hat{\boldsymbol{n}} \cdot \hat{\boldsymbol{z}} \neq 0$, with this assumption based on convenience. The stress condition is general and so does not require this assumption.*

19.8.3 No-slip boundary condition

At solid boundaries, the kinematic boundary condition from Section 14.4.1 sets the normal component of the velocity to zero

$$\boldsymbol{v} \cdot \hat{\boldsymbol{n}} = 0 \qquad \text{kinematic no-flux condition on static material boundary.} \tag{19.53}$$

However, nowhere in the present discussion have we specified the tangential component of the velocity along a solid boundary. What is it?

Figure 19.7: *The no-slip boundary condition means that fluid exhibits zero relative motion at the solid-fluid boundary, $\boldsymbol{v} \cdot \hat{\boldsymbol{t}} = 0$. That is, the fluid sticks to the solid boundary. The kinematic no-normal flow boundary condition, $\boldsymbol{v} \cdot \hat{\boldsymbol{n}} = 0$, 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.*

Careful laboratory experiments over the 19th and 20th centuries indicate that there is no relative motion of molecules at solid-fluid interfaces. That is, a fluid at the fluid-solid interface has a velocity

matching that of the solid so that the fluid sticks to the solid boundary as depicted in Figure 19.7. The no-slip boundary condition means that both the normal and tangential components of the fluid velocity vanish next to static solid boundaries

$$\boldsymbol{v} \cdot \hat{\boldsymbol{n}} = \boldsymbol{v} \cdot \hat{\boldsymbol{t}} = 0 \qquad \text{no-slip condition on static solid boundaries.} \tag{19.54}$$

The no-slip boundary condition gives rise to an exchange of momentum between the solid and fluid, with this exchange mediated by friction. In the absence of friction, as per an inviscid 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.

19.8.4 Lagrangian interface

Consider a Lagrangian interface within the fluid, with this interface defined so that

$$(\boldsymbol{v}_{\mathrm{A}} - \boldsymbol{v}^{(b)}) \cdot \hat{\boldsymbol{n}} = (\boldsymbol{v}_{\mathrm{B}} - \boldsymbol{v}^{(b)}) \cdot \hat{\boldsymbol{n}} = 0. \tag{19.55}$$

Again the kinetic stress contribution to equation (19.49) vanishes. We thus have the frictional stress condition (19.51) and a Newton's third law interpretation (19.52), yet now the frictional transfer takes place between two regions of the same fluid.

19.8.5 Permeable interface

Now allow for the interface to be permeable and introduce the dia-surface transport defined by equation (17.40)

$$(\boldsymbol{v}_{\mathrm{A}} - \boldsymbol{v}^{(b)}) \cdot \hat{\boldsymbol{n}} = (\boldsymbol{v}_{\mathrm{B}} - \boldsymbol{v}^{(b)}) \cdot \hat{\boldsymbol{n}} = w^{\mathrm{dia}}, \tag{19.56}$$

where we assumed $\boldsymbol{v}_{\mathrm{A}} = \boldsymbol{v}_{\mathrm{B}}$ at the interface. The kinetic stress thus adds to the frictional contribution, with the kinetic stress representing the transfer of momentum across the interface through the dia-surface advection of fluid elements

$$\mathbb{T}_{\mathrm{A}} \cdot \hat{\boldsymbol{n}} + \rho_{\mathrm{A}}\, \boldsymbol{v}_{\mathrm{A}}\, w^{\mathrm{dia}} = \mathbb{T}_{\mathrm{B}} \cdot \hat{\boldsymbol{n}} + \rho_{\mathrm{B}}\, \boldsymbol{v}_{\mathrm{B}}\, w^{\mathrm{dia}}. \tag{19.57}$$

We consider two cases.

Single continuous fluid media

If the interface is within a single continuous fluid media, then $\rho_{\mathrm{A}} = \rho_{\mathrm{B}}$ and $\boldsymbol{v}_{\mathrm{A}} = \boldsymbol{v}_{\mathrm{B}}$, 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 (19.51) 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 (14.60)

$$\rho_{\mathrm{A}}\, w^{\mathrm{dia}} = \rho_{\mathrm{B}}\, w^{\mathrm{dia}} = -\mathcal{Q}_{\mathrm{m}}, \tag{19.58}$$

where $\mathcal{Q}_{\mathrm{m}}$ is the mass per time per surface area crossing the boundary. The minus sign is implied by the convention that $\mathcal{Q}_{\mathrm{m}} > 0$ 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 (19.57) takes the form

$$\mathbb{T}_{\text{atm}} \cdot \hat{\boldsymbol{n}} - \boldsymbol{v}_{\text{atm}} \mathcal{Q}_{\text{m}} = \mathbb{T}_{\text{ocn}} \cdot \hat{\boldsymbol{n}} - \boldsymbol{v}_{\text{ocn}} \mathcal{Q}_{\text{m}}. \tag{19.59}$$

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 (19.56) we assumed that $\boldsymbol{v}_{\text{A}} = \boldsymbol{v}_{\text{B}}$ at the interface. That is, the linear momentum transferred from the atmosphere to the ocean is equal to that received by the ocean, and vice-versa. Consequently, the kinetic stress again balances across the boundary so that we are left with the friction boundary condition (19.51).

19.8.6 Comments

There are three terms in the general expression for the stress boundary condition (19.48), with contributions from pressure, friction, and kinetic stresses. 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. Furthermore, assuming a continuous velocity along with mass conservation, we find a continuous kinetic stress at the interface. By inference, we are led to a continuous frictional stress. We conclude that each of the three stresses separately match on the two sides of the interface. We thus find no coupling between the three contributions to the boundary stress. We might have guessed that there is a separate balance of the three stress terms, given the physically distinct properties of the three stresses. Even so, it is useful to see how this separate balance can be deduced by an analysis of the boundary conditions and the underlying physical principles.

19.9 Exercises

EXERCISE 19.1: STEADY STATE OCEAN FORCE BALANCE

Consider a hydrostatic fluid in an ocean basin, $\mathcal{R}$. The bottom interface separates the ocean from the solid-earth, and the upper interface separates the ocean from the atmosphere with a nonzero mass. Apply a horizontal stress over the ocean surface with a stress vector $\boldsymbol{\tau}^{\text{surf}}$, and allow for the ocean bottom to exchange momentum with the solid-earth through a horizontal bottom stress, $\boldsymbol{\tau}^{\text{bott}}$.

(a) What is the force balance for the full ocean domain at steady state? Express this balance in words and in equations. The answer should be generally stated; no need for specific details.

(b) What is the vertical component of the force balance, again computed over the full ocean domain? Hint: recall the fluid is assumed to be in hydrostatic balance.

(c) Assume an f-plane (i.e., the Coriolis parameter f is a constant; see Section 24.2) and that the center of mass motion vanishes, $\int \rho \, \boldsymbol{v} \, \mathrm{d}V = 0$. Discuss the zonal and meridional steady state force balance.

20

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. It is worth emphasizing this point; namely, form stresses act on sloped surfaces and as such have dimensions of force per area just as all stresses. As for any contact force, Newton's Third law describes how form stress renders a transfer of pressure forces across interfaces. Pressure form stresses thus affect a vertical transfer of horizontal forces and, in turn, provide an inviscid/reversible mechanism for the vertical transfer of horizontal momentum. In this chapter we develop an understanding of how form stresses act on the variety of interfaces encountered in geophysical fluids. This understanding has great value in many contexts, such as understanding how pressure acts between layers of fluid (*interfacial form stress*); how it acts between a fluid and a solid boundary (*topographic form stress*); and how it acts between two fluids such as the atmosphere and ocean (*atmospheric form stress* and *oceanic form stress*).

READER'S GUIDE TO THIS CHAPTER

This chapter builds from our study of stresses in Chapter 19. An understanding of pressure form stress greatly enhances our understanding of how geophysical fluids are forced. We will thus encounter form stresses in many contexts within subsequent chapters of this book.

20.1 Pressure form stresses at an interface **336**
20.1.1 Concerning the sign of a form stress 336
20.1.2 Mathematical expression for form stress 337
20.1.3 Form stresses in exact hydrostatic equilibrium 338
20.1.4 Comments 339
20.2 Topographic and atmospheric form stresses on the ocean **340**
20.2.1 Zonally symmetric ridge example 340
20.2.2 Form stress transfer between the fluid and its boundaries 341
20.3 Interfacial form stress **342**
20.3.1 Interfacial form stresses transferred between layers 343
20.3.2 Zonally integrated interfacial form stress 344
20.3.3 Comments 345
20.4 Axial angular momentum budget **345**
20.4.1 Anticipating the budget 346
20.4.2 Axial angular momentum 346
20.4.3 Depth integrated budget 347

20.4.4 Atmospheric and topographic form stresses 347
20.4.5 Frictional stresses at the surface and bottom 348
20.4.6 Summary budget for column integrated axial angular momentum . . 348
20.4.7 Zonal domain topology/geometry and zonal integration 349
20.4.8 Steady zonal and depth integrated budget 350
20.4.9 Remarks on a modified expression for the form stress 351
20.4.10 Southern Ocean balances 351
20.4.11 Further study 352

20.1 Pressure form stresses at an interface

As depicted in Figure 20.1, there are three surfaces or interfaces across which we commonly encounter 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 material surfaces in a perfect fluid, and more generally their slopes are directly connected to the geostrophic motion studied in Chapter 26. We study form stresses associated with buoyancy interfaces found in geostrophic flows as part of Section 26.5.

- 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 when reading the literature if it is unclear who is the giver of the form stress and who is the receiver. To help solidfy our understanding of 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 20.1: *Illustrating the three interfaces of concern in geophysical fluid mechancs for the discussion of form stresses. Left panel: a curved atmosphere-ocean interface leads to an atmospheric form stress acting on the ocean,* $\boldsymbol{F}^{form}_{atm\ to\ ocn}$*, and its equal and opposite oceanic form stress acting on the atmosphere,* $\boldsymbol{F}^{form}_{ocn\ to\ atm} = -\boldsymbol{F}^{form}_{atm\ to\ ocn}$*. Middle panel: a curved interior ocean interface (e.g., a buoyancy surface) leads to an interfacial form stress acting on the lower layer,* $\boldsymbol{F}^{form}_{k\to k+1}$*, and its equal and opposite interfacial form stress acting on the upper layer,* $\boldsymbol{F}^{form}_{k+1\to k} = -\boldsymbol{F}^{form}_{k\to k+1}$*. Right panel: a curved fluid solid earth interface leads to a fluid form stress acting on the solid earth,* $\boldsymbol{F}^{form}_{fluid\ to\ earth}$*, and its equal and opposite oceanic form stress acting on the fluid,* $\boldsymbol{F}^{form}_{earth\ to\ fluid} = -\boldsymbol{F}^{form}_{fluid\ to\ earth}$*. 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.*

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,* $\boldsymbol{F}^{A\ onto\ B}$*, is equal and opposite to the force imparted by region B onto region A so that* $\boldsymbol{F}^{A\ onto\ B} = -\boldsymbol{F}^{B\ onto\ A}$*. In the case where A=fluid (ocean or atmosphere) and B=solid earth, we generally to refer to the horizontal portion of* $\boldsymbol{F}^{earth\ onto\ fluid}$*, 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 as[1]

$$z = \eta(x, y, t). \tag{20.1}$$

The outward normal pointing away from the top side of the surface is given by

$$\hat{\boldsymbol{n}}_{\text{top}} = \frac{\nabla(z-\eta)}{|\nabla(z-\eta)|} = \frac{\hat{\boldsymbol{z}} - \nabla\eta}{\sqrt{1+|\nabla\eta|^2}}. \tag{20.2}$$

Multiplying the pressure times the horizontal area element on the surface, $\mathrm{d}\mathcal{S}$, leads to the net

[1]For more on the maths of such surfaces, see the geometry discussion in Chapter 4. See also the discussion of generalized vertical coordinates in Chapters 8 and 17.

pressure force acting at a point on the top side of the surface

$$\boldsymbol{F}^{\text{press}} = -p\,\hat{\boldsymbol{n}}_{\text{top}}\,\mathrm{d}\mathcal{S} = -p\,(\hat{\boldsymbol{z}} - \nabla\eta)\,\mathrm{d}A = -p\,(-\partial_x\eta\,\hat{\boldsymbol{x}} - \partial_y\eta\,\hat{\boldsymbol{y}} + \hat{\boldsymbol{z}})\,\mathrm{d}A. \tag{20.3}$$

In this equation we used the identity[2]

$$\mathrm{d}\mathcal{S} = |\nabla(z-\eta)|\,\mathrm{d}A, \tag{20.4}$$

with

$$\mathrm{d}A = \mathrm{d}x\,\mathrm{d}y \tag{20.5}$$

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

$$\text{pressure form stress acting on top side of interface} \equiv p\,\nabla\eta. \tag{20.6}$$

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

$$\boldsymbol{F}^{\text{press}}_{\text{top}} = \hat{\boldsymbol{z}}\cdot\boldsymbol{F}^{\text{press}}_{\text{top}} + \boldsymbol{F}^{\text{form}}_{\text{top}} = p\,(-\hat{\boldsymbol{z}} + \nabla\eta)\,\mathrm{d}A \qquad \text{pressure force on top of interface.} \tag{20.7}$$

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 19.8, 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 19.6.2)

$$\boldsymbol{F}^{\text{press}}_{\text{bot}} = \hat{\boldsymbol{z}}\cdot\boldsymbol{F}^{\text{press}}_{\text{bot}} + \boldsymbol{F}^{\text{form}}_{\text{bot}} = p\,(+\hat{\boldsymbol{z}} - \nabla\eta)\,\mathrm{d}A \qquad \text{pressure force on bottom of interface.} \tag{20.8}$$

20.1.3 Form stresses in exact hydrostatic equilibrium

As discussed in Section 18.3, static equilibrium of a fluid in a gravitational field results in a hydrostatic pressure field satisfying

$$\nabla p = -\rho\,\nabla\Phi, \tag{20.9}$$

where Φ is the geopotential. All other stresses vanish for the static fluid. With $\Phi = g\,z$, hydrostatic balance is realized for any finite region in the fluid, $\mathcal{R}$, means that

$$\oint_{\partial\mathcal{R}} \hat{\boldsymbol{z}}\cdot\hat{\boldsymbol{n}}\,p\,\mathrm{d}\mathcal{S} = -\int_{\mathcal{R}} \rho\,g\,\mathrm{d}V \tag{20.10a}$$

$$\oint_{\partial\mathcal{R}} \hat{\boldsymbol{x}}\cdot\hat{\boldsymbol{n}}\,p\,\mathrm{d}\mathcal{S} = 0 \tag{20.10b}$$

$$\oint_{\partial\mathcal{R}} \hat{\boldsymbol{y}}\cdot\hat{\boldsymbol{n}}\,p\,\mathrm{d}\mathcal{S} = 0. \tag{20.10c}$$

Equation (20.10a) says the vertical pressure force acting on a region is balanced by the weight of that region. Equations (20.10b) and (20.10c) say that the net form stress integrated over the region boundaries vanishes.

[2]The identity (20.4) follows from trigonometry summarized in Figure 20.3. See further details in the kinematic boundary conditions of Section 14.4 and the analogous dia-surface transport in Section 17.3.

Figure 20.3: *The pressure force acting on an infinitesimal portion of an arbitrary surface is given by* $\boldsymbol{F}^{\text{press}} = -p\,\hat{\boldsymbol{n}}\,\mathrm{d}\mathcal{S}$*, where* $\mathrm{d}\mathcal{S}$ *is the surface area of the tiny region. We here depict the pressure acting on the top side of a surface,* $\boldsymbol{F}^{\text{press}}_{top}$*. 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:* $\boldsymbol{F}^{\text{press}}_{top} = -\boldsymbol{F}^{\text{press}}_{bot}$*. The horizontal component of this force vector arises from the slope or the geometric form of the surface. We refer to the horizontal pressure force per area as the form stress,* $\boldsymbol{F}^{\text{form}}_{top} = -\boldsymbol{F}^{\text{form}}_{bot}$*. The area element on the surface,* $\mathrm{d}\mathcal{S}$*, has a horizontal projection given by* $\mathrm{d}A = \mathrm{d}x\,\mathrm{d}y = \cos\vartheta\,\mathrm{d}\mathcal{S}$*, with the angle assumed to be within the range* $-\pi/2 < \vartheta < \pi/2$ *so that the surface is nowhere vertical.*

Equations (20.10a)–(20.10c) offer an expression of static force balance over a finite region in a hydrostatic fluid. They are integral expressions rather than point-wise expressions. Therefore, as illustrated by Figure 20.4, the force acting at one depth across the region need not be in exact balance. Correspondingly, the form stresses acting over the body need not balance at any particular depth. This imbalance might lead one to incorrectly conclude that there is a net horizontal acceleration merely due to the shape of the region. That conclusion is wrong since the force balance for a finite region is computed as an area integral over the full boundary of the region, not just at a single point or depth along the boundary. This is a rather trivial conclusion, and yet it is one that can be readily missed when enmeshed in a detailed analysis, such as that considered in the following sections.

20.1.4 Comments

Form stress can appear mysterious when encountered in the geophysical fluid mechanics literature. Part of the reason is that it often appears "out of the blue" as part of mathematical manipulations of the momentum equation. We illustrate these manipulations in Sections 20.2, 20.3, and 20.4. Another reason for the mystery is that the signs are often not clearly specified, as per the discussion in Section 20.1.1. From the first principles perspective presented here, form stress is simply the horizontal projection of the pressure stress acting along a sloping surface. Keeping this physical picture in mind goes a long way to help in understanding form stress and reading the literature. Besides our further discussions of form stress in this chapter, we further encounter form stress when studying buoyancy/isopycnal form stresses for geostrophic motions in Section 26.5, and for shallow water fluids in Section 30.2.

Figure 20.4: *Here we depict a right triangle submerged in a hydrostatic ocean along with bottom topography. In exact hydrostatic balance, the area integrated pressure force acting over the triangle boundary vanishes,* $-\oint_{\partial\mathcal{R}} p\,\hat{\boldsymbol{n}}\,\mathrm{d}\mathcal{S} = 0$. *However, at any particular depth, the force acting on the triangle will not generally be zero. For example, consider the forces acting at the depth of the horizontal black line. On the vertical left side, the form stress equals to the pressure force,* $\boldsymbol{F}^{form}_{vertical} = \boldsymbol{F}^{press}$. *At the same depth on the sloped side, the pressure force has the same magnitude since the pressure is exactly hydrostatic and so is only a function of* z. *However, the form stress acting on the sloped side has a magnitude smaller than the pressure force,* $|\boldsymbol{F}^{form}_{sloped}| = |\boldsymbol{F}^{press}|\,\sin\theta < |\boldsymbol{F}^{press}|$. *Even though the form stresses are not in balance across this particular depth, there is no net horizontal acceleration of the triangle since it is the net area integrated force that determines the acceleration of any region, not just the force at any particular depth.*

20.2 Topographic and atmospheric form stresses on the ocean

In this section we focus on the form stress associated with the shape of the solid earth boundary 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 example

Figure 20.5 provides an idealized ridge configuration 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 20.4, 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 20.4.10 when studying the force balances for steady circulation in a zonally periodic channel.

Figure 20.5: *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 20.9 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.*

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 $z = -H(x, y)$ to the free surface at $z = \eta(x, y, t)$, and focus on the zonal force balance such as depicted in Figure 20.5. In computing the acceleration acting on this column at a particular horizontal position, we need to determine the depth integrated zonal pressure gradient

$$\text{depth integrated zonal pressure gradient} = -\int_{-H}^{\eta} \frac{\partial p}{\partial x}\, \mathrm{d}z. \tag{20.11}$$

We expose the contact force version of the pressure force by making use of Leibniz's Rule (e.g., Section 14.5) to write

$$-\int_{-H}^{\eta} \frac{\partial p}{\partial x}\, \mathrm{d}z = \underbrace{-\frac{\partial}{\partial x}\int_{-H}^{\eta} p\, \mathrm{d}z}_{\text{zonal deriv depth integrated pressure}} + \underbrace{\frac{\partial \eta}{\partial x}\, p_{\text{atm}}}_{\text{atmospheric form stress}} + \underbrace{\frac{\partial H}{\partial x}\, p_{\text{bot}}}_{\text{topographic form stress}}, \tag{20.12}$$

where we introduced the atmospheric and bottom pressures, p_{atm} and p_{bot}, 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

$$\text{zonal derivative of layer integrated pressure} = -\frac{\partial}{\partial x}\int_{-H}^{\eta} p\,\mathrm{d}z. \tag{20.13}$$

  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_{atm}, imparts an atmospheric form stress onto the ocean

$$\text{zonal atmospheric form stress acting on ocean} = \frac{\partial \eta}{\partial x}\, p_{\text{atm}}. \tag{20.14}$$

  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 pressure, p_{bot} present at $z = -H$ imparts an oceanic form stress to the solid earth

$$\text{zonal oceanic form stress acting on solid earth} = -\frac{\partial H}{\partial x}\, p_{\text{bot}}. \tag{20.15}$$

  In turn, through Newton's Third Law, the topographic form stress acting on the ocean is equal in magnitude but oppositely directed

$$\text{zonal topographic form stress acting on ocean} = +\frac{\partial H}{\partial x}\, p_{\text{bot}}. \tag{20.16}$$

  For example, if the bottom rises to the east, so that $\partial H/\partial x < 0$, 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 sanity check, we verify that the signs of these form stresses are consistent with those in Figure 20.5.

20.3 Interfacial form stress

In this section we focus on the form stress associated with the form of 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 20.6. These manipulations are analogous to those considered in Section 20.2 for the topographic and atmospheric form stresses.

Figure 20.6: *Fluid layers for illustrating interfacial form stress. The layer of focus is labelled k, with the layer above labelled $k-1$ and layer below labeled $k+1$. The layer bounds are specified by vertical positions $z = \eta_{k-1/2}(x, y, t)$ and $z = \eta_{k+1/2}(x, y, t)$. It is common to define the layers according to buoyancy (see Section 26.5), given its direct connection to pressure and dynamics, and furthermore given that buoyancy surfaces are material in perfect fluids. Even so, the ideas of interfacial form stress are generic and thus hold for layers defined arbitrarily. In formulating the layer momentum budget at a fixed horizontal position, we integrate the zonal pressure gradient over a column living within the layer. According to equation (20.18), this integrated zonal pressure force can be decomposed into interfacial form stresses (IFS) acting on the upper and lower layer interfaces, plus the zonal gradient of the pressure acting along vertical sides of the column. 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.*

20.3.1 Interfacial form stresses transferred between layers

When studying the momentum of a column of fluid within a chosen layer, we will need to compute the depth integrated zonal pressure gradient over a layer at a particular horizontal point

$$\text{layer integrated zonal pressure gradient} = -\int_{\eta_{k+1/2}}^{\eta_{k-1/2}} \frac{\partial p}{\partial x}\, \mathrm{d}z, \tag{20.17}$$

where $z = \eta_{k-1/2}(x, y, t)$ is the vertical position for the interface at the top of the fluid layer, $z = \eta_{k+1/2}(x, y, t)$ 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 studying topographic and atmospheric form stresses, we make use of Leibniz' Rule for a fluid layer

$$-\int_{\eta_{k+1/2}}^{\eta_{k-1/2}} \frac{\partial p}{\partial x}\, \mathrm{d}z = \underbrace{-\frac{\partial}{\partial x}\int_{\eta_{k+1/2}}^{\eta_{k-1/2}} p\, \mathrm{d}z}_{\text{zonal deriv layer integrated pressure}} + \underbrace{\frac{\partial \eta_{k-1/2}}{\partial x}\, p_{k-1/2}}_{\text{IFS at k-1/2 interface}} - \underbrace{\frac{\partial \eta_{k+1/2}}{\partial x}\, p_{k+1/2}}_{\text{IFS at k+1/2 interface}}, \tag{20.18}$$

where we introduced the pressures acting at a point on the interfaces

$$p_{k-1/2} = p(x, y, z = \eta_{k-1/2}, t) \qquad \text{and} \qquad p_{k+1/2} = p(x, y, z = \eta_{k+1/2}, t). \tag{20.19}$$

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

$$\text{zonal derivative of layer integrated pressure} = -\frac{\partial}{\partial x} \int_{\eta_{k+1/2}}^{\eta_{k-1/2}} p \, \mathrm{d}z. \tag{20.20}$$

- 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:

$$\text{IFS on layer-}k \text{ from the } \eta_{k-1/2} \text{ interface} = \frac{\partial \eta_{k-1/2}}{\partial x} \, p_{k-1/2}. \tag{20.21}$$

  For example, if the upper layer interface slopes up to the east, $\partial \eta_{k-1/2}/\partial x > 0$, then the IFS provides a positive (eastward) zonal force to layer k. In turn, through Newton's Third law, the layer above, labelled $k-1$, feels an IFS 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

$$\text{IFS on layer-}k \text{ from } \eta_{k+1/2} \text{ interface} = -\frac{\partial \eta_{k+1/2}}{\partial x} \, p_{k+1/2}. \tag{20.22}$$

  For example, if the layer slopes down to the east, $\partial \eta_{k+1/2}/\partial x < 0$, then the IFS accelerates layer-k to the east. In turn, through Newton's Third law, the IFS acts to accelerate the layer below, labelled $k+1$, to the west.

Extending the above to include the full ocean column, starting from the free upper surface to the rigid solid earth bottom, all the intermediate interfacial form stresses vanish in the depth integral due to Newton's Third law. Hence, accumulation of the interfacial form stresses throughout the column leaves only the IFS at the top and that 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

$$-\int \left[\int_{\eta_{k+1/2}}^{\eta_{k-1/2}} \frac{\partial p}{\partial x} \, \mathrm{d}z \right] \mathrm{d}x = \int \left[-\frac{\partial}{\partial x} \int_{\eta_{k+1/2}}^{\eta_{k-1/2}} p \, \mathrm{d}z + \frac{\partial \eta_{k-1/2}}{\partial x} \, p_{k-1/2} - \frac{\partial \eta_{k+1/2}}{\partial x} \, p_{k+1/2} \right] \mathrm{d}x. \tag{20.23}$$

If the domain is zonally periodic or is bounded by sloping shorelines (see Figure 20.8 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

$$-\int \left[\int_{\eta_{k+1/2}}^{\eta_{k-1/2}} \frac{\partial p}{\partial x} \, \mathrm{d}z \right] \mathrm{d}x = \int \left[\frac{\partial \eta_{k-1/2}}{\partial x} \, p_{k-1/2} - \frac{\partial \eta_{k+1/2}}{\partial x} \, p_{k+1/2} \right] \mathrm{d}x. \tag{20.24}$$

This zonal integral is only affected by zonal anomalies for the layer vertical positions and pressures

$$-\int \left[\int_{\eta_{k+1/2}}^{\eta_{k-1/2}} \frac{\partial p}{\partial x} \, \mathrm{d}z \right] \mathrm{d}x = \int \left[\frac{\partial \eta'_{k-1/2}}{\partial x} \, p'_{k-1/2} - \frac{\partial \eta'_{k+1/2}}{\partial x} \, p'_{k+1/2} \right] \mathrm{d}x \tag{20.25a}$$

$$= \int \left[-\eta'_{k-1/2} \frac{\partial p'_{k-1/2}}{\partial x} + \eta'_{k+1/2} \frac{\partial p'_{k+1/2}}{\partial x} \right] \mathrm{d}x, \tag{20.25b}$$

where primes denote deviations from the zonal mean. We spend some time in Section 20.4.8 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 is a notion that holds for an arbitrary surface drawn in a fluid. Interfaces defined by buoyancy isosurfaces make the connection between the general notions presented here to geostrophic mechanics given the connection between buoyancy slopes and thermal wind. Most studies of interfacial form stress are thus concerned with isopycnal interfacial form stress, and we study these stresses in Section 26.5, i

20.4 Axial angular momentum budget

We here develop the column integrated budget for axial angular momentum over a finite region of the ocean, such as shown in Figure 20.7. 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. For some generality we make use of spherical coordinates, though doing so offers only a modest degree of extra details beyond Cartesian. 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 detail for the connection provided in our study of geophysical particle dynamics in Sections 10.6 and 10.7. 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.

Figure 20.7: *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. Under certain circumstances, such as for the large-scale Southern Ocean circulation, the steady state balance becomes quite simple, as indicated in Figure 20.9.*

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 rendered a physically sensible budget.

For this purpose, consider a column of fluid such as shown in Figure 20.7. 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 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. In so doing, they modify the axial angular momentum of the fluid column. In the following development, we mathematically express the variety of forces and corresponding torques. A prominant ocean example where the axial angular momentim budget proves of value is in studying dynamical balances in the Southern Ocean. We briefly consider that example at the end of the section.

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

$$\rho \frac{\mathrm{D} l^z}{\mathrm{D} t} = -\frac{\partial p}{\partial \lambda} + r_\perp \, \rho \, F^\lambda, \tag{20.26}$$

where

$$l^z = r_\perp \, (u + r_\perp \, \Omega) \tag{20.27}$$

is the axial angular momentum per unit mass, and

$$r_\perp = r \cos \phi \tag{20.28}$$

is the distance to the polar rotation axis. Use of the Eulerian form of mass conservation (equation (14.9)) leads to the Eulerian flux-form budget

$$\frac{\partial(\rho\, l^z)}{\partial t} + \nabla\cdot(\rho\, \boldsymbol{v}\, l^z) = -\frac{\partial p}{\partial \lambda} + r_\perp\, \rho\, F^\lambda, \tag{20.29}$$

with $l^z \rho\, \mathrm{d}z$ 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 the budget (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

$$\int_{-H}^{\eta} \frac{\partial(\rho\, l^z)}{\partial t}\, \mathrm{d}z = \frac{\partial}{\partial t}\left[\int_{-H}^{\eta} \rho\, l^z \mathrm{d}z\right] - \left[\rho\, l^z\, \frac{\partial \eta}{\partial t}\right]_{z=\eta} \tag{20.30a}$$

$$\int_{-H}^{\eta} \nabla_z\cdot(\rho\, \boldsymbol{u}\, l^z)\mathrm{d}z = \nabla_z\cdot\left[\int_{-H}^{\eta} \rho\, \boldsymbol{u}\, l^z \mathrm{d}z\right] - [\rho\, l^z \boldsymbol{u}\cdot\nabla\eta]_{z=\eta} - [\rho\, l^z \boldsymbol{u}\cdot\nabla H]_{z=-H} \tag{20.30b}$$

$$\int_{-H}^{\eta} \frac{\partial(\rho\, w\, l^z)}{\partial z}\, \mathrm{d}z = [w\, \rho\, l^z]_{z=\eta} - [w\, \rho\, l^z]_{z=-H}\,. \tag{20.30c}$$

The surface kinematic boundary condition (14.73) and bottom kinematic boundary condition (14.38) allow us to reach a reasonably tidy expression

$$\frac{\partial}{\partial t}\left[\int_{-H}^{\eta} l^z \rho\, \mathrm{d}z\right] = -\nabla_z\cdot\left[\int_{-H}^{\eta} l^z \boldsymbol{u}\, \rho\, \mathrm{d}z\right] + [l^z Q_{\mathrm{m}}]_{z=\eta} + \int_{-H}^{\eta}\left[-\frac{\partial p}{\partial \lambda} + r_\perp\, \rho\, F^\lambda\right]\mathrm{d}z. \tag{20.31}$$

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

$$\int_{-H}^{\eta} \frac{\partial p}{\partial \lambda}\, \mathrm{d}z = \frac{\partial P}{\partial \lambda} - p_{\mathrm{a}}\, \frac{\partial \eta}{\partial \lambda} - p_{\mathrm{b}}\, \frac{\partial H}{\partial \lambda}, \tag{20.32a}$$

where

$$P = \int_{-H}^{\eta} p\, \mathrm{d}z \tag{20.33}$$

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, $z = -H$. 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 Frictional stresses at the surface and bottom

For friction, we are uninterested in the contribution from viscous friction acting on horizontal strains. Instead, we are interested in the vertical transfer of zonal momentum arising from the vertical shear of horizontal stresses

$$\rho\, F^{\lambda} = \frac{\partial \tau^{\lambda}}{\partial z}, \tag{20.34}$$

where τ^{λ} is the zonal component to the stress vector. When integrated vertically over an ocean column, this friction reduces to 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

$$r_{\perp} = r\cos\phi = (z+R)\cos\phi \approx R\cos\phi = R_{\perp}, \tag{20.35}$$

with this assumption holding for the shallow fluid approximation built into the hydrostatic primitive equations discussed in Section 24.1. This assumption allows us to write the frictional contribution to the angular momentum budget (20.31) in the form

$$\int_{-H}^{\eta} r_{\perp}\, \rho\, F^{\lambda}\, \mathrm{d}z \approx R_{\perp} \int_{-H}^{\eta} \rho\, F^{\lambda}\, \mathrm{d}z = R_{\perp}\, (\tau_{\mathrm{a}}^{\lambda} - \tau_{\mathrm{b}}^{\lambda}). \tag{20.36}$$

The final expression introduced $\tau_{\mathrm{a}}^{\lambda}$, 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 $\tau_{\mathrm{a}}^{\lambda} > 0$ transfers an eastward momentum to the ocean such as via a westerly wind stress. Likewise, the stress $\tau_{\mathrm{b}}^{\lambda}$ is the zonal stress at the ocean bottom imparted through irreversible interactions between the ocean and the solid-earth. The signs are such that $\tau_{\mathrm{b}}^{\lambda} > 0$ 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

$$\frac{\partial}{\partial t}\left[\int_{-H}^{\eta} l^{z} \rho\, \mathrm{d}z\right] = -\nabla_{z}\cdot\left[\int_{-H}^{\eta} l^{z}\boldsymbol{u}\,\rho\,\mathrm{d}z\right] - \frac{\partial P}{\partial \lambda} + [l^{z} Q_{\mathrm{m}}]_{z=\eta} + p_{\mathrm{a}}\frac{\partial \eta}{\partial \lambda} + p_{\mathrm{b}}\frac{\partial H}{\partial \lambda} + R_{\perp}\,(\tau_{\mathrm{a}}^{\lambda} - \tau_{\mathrm{b}}^{\lambda}). \tag{20.37}$$

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.

Removing zonal means

We further isolate the processes contributing to the budget (20.37) by introducing the zonal mean operator

$$\overline{A} \equiv \frac{1}{L(\phi)} \int A\, \mathrm{d}\lambda, \tag{20.38}$$

where

$$L(\phi) = R\,\cos\phi\,\Delta\lambda = R_{\perp}\,\Delta\lambda \tag{20.39}$$

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 20.8. The corresponding zonal anomalies to the depth integrated pressure, sea surface height, and bottom topography are thus given by

$$P' = P - \overline{P} \qquad \text{and} \qquad \eta' = \eta - \overline{\eta} \qquad \text{and} \qquad H' = H - \overline{H}, \tag{20.40}$$

in which case equation (20.37) takes the form

$$\frac{\partial}{\partial t}\left[\int_{-H}^{\eta} l^z \rho \, \mathrm{d}z\right] = -\nabla_z \cdot \left[\int_{-H}^{\eta} l^z \boldsymbol{u} \, \rho \, \mathrm{d}z\right] - \frac{\partial P'}{\partial \lambda} + [l^z Q_{\mathrm{m}}]_{z=\eta} + p_{\mathrm{a}} \frac{\partial \eta'}{\partial \lambda} + p_{\mathrm{b}} \frac{\partial H'}{\partial \lambda} + R_{\perp}\,(\tau_{\mathrm{a}}^{\lambda} - \tau_{\mathrm{b}}^{\lambda}). \tag{20.41}$$

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

$$\nabla_z \cdot \left[\int_{-H}^{\eta} l^z \boldsymbol{u} \, \rho \, \mathrm{d}z\right] = -\frac{\partial P'}{\partial \lambda} + [l^z Q_{\mathrm{m}}]_{z=\eta} + p_{\mathrm{a}} \frac{\partial \eta'}{\partial \lambda} + p_{\mathrm{b}} \frac{\partial H'}{\partial \lambda} + R_{\perp}\,(\tau_{\mathrm{a}}^{\lambda} - \tau_{\mathrm{b}}^{\lambda}). \tag{20.42}$$

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

Figure 20.8: *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 along 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.*

Zonal integration and averaging is 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. 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 domains, it can be of interest to zonally integrate across the domain in order to emphasize zonal boundary effects.

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 20.8. For the zonally periodic domain, the zonal integral of any zonal derivative vanishes

$$\oint \frac{\partial P}{\partial \lambda}\, \mathrm{d}\lambda = 0, \tag{20.43}$$

where we write $\oint$ for integration over a zonally periodic domain. Additionally, zonal derivatives vanish for a zonally bounded domain with sloping shorelines

$$\int_{\text{shorelines}} \frac{\partial P}{\partial \lambda}\, \mathrm{d}\lambda = 0. \tag{20.44}$$

The reason this integral vanishes is that quanties such as P are depth integrated, and so $P = 0$ 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 20.8) 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, we mostly focus on the periodic or sloping shoreline configuration in the following.

20.4.8 Steady zonal and depth integrated budget

For a zonal domain that has a either periodic topology or sloping shoreline geometry (Figure 20.8), we are left with the zonal and depth integrated steady angular momentum budget

$$\frac{1}{R} \int \frac{\partial}{\partial \phi} \left[\int_{-H}^{\eta} l^z\, v\, \rho\, \mathrm{d}z \right] \mathrm{d}\lambda = \int \left[[l^z Q_{\mathrm{m}}]_{z=\eta} + p_{\mathrm{a}}' \frac{\partial \eta'}{\partial \lambda} + p_{\mathrm{b}}' \frac{\partial H'}{\partial \lambda} + R_{\perp} \left(\tau_{\mathrm{a}}^{\lambda} - \tau_{\mathrm{b}}^{\lambda} \right) \right] \mathrm{d}\lambda. \tag{20.45}$$

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. Furthermore, we introduced the zonal anomalies for the surface and bottom pressures

$$p_{\mathrm{a}}' = p_{\mathrm{a}} - \overline{p_{\mathrm{a}}} \qquad \text{and} \qquad p_{\mathrm{b}}' = p_{\mathrm{b}} - \overline{p_{\mathrm{b}}}. \tag{20.46}$$

We can introduce these anomalous fields since their zonal means do not contribute to either the periodic or sloping shoreline domains. To verify this property, note that

$$\int \overline{p_{\mathrm{a}}} \frac{\partial \eta'}{\partial \lambda}\, \mathrm{d}\lambda = \int \frac{\partial (\overline{p_{\mathrm{a}}}\, \eta')}{\partial \lambda}\, \mathrm{d}\lambda = 0. \tag{20.47}$$

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

$$\int \overline{p_{\mathrm{b}}} \frac{\partial H'}{\partial \lambda}\, \mathrm{d}\lambda = \int \frac{\partial (\overline{p_{\mathrm{b}}}\, H')}{\partial \lambda}\, \mathrm{d}\lambda = 0, \tag{20.48}$$

which follows either by periodicity or since $H' = 0$ 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 momenum balance (20.45) 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 (20.45) in the alternative form

$$\int \left[p_{\mathrm{a}}' \, \frac{\partial \eta'}{\partial \lambda} + p_{\mathrm{b}}' \, \frac{\partial H'}{\partial \lambda} \right] \mathrm{d}\lambda = \int \left[-\frac{\partial p_{\mathrm{a}}'}{\partial \lambda} \, \eta' - \frac{\partial p_{\mathrm{b}}'}{\partial \lambda} \, H' \right] \mathrm{d}\lambda. \tag{20.49}$$

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. 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 in question.

- 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 Southern Ocean balances

Under certain cases the primary balance in equation (20.45) is between the form stress and friction stress, whereby

$$\oint \left[p_{\mathrm{a}}' \, \frac{\partial \eta'}{\partial \lambda} + R_{\perp} \, \tau_{\mathrm{a}}^{\lambda} \right] \mathrm{d}\lambda \approx \oint \left[-p_{\mathrm{b}}' \, \frac{\partial H'}{\partial \lambda} + R_{\perp} \, \tau_{\mathrm{b}}^{\lambda} \right] \mathrm{d}\lambda. \tag{20.50}$$

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

$$\oint \tau_{\mathrm{a}}^{\lambda} \, \mathrm{d}\lambda \approx - \oint p_{\mathrm{b}}' \, \frac{1}{R_{\perp}} \, \frac{\partial H'}{\partial \lambda} \, \mathrm{d}\lambda = \oint H' \, \frac{1}{R_{\perp}} \, \frac{\partial p_{\mathrm{b}}'}{\partial \lambda} \, \mathrm{d}\lambda, \tag{20.51}$$

with this balance exemplified in Figure 20.9. We now state in words what this balance means in the presence of a net eastward wind stress, $\oint \tau_{\mathrm{a}}^{\lambda} \, \mathrm{d}\lambda > 0$. 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, and 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 (20.51) reveals that for $H' < 0$ (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 20.9.

Figure 20.9: *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 30.6.*

20.4.11 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). 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 30.4 for the shallow water system.

21

Buoyancy

Buoyancy is the acceleration felt by a massive body within a fluid due to the gravitational acceleration of the body relative to the gravitational acceleration of the fluid displaced by the body. As shown in Section 21.2, this statement of *Archimedes' Principle* is a basic consequence of hydrostatics. A body's buoyancy is measured by its density relative to that of the displaced fluid. Although the concept of buoyancy has wide applications for marine engineering (e.g., design of ships), we are interested in the buoyancy of fluid itself. That is, the "body" of interest for geophysical fluid mechanics is a test fluid element, one that does not disturb the environment into which it is placed.

We are thus interested in the buoyancy of a fluid element relative to the fluid that surrounds the element. A fluid element that is less dense than its local environment rises, whereas it sinks if less dense. We are furthermore interested in how density is measured and how it changes, with Section 21.3 just the beginning of our investigations into how density is modified by physical processes. In Section 21.4 we examine the resistance to motion within a gravitating fluid as measured by the field of buoyancy and its spatial gradients. As part of that study, we conceive of *neutral directions* where infinitesimal displacements of a fluid element lead to no buoyancy forces; i.e., directions where the fluid element is neutrally buoyant.

READER'S GUIDE TO THIS CHAPTER

This chapter builds from our ideas of pressure in Chapters 18 and 19. Buoyancy is a feature of many of the subsequent chapters in this book given its central role in the motion of geophysical fluids.

21.1 Loose threads **354**
21.2 Buoyancy force and Archimedes' Principle **354**
21.2.1 Hydrostatic pressure force 355
21.2.2 Vertical buoyancy force 355
21.3 Mass density and its flavors **356**
21.3.1 Realistic fluids 356
21.3.2 Equation of state in terms of potential temperature 356
21.3.3 Equation of state in terms of Consevative Temperature 357
21.3.4 Infinitesimal density increments and material time changes 357
21.3.5 Potential density 357
21.3.6 Linear equation of state for the ocean 359
21.3.7 Further study 359
21.4 Buoyancy stratification **359**

21.4.1 Buoyancy as a field . . . 359
21.4.2 Physical ideas underlying neutral directions . . . 360
21.4.3 Comparing density under two forms of displacement . . . 361
21.4.4 Buoyancy frequency . . . 362
21.4.5 Buoyancy frequency and locally referenced potential density . . . 363
21.5 Gravitational stability of a dry ideal gas atmosphere . . . **363**
21.6 Neutral directions . . . **364**
21.6.1 Neutral directions and the neutrality condition . . . 365
21.6.2 Comments on neutral directions . . . 365
21.7 Neutral helicity . . . **365**
21.7.1 Mathematical preliminaries . . . 366
21.7.2 Helical nature of neutral displacements . . . 366
21.7.3 Comments and further study . . . 367
21.8 Exercises . . . **367**

21.1 Loose threads

- effective buoyancy as per Jeevanjee's work
- pressure forces implied by buoyant motion

21.2 Buoyancy force and Archimedes' Principle

Consider an arbitrary massive body or region contained within a fluid at rest, such as that shown in Figure 21.1. The body can be a rigid solid or perhaps a human swimmer. For our studies, we are most interested in arbitrary region of the fluid itself, such as an infinitesimal fluid element or a finite volume. In general, the massive body displaces a volume of fluid, with the mass of the displaced fluid given by the integral of the density over the region

$$m_{\text{fluid}} = \int_{\mathscr{R}} \rho \, \mathrm{d}V. \tag{21.1}$$

We now seek an understanding of the forces acting on the massive body within the fluid.

Figure 21.1: *An arbitrary massive body, $\mathscr{R}$, within a fluid experiences a gravitational force acting down and a buoyant force acting up. The body is arbitrary, such as a rigid solid, a swimming creature, a fluid element, or a finite region of fluid. When the fluid is at rest, 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 buoyancy force acting on the body is determined by the difference in weight between the body and the displaced fluid. A positive buoyancy force acts on a body less dense than the displaced fluid, and a negative buoyancy force acts on a body more dense.*

21.2.1 Hydrostatic pressure force

As we will discuss in Chapter 19, any surface (even an imaginary surface) contained in 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 surface of $\mathcal{R}$ leads to the pressure force acting on the region

$$\boldsymbol{F}_{\text{pressure}} = -\oint_{\partial\mathcal{R}} p\,\hat{\boldsymbol{n}}\,\mathrm{d}\mathcal{S}, \tag{21.2}$$

where p is the hydrostatic pressure, $\hat{\boldsymbol{n}}$ is the outward normal on the boundary, and $\mathrm{d}\mathcal{S}$ is the area element. Making use of Gauss's Law in the form of equation (2.63) leads to the equivalent expression in terms of the volume integral of the pressure gradient

$$\boldsymbol{F}_{\text{pressure}} = -\int_{\mathcal{R}} \nabla p\,\mathrm{d}V. \tag{21.3}$$

This expression holds regardless the composition of the massive body. However, it is most useful when the body is a region of fluid so that the pressure gradient can be computed within the volume.

Since the fluid is at rest, the pressure field is hydrostatic and only has a dependence on the vertical position within the fluid, $p = p(z)$. The hydrostatic pressure equals to the weight per area of fluid sitting above any point in the fluid so that its vertical derivative given by

$$\frac{\partial p}{\partial z} = -\rho\, g. \tag{21.4}$$

Hence, the force acting on the region is

$$\boldsymbol{F}_{\text{pressure}} = -\int_{\mathcal{R}} \nabla p\,\mathrm{d}V = -\int_{\mathcal{R}} \hat{\boldsymbol{z}}\,(\partial p/\partial z)\,\mathrm{d}V = \hat{\boldsymbol{z}}\,g\int_{\mathcal{R}} \rho\,\mathrm{d}V = \hat{\boldsymbol{z}}\,g\,\mathcal{M}_{\text{fluid}}, \tag{21.5}$$

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 body. This result is a mathematical expression of Archimedes' Principle.

Note that there are no net horizontal forces acting on the body. 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, thus leaving zero net horizontal acceleration.

21.2.2 Vertical buoyancy force

The vertical pressure force (21.6) acts to push the body vertically upward against gravity. It therefore appears to reduce 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.

We are thus motivated to define a *buoyancy force* acting on a massive body relative to the displaced fluid according to

$$\boldsymbol{F}_{\text{buoyancy}} = \hat{\boldsymbol{z}}\,g\int_{\mathcal{R}} (\rho_{\text{fluid}} - \rho_{\text{body}})\,\mathrm{d}V = \hat{\boldsymbol{z}}\,g\,(\mathcal{M}_{\text{fluid}} - \mathcal{M}_{\text{body}}). \tag{21.6}$$

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.

21.3 Mass density and its flavors

The density of a fluid element plays a central role in the study of buoyancy in geophysical fluids. More precisely, it is the density of the fluid element relative to that of its surroundings. As commonly referred to in geophysical fluid mechanics, the equation of state 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). This equation is called the *thermal equation of state* in the thermodynamics literature, and we will study it more formally in Chapter 22. We here discuss the equation of state as well as the related flavors of mass density used to study stratified fluids.

21.3.1 Realistic fluids

To allow for general situations of multi-component fluids, we write the *in situ* density as a general function

$$\rho = \rho(S, T, p). \tag{21.7}$$

We made use of the standard oceanographic notation where $S = 1000\, C$ is the salinity as defined by equation (22.36). This equation of state can be derived from one of the more general *fundamental equations of state* discussed in Section 22.4.

The thermal equation of state for fluid density is generally not as simple as that for an ideal gas (Section 22.8.1). Indeed, liquids such as seawater have rather complex empirical expressions obtained from statistical fits to data. Part of the complexity arises from the multi-component nature of seawater (salt plus freshwater) as well as the nontrivial inter-molecular forces. In contrast, the equation of state for moist air can be written much like that for an ideal gas, thus making the equation of state for air far less complex than for seawater.

The equation of state (21.7) is a function of the *in situ* temperature, T, the *in situ* pressure, p, and the *in situ* salinity, S (ocean) or humidity (atmosphere). The term *in situ* means that a property is measured locally at a point in the fluid, with the resulting density also that measured at that point. Such *in situ* properties contrast to *potential* properties, which are based on referencing to a chosen pressure (e.g., potential temperature described in Section 22.9).

21.3.2 Equation of state in terms of potential temperature

Equation (22.113) provides a unique relation between potential temperature θ and *in situ* temperature, T, salinity and pressure. Furthermore, as discussed in Section 23.2.7, potential temperature is more convenient for modeling than *in situ* temperature. Hence, it is useful to express the *in situ* density as a function of salinity, potential temperature, and pressure

$$\rho = \rho(S, \theta, p). \tag{21.8}$$

One comment on notation is key here. Namely, the functions $\rho(S, T, p)$ and $\rho(S, \theta, p)$ are distinct, so that their respective polynomial coefficients are different. Hence, one may wish to introduce distinct notation to distinguish the two functions; e.g., $\rho = F(S, T, p) = G(S, \theta, p)$. However, we choose brevity in notation by allowing the functional dependence to make the distinction. Doing so is standard in the oceanography literature.

21.3.3 Equation of state in terms of Consevative Temperature

For seawater, potential enthalpy is a more accurate means to measure the heat transfer through the ocean than potential temperature *McDougall* (2003). Consequently, since the advent of the Thermodynamic Equation of State 2010 (*IOC et al.*, 2010), oceanographers make use of Conservative Temperature, Θ, rather than potential temperature for modeling and observations. For this reason we often compute the density as a function

$$\rho = \rho(S, \Theta, p). \tag{21.9}$$

For our purposes it is not important to make the distinction between potential temperature and Conservative Temperature. We use θ for brevity throughout this chapter.

21.3.4 Infinitesimal density increments and material time changes

Given the functional dependence for the equation of state written in terms of S, θ, p (equation (21.8)), an infinitesimal density increment is given by

$$\mathrm{d}\rho = \left[\frac{\partial \rho}{\partial S}\right] \mathrm{d}S + \left[\frac{\partial \rho}{\partial \theta}\right] \mathrm{d}\theta + \left[\frac{\partial \rho}{\partial p}\right] \mathrm{d}p \tag{21.10a}$$

$$\equiv \rho\,\beta\,\mathrm{d}S - \rho\,\alpha\,\mathrm{d}\theta + c_{\mathrm{sound}}^{-2}\,\mathrm{d}p. \tag{21.10b}$$

The second line introduced the following thermodynamic properties of the fluid

$$\beta = \frac{1}{\rho}\left[\frac{\partial \rho}{\partial S}\right]_{\theta,p} \qquad \text{haline contraction coefficient} \tag{21.11}$$

$$\alpha = -\frac{1}{\rho}\left[\frac{\partial \rho}{\partial \theta}\right]_{S,p} \qquad \text{thermal expansion coefficient} \tag{21.12}$$

$$c_{\mathrm{sound}}^2 = \left[\frac{\partial p}{\partial \rho}\right]_{S,\theta} \qquad \text{squared sound speed.} \tag{21.13}$$

The haline contraction coefficient, β, is considered for the ocean, where *haline* refers to salinity.[1]

The infinitesimal density increment (21.10b) leads to the expression for the material change

$$\frac{1}{\rho}\frac{\mathrm{D}\rho}{\mathrm{D}t} = \beta\,\frac{\mathrm{D}S}{\mathrm{D}t} - \alpha\,\frac{\mathrm{D}\theta}{\mathrm{D}t} + \frac{1}{\rho\,c_{\mathrm{sound}}^2}\,\frac{\mathrm{D}p}{\mathrm{D}t}. \tag{21.14}$$

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

$$\frac{\mathrm{D}\rho}{\mathrm{D}t} = \frac{1}{c_{\mathrm{sound}}^2}\,\frac{\mathrm{D}p}{\mathrm{D}t} \qquad \text{adiabatic and isohaline changes.} \tag{21.15}$$

21.3.5 Potential density

Isentropic motion of a frictionless fluid element generally occurs at materially constant potential temperature and materially constant tracer concentration (e.g., salinity or humidity) (see Section 22.9.3). We 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.

[1] Note that in many chapters of this book, $\beta = \partial f/\partial y$ is the meridional derivative of the Coriolis parameter. We are careful to keep the two usages for β distinct.

Defining the potential density

We start by recalling the definition of potential temperature, which is the temperature of a fluid element that is isentropically moved to the reference pressure, p_{R}. This isentropic displacement leads to the implicit expression for the potential temperature (equation (22.118))

$$\mathcal{S}(S, T, p) = \mathcal{S}(S, \theta, p_{\mathrm{R}}). \tag{21.16}$$

We define potential density similarly, by computing the density a fluid element would have if isentropically moved to the same reference pressure as the potential temperature

$$\rho_{\mathrm{pot}} = \rho(S, \theta, p_{\mathrm{R}}). \tag{21.17}$$

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 (21.17), the material evolution of potential density is given by

$$\frac{1}{\rho_{\mathrm{R}}} \frac{\mathrm{D}\rho_{\mathrm{pot}}}{\mathrm{D}t} = \beta_{\mathrm{R}} \frac{\mathrm{D}S}{\mathrm{D}t} - \alpha_{\mathrm{R}} \frac{\mathrm{D}\theta}{\mathrm{D}t}, \tag{21.18}$$

where

$$\beta_{\mathrm{R}} = \frac{1}{\rho_{\mathrm{R}}} \left[\frac{\partial \rho(S, \theta, p_{\mathrm{R}})}{\partial S} \right]_{\theta} \qquad \text{haline contraction at } p = p_{\mathrm{R}} \tag{21.19}$$

$$\alpha_{\mathrm{R}} = -\frac{1}{\rho_{\mathrm{R}}} \left[\frac{\partial \rho(S, \theta, p_{\mathrm{R}})}{\partial \theta} \right]_{S} \qquad \text{thermal expansion at } p = p_{\mathrm{R}} \tag{21.20}$$

are the haline contraction and thermal expansion coefficients evaluated at the reference pressure $p = p_{\mathrm{R}}$. Potential temperature and salinity are materially constant for adiabatic motion that also maintains constant matter content (e.g., isohaline) for fluid elements. By construction, potential density is also materially constant for this motion. This behavior is in contrast to *in situ* density, whose evolution is affected by pressure changes as seen by equations (21.14) and (21.15).

Reference pressures for ρ_{pot} and θ

As defined by equation (21.17), the reference pressure for the potential density is assumed to be the same as for the potential temperature. This assumption is common 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, it is common to consider potential density with larger reference pressures, such as found in 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.

21.3.6 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

$$\rho_{\text{pot}} = \rho_0 \left[1 - \alpha \left(\theta - \theta_0\right) + \beta \left(S - S_0\right)\right], \tag{21.21}$$

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.

21.3.7 Further study

Chapter 1 of *Vallis* (2017) provides a pedagogical discussion of the equation of state for both a moist atmosphere and for seawater, as well as a discussion of the various flavors of density. The seawater equation of state is far more complex than the atmosphere, with *IOC et al.* (2010) providing the authoritative treatment.

21.4 Buoyancy stratification

We now return to the notions of buoyancy as introduced in Section 21.2, with a particular focus on how buoyancy is stratified.

21.4.1 Buoyancy as a field

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 geophysical fluid mechanics, we consider the massive body to be a fluid element whose presence does not alter the flow field; i.e., a "test" fluid element.

Locally defined environment

Consider a local definition of fluid buoyancy according to

$$b_{\text{local}} = g \left(\rho_{\text{environ}} - \rho_{\text{element}}\right)/\rho_{\text{environ}} = g \left(1 - \rho_{\text{element}}/\rho_{\text{environ}}\right), \tag{21.22}$$

where ρ_{environ} is the local density of the fluid environment, and ρ_{element} 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 buoyancy, and vice versa.

We determine ρ_{element} 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 (21.22) 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.*

Globally defined environment

The definition (21.22) accepts that buoyancy is a relative field in which at each point in the fluid one must redefine the environment to which to compare the density of the test fluid element. However, there are cases in which it is sufficient to define a globally constant environment with a constant density, ρ_{ref}. In this case we consider the global buoyancy as

$$b_{\text{global}} = g\,(1 - \rho/\rho_{\text{ref}}), \tag{21.23}$$

where we compute ρ according to the local environmental density. This definition is 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 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, in which case the reference density is irrelevant since it is a global constant. 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 just at that point. Correspondingly, we can map b_{global} and determine the relative buoyancy of fluid elements anywhere in the fluid.

Comments

Although appealing, the use of a globally constant reference density has its limitations for real fluids whose density is a function of pressure. Hence, for the remainder of this section we retain the local definition of buoyancy given by equation (21.22) and make use of the local definition to determine neutral directions as defined below.

21.4.2 Physical ideas underlying neutral directions

As a constant mass fluid element moves through the ocean and atmosphere, it is exposed to a suite of physical processes that 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 (Chapter 19). Modification of the thermal and material properties occurs through the exchange of heat and matter with adjacent fluid elements. The exchange of heat and matter occurs only in the presence of irreversible processes such as mixing, whereas mechanical exchanges occur either reversibly (pressure exchange) or irreversibly (viscous exchange; Section 19.2).

The exchange of heat and matter generally alters the density of a fluid element relative to its local environment, thus affecting the buoyancy of the fluid element. However, it is possible to modify θ and S without modifying its buoyancy defined local to the environment surrounding the fluid element. To do so requires changes in θ to precisely compensate changes in S. Directions in space determined by such compensated changes define *neutral directions*.

The *in situ* density generally changes when a fluid element is displaced an infinitesimal distance, $\mathrm{d}\boldsymbol{x}$. How the density changes is determined by how the element interacts with the surrounding environment. We examine two cases.

- Displace the fluid element allowing for θ, S, and p to equilibrate with the local environment; i.e., full mixing of the element with the environment.

- Displace the fluid element without changing θ and S yet allowing p it to equilibrate with the local environment; i.e., no mixing of the element and the environment.

Notably, the pressure of the fluid element is modified the same amount under both displacements since in both cases the element reaches the same mechanical equilibrium with the local environment. Hence, subtracting the *in situ* density of the above two displaced elements removes the effects from pressure changes, leaving only the effects on density from changes to θ and S. Directions where the *in situ* density is the same for the two forms of parcel displacements determine neutral directions.

The above thought experiment is identical to that used to determine the gravitational stability of a vertical fluid column (e.g., Section 3.6 of *Gill*, 1982), whereby $\mathrm{d}\boldsymbol{x} = \hat{\boldsymbol{z}}\,\mathrm{d}z$ is a vertical displacement. The only difference is that neutral directions consider arbitrary three dimensional displacements rather than just vertical displacements.

21.4.3 Comparing density under two forms of displacement

Consider an infinitesimal displacement, $\mathrm{d}\boldsymbol{x}$, of a fluid element and examine how its *in situ* density $\rho = \rho(S, \theta, p)$ changes under two different displacements. First, assume that the element exchanges heat and salt with the surroundings as it moves, and that it feels the local pressure. In equilibrium, density at the new location agrees with that of the local environment, $\rho(\boldsymbol{x} + \mathrm{d}\boldsymbol{x})$, since the salinity, potential temperature, and pressure are those of the local environment. To leading order, the density difference between the two positions is computed according to

$$\mathrm{d}\rho = \rho(\boldsymbol{x} + \mathrm{d}\boldsymbol{x}) - \rho(\boldsymbol{x}) \tag{21.24a}$$

$$= \rho[S(\boldsymbol{x} + \mathrm{d}\boldsymbol{x}), \theta(\boldsymbol{x} + \mathrm{d}\boldsymbol{x}), p(\boldsymbol{x} + \mathrm{d}\boldsymbol{x})] - \rho[S(\boldsymbol{x}), \theta(\boldsymbol{x}), p(\boldsymbol{x})] \tag{21.24b}$$

$$= \mathrm{d}\boldsymbol{x} \cdot \left[\frac{\partial \rho}{\partial \theta}\nabla\theta + \frac{\partial \rho}{\partial S}\nabla S + \frac{\partial \rho}{\partial p}\nabla p\right] \tag{21.24c}$$

$$= \rho\,\mathrm{d}\boldsymbol{x} \cdot \left[-\alpha\,\nabla\theta + \beta\,\nabla S + \frac{1}{\rho\, c_{\text{sound}}^2}\nabla p\right]. \tag{21.24d}$$

For the second displacement, do not allow the fluid element to exchange (mix) heat or salt with the environment, thus undergoing an adiabatic and isohaline motion. In this case, the element's density change is associated just with pressure changes

$$(\mathrm{d}\rho)_{(\text{no mix})} = \rho(\boldsymbol{x} + \mathrm{d}\boldsymbol{x})_{(\text{no mix})} - \rho(\boldsymbol{x}) \tag{21.25a}$$

$$= \rho[S(\boldsymbol{x}), \theta(\boldsymbol{x}), p(\boldsymbol{x} + \mathrm{d}\boldsymbol{x})] - \rho[S(\boldsymbol{x}), \theta(\boldsymbol{x}), p(\boldsymbol{x})] \tag{21.25b}$$

$$= \rho\,\mathrm{d}\boldsymbol{x} \cdot \left[\frac{1}{\rho\, c_{\text{sound}}^2}\nabla p\right]. \tag{21.25c}$$

That is, when the fluid element moves through the fluid without exchange of heat or salt, then the only way for the *in situ* density to change is via pressure changes. Comparing the two densities renders

$$\rho(\boldsymbol{x} + \mathrm{d}\boldsymbol{x}) - \rho(\boldsymbol{x} + \mathrm{d}\boldsymbol{x})_{(\text{no mix})} = \mathrm{d}\rho - (\mathrm{d}\rho)_{(\text{no mix})} = \rho\,\mathrm{d}\boldsymbol{x} \cdot [-\alpha\,\nabla\theta + \beta\,\nabla S]\,. \tag{21.26}$$

Figure 21.2 offers a schematic of the calculation used to determine neutral directions for the special case of a vertical displacement.

$\rho[S(\boldsymbol{x}+\mathrm{d}\boldsymbol{x}), \theta(\boldsymbol{x}+\mathrm{d}\boldsymbol{x}), p(\boldsymbol{x}+\mathrm{d}\boldsymbol{x})] < \rho[S(\boldsymbol{x}), \theta(\boldsymbol{x}), p(\boldsymbol{x}+\mathrm{d}\boldsymbol{x})] \Longrightarrow$ **gravitationally stable**

$\rho[S(\boldsymbol{x}+\mathrm{d}\boldsymbol{x}), \theta(\boldsymbol{x}+\mathrm{d}\boldsymbol{x}), p(\boldsymbol{x}+\mathrm{d}\boldsymbol{x})] > \rho[S(\boldsymbol{x}), \theta(\boldsymbol{x}), p(\boldsymbol{x}+\mathrm{d}\boldsymbol{x})] \Longrightarrow$ **gravitationally unstable**

Figure 21.2: *Schematic of the calculation used to examine whether a fluid column is gravitationally stable under vertical displacement of a fluid element. A fluid element is displaced from its original location at a position $\boldsymbol{x}$ to a position $\boldsymbol{x}+\mathrm{d}\boldsymbol{x}$, with $\mathrm{d}\boldsymbol{x} = \hat{\boldsymbol{z}}\,\mathrm{d}z$ and $\mathrm{d}z > 0$ in this figure. If the displacement occurs without mixing, the fluid element carries its original potential temperature and matter concentration, whereas the pressure acting on the element generally changes to that of the local environment (right path). Comparing this "no-mixing" displaced fluid element to the density of the local environment (left path) determines whether the density stratification of the fluid environment is gravitationally stable or unstable.*

21.4.4 Buoyancy frequency

The special case of a vertical displacement (Figure 21.2) yields

$$\rho(z+\mathrm{d}z) - \rho(z+\mathrm{d}z)_{(\text{no mix})} = \rho\,\mathrm{d}z\left[-\alpha\,\frac{\partial\theta}{\partial z} + \beta\,\frac{\partial S}{\partial z}\right]. \tag{21.27}$$

Consider a vertically upward displacement so that $\mathrm{d}z > 0$. If the surrounding environment has a lower density than the adiabatic and isohaline displaced fluid element, $\rho(z+\mathrm{d}z) < \rho(z+\mathrm{d}z)_{(\text{no mix})}$, then the element will feel a buoyancy force returning it to the original depth. The restorative buoyancy force per volume is written

$$g\left[\rho(z+\mathrm{d}z) - \rho(z+\mathrm{d}z)_{(\text{no mix})}\right] = g\,\rho\,\mathrm{d}z\left[-\alpha\,\frac{\partial\theta}{\partial z} + \beta\,\frac{\partial S}{\partial z}\right] \equiv -N^2\,\rho\,\mathrm{d}z, \tag{21.28}$$

where we defined the squared buoyancy frequency

$$N^2 = g\left[\alpha\,\frac{\partial\theta}{\partial z} - \beta\,\frac{\partial S}{\partial z}\right]. \tag{21.29}$$

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 fluid element results in an exponential growth associated with a *gravitational instability*. That is, when the fluid column is unstably stratified, an initially tiny vertical displacement will lead to an even larger displacement, thus causing the perturbation to grow unbounded, which is the sign of an instability.

21.4.5 Buoyancy frequency and locally referenced potential density

Equation (21.29) 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 (21.17), 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

$$N^2 = -g \left[\frac{\partial \ln \rho_{\rm pot}}{\partial z} \right]_{p_{\rm R}=p} = g \left[\alpha \frac{\partial \theta}{\partial z} - \beta \frac{\partial S}{\partial z} \right]. \tag{21.30}$$

Note that at a point in the fluid, the locally referenced potential density equals to the *in situ* density. However, when probing nearby points, and thus in taking spatial gradients, the two 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.

21.5 Gravitational stability of a dry ideal gas atmosphere

We introduced the adiabatic lapse rate in Section 22.7 as a measure of how temperature varies as a function of pressure or depth. For an ideal gas atmosphere, the squared buoyancy frequency can be written (see exercise 21.1)

$$N^2 = \frac{g}{\theta} \frac{\partial \theta}{\partial z}. \tag{21.31}$$

The potential temperature for an ideal gas is given by equation (22.128)

$$\theta = T \left[\frac{p_{\rm R}}{p} \right]^{\varphi} \tag{21.32}$$

where

$$\varphi = \frac{R^{\rm M}}{c_p} \tag{21.33}$$

is a constant for a simple ideal gas. Consequently, the squared buoyancy frequency takes the form

$$g^{-1} N^2 = \frac{\partial \ln \theta}{\partial z} \tag{21.34a}$$

$$= \frac{\partial \ln T}{\partial z} - \varphi \frac{\partial \ln p}{\partial z}. \tag{21.34b}$$

For a hydrostatic fluid with a constant gravitational acceleration, the vertical derivative of pressure is given by

$$\frac{\partial p}{\partial z} = -\rho g, \tag{21.35}$$

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

$$g^{-1} N^2 = \frac{\partial \ln T}{\partial z} + \frac{\varphi g \rho}{p} \tag{21.36a}$$

$$= \frac{1}{T} \frac{\partial T}{\partial z} + \frac{g}{c_p T}, \tag{21.36b}$$

where we used the ideal gas relation $p = \rho T R^{\rm M}$ for the final step.

Figure 21.3: *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* $\Gamma_d = -g/c_p \approx -9.8$ *K/(1000 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* $\Gamma_{stable} > \Gamma_d = -g/c_p$*.*

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

$$N^2 = 0 \Longleftrightarrow \frac{\partial T}{\partial z} = \Gamma_d, \tag{21.37}$$

where (see equation (22.111))

$$\Gamma_d = -\frac{g}{c_p} \approx -9.8 \text{ K/(1000 m)}. \tag{21.38}$$

That is, if the *in situ* temperature decreases upon ascent more strongly than the dry adiabatic lapse rate, then the vertical column is gravitationally unstable. In effect, the column becomes top heavy and subject to overturning. We summarize this stability criteria as

$$\text{stable} \qquad N^2 > 0 \Longleftrightarrow -\frac{\partial T}{\partial z} < \frac{g}{c_p} \tag{21.39}$$

$$\text{unstable} \qquad N^2 < 0 \Longleftrightarrow -\frac{\partial T}{\partial z} > \frac{g}{c_p}, \tag{21.40}$$

with Figure 21.3 providing an illustration for three linear temperature profiles.

21.6 Neutral directions

Rather than specializing to a vertical displacement as for the buoyancy frequency in Section 21.4.4, consider an arbitrary three-dimensional displacement. The resulting directions are known as *neutral directions*.

21.6.1 Neutral directions and the neutrality condition

To determine neutral directions, return to equation (21.26) to write

$$\rho(\boldsymbol{x}+\mathrm{d}\boldsymbol{x})-\rho(\boldsymbol{x}+\mathrm{d}\boldsymbol{x})_{(\text{no mix})}=\rho(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x}\cdot[-\alpha\,\nabla\theta+\beta\,\nabla S] \tag{21.41a}$$

$$=\rho(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x}\cdot\hat{\boldsymbol{\gamma}}\,|-\alpha\,\nabla\theta+\beta\,\nabla S|. \tag{21.41b}$$

The second expression introduced the *dianeutral unit vector*

$$\hat{\boldsymbol{\gamma}}=\frac{-\alpha\,\nabla\theta+\beta\,\nabla S}{|-\alpha\,\nabla\theta+\beta\,\nabla S|}. \tag{21.42}$$

Displacements, $\mathrm{d}\boldsymbol{x}$, orthogonal to $\hat{\boldsymbol{\gamma}}$ lead to no difference in the density between the environment and the unmixed fluid element

$$\mathrm{d}\boldsymbol{x}\cdot\hat{\boldsymbol{\gamma}}=0\Rightarrow\rho(\boldsymbol{x}+\mathrm{d}\boldsymbol{x})=\rho(\boldsymbol{x}+\mathrm{d}\boldsymbol{x})_{(\text{no mix})}\Longleftrightarrow\text{neutral displacements}. \tag{21.43}$$

Such displacements are said to occur along a *neutral direction*. Neutral directions generalize the notion of buoyancy surfaces or statification surfaces within the ocean. Motion perpendicular to such surfaces is suppressed through the restoring force from buoyancy.

For a neutral displacement let us write

$$\mathrm{d}\boldsymbol{x}=\hat{\boldsymbol{t}}\,\mathrm{d}s, \tag{21.44}$$

where $\mathrm{d}s$ is the arc-distance along the neutral direction and $\hat{\boldsymbol{t}}$ is the local tangent parallel to the neutral direction. The neutral displacement condition (21.43) leads to the *neutrality condition*

$$\alpha\,\nabla_{\text{neut}}\theta=\beta\,\nabla_{\text{neut}}S, \tag{21.45}$$

where

$$\nabla_{\text{neut}}=\hat{\boldsymbol{t}}\cdot\nabla \tag{21.46}$$

is the gradient operator oriented parallel to the neutral direction.

21.6.2 Comments on neutral directions

As defined, neutral displacements generally occur via the irreversible mixing of θ and S. To be displaced along a neutral direction requires the mixing of θ to precisely balance that of S so that the fluid element's *in situ* density remains identical to that of the local environment. That is, a fluid element displaced along a neutral direction leaves the *in situ* density of the element equal to that of the local environment, thus engendering no local buoyancy force on the element.

21.7 Neutral helicity

As discussed in Section 21.6, movement of a fluid element along a neutral direction requires the mixing of θ and S, with θ mixing precisely balanced by S mixing so that the fluid element's *in situ* density remains identical to that of the local environment. In so doing, the fluid element encounters no locally defined buoyancy force, thus prompting the name "neutral direction". We make the balance of θ mixing and S mixing precise when presenting the neutrality condition in Section 49.3.4. What we ask here concerns the path taken when undergoing a suite of neutral displacements. In particular, if a suite of neutral displacements close in latitude/longitude space, will they also close in depth? As we show here, neutral displacements generally do not close due to a property of seawater known as *neutral helicity*, thus revealing a nontrivial helical topology.

21.7.1 Mathematical preliminaries

Consider a simply connected smooth surface with outward normal written in the form

$$\hat{\boldsymbol{n}} = |\boldsymbol{N}|^{-1}\,\boldsymbol{N}. \tag{21.47}$$

Likewise, consider a unit vector $\hat{\boldsymbol{t}}$ that lives within the surface and is directed tangent to an arbitrary closed loop. Since $\hat{\boldsymbol{n}}\cdot\hat{\boldsymbol{t}} = 0$ by construction, we can integrate around an arbitrary closed loop within the surface and still maintain the trivial result

$$\oint_{\partial\mathcal{S}} \boldsymbol{N}\cdot\hat{\boldsymbol{t}}\,\mathrm{d}l = 0. \tag{21.48}$$

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 (21.48) to yield

$$\int_{\mathcal{S}} (\nabla\wedge\boldsymbol{N})\cdot\hat{\boldsymbol{n}}\,\mathrm{d}\mathcal{S} = 0, \tag{21.49}$$

where $\mathrm{d}\mathcal{S}$ is the area element in the surface $\mathcal{S}$ with outward normal $\hat{\boldsymbol{n}}$. Since the closed path is arbitrary, the area integral (21.49) vanishes only if the integrand is identically zero. We conclude that for the surface to be simply connected requires that the helicity must vanish

$$\mathcal{H} = \boldsymbol{N}\cdot(\nabla\wedge\boldsymbol{N}) = 0 \Longrightarrow \text{simply connected surface.} \tag{21.50}$$

21.7.2 Helical nature of neutral displacements

Now apply the above mathematical results towards the question of whether a neutral surface is simply connected. For that purpose we set

$$\boldsymbol{N} = -\alpha\,\nabla\theta + \beta\,\nabla S \tag{21.51}$$

rendering the neutral helicity

$$\mathcal{H}_{\text{neutral}} = (-\alpha\,\nabla\theta + \beta\,\nabla S)\cdot\nabla\wedge(-\alpha\,\nabla\theta + \beta\,\nabla S), \tag{21.52}$$

which can be written

$$\begin{aligned}
\mathcal{H}_{\text{neutral}} &= (-\alpha\,\nabla\theta + \beta\,\nabla S)\cdot\nabla\wedge(-\alpha\,\nabla\theta + \beta\,\nabla S) && (21.53\text{a})\\
&= -\alpha\,\nabla\theta\cdot(\nabla\wedge\beta\,\nabla S) - \beta\,\nabla S\cdot(\nabla\wedge\alpha\,\nabla\theta) && (21.53\text{b})\\
&= -\alpha\,\nabla\theta\cdot(\nabla\beta\wedge\nabla S) - \beta\,\nabla S\cdot(\nabla\alpha\wedge\nabla\theta). && (21.53\text{c})
\end{aligned}$$

Expand the gradients of α and β according to

$$\begin{aligned}
\nabla\alpha &= \alpha_\theta\,\nabla\theta + \alpha_S\,\nabla S + \alpha_p\,\nabla p && (21.54\text{a})\\
\nabla\beta &= \beta_\theta\,\nabla\theta + \beta\,\nabla S + \beta_p\,\nabla p, && (21.54\text{b})
\end{aligned}$$

so that

$$\begin{aligned}
-\mathcal{H}_{\text{neutral}} &= \alpha\,\nabla\theta\cdot(\beta_\theta\,\nabla\theta + \beta_p\,\nabla p) + \beta\,\nabla S\cdot(\alpha_S\,\nabla S + \alpha_p\,\nabla p) && (21.55\text{a})\\
&= \alpha\,\nabla\theta\cdot(\nabla p\wedge\nabla S)\,\beta_p + \beta\,\nabla S\cdot(\nabla p\wedge\nabla\theta)\,\alpha_p && (21.55\text{b})\\
&= \nabla p\cdot(\nabla S\wedge\nabla\theta)\,(\alpha\,\beta_p - \beta\,\alpha_p). && (21.55\text{c})
\end{aligned}$$

Introducing the thermobaricity parameter from Section 49.4

$$\mathcal{T} = \beta\, \partial_p(\alpha/\beta) \tag{21.56}$$

renders the tidy result

$$\mathcal{H}_{\text{neutral}} = \beta\, \mathcal{T}\, \nabla p \cdot (\nabla S \wedge \nabla \theta). \tag{21.57}$$

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)

$$\nabla p \cdot (\nabla S \wedge \nabla \theta) = \nabla \theta \cdot (\nabla p \wedge \nabla S) = \nabla S \cdot (\nabla \theta \wedge \nabla p). \tag{21.58}$$

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.

21.7.3 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 stratification. *McDougall et al.* (2014) offer another presentation of why neutral directions are relevant for the ocean, taking into account measurements of dianeutral mixing. Section 2.7.2 of *Olbers et al.* (2012) offers a concise and pedagogical summary of neutral directions.

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) were the first to make note of the helical nature of 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 some general mathematical properties encountered in a variety of thermodynamic systems. The helical property of the seawater equation of state, though somewhat exotic, has some very practical implications on the choice for vertical coordinate used in realistic numerical ocean climate models. *Stanley* (2019) offers recent insights into the mathematics of neutral directions.

21.8 Exercises

EXERCISE 21.1: 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 (21.30).

EXERCISE 21.2: EXAMPLES OF BUOYANCY PERIOD

Using approximate but realistic values for the observed stratification, determine the buoyancy period ($T_{\text{b}} = 2\pi/N$) 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 21.3: VERTICAL INTEGRAL OF N^2

The expression for squared buoyancy frequency

$$N^2 = -g \left[\frac{\partial \ln \rho_{\text{pot}}}{\partial z} \right]_{p_{\text{R}}=p} \tag{21.59}$$

makes is tempting to consider its vertical integral according to

$$-\, g^{-1} \int_{-H}^{\eta} N^2 \, \mathrm{d}z \stackrel{?}{=} \left[\ln \rho_{\text{pot}} \right]_{\eta} - \left[\ln \rho_{\text{pot}} \right]_{-H} . \tag{21.60}$$

Discuss what is wrong with this equation. Under what conditions is it correct?

22

Equilibrium thermodynamics

Thermodynamics is a phenomenological discipline whose fundamentals lie in statistical mechanics. In this chapter we develop those elements of equilibrium thermodynamics relevant to a single-phase multi-component fluid system in near equilibrium. We pick up the topic of linear irreversible thermodynamics in Chapter 23 when applying thermodynamics to a moving fluid.

READER'S GUIDE FOR THIS CHAPTER

One aim for this chapter is to expose the basics of equilibrium thermodynamics without diving too deeply into details. Even so, the concepts and associated mathematical formalism can be rather dense and abstruse on first reading, with the beauty and power of thermodynamics often unveiled only after further study. In this chapter the reader is assumed to have a working understanding of partial differential calculus at the level of Chapter 2.

Thermodynamics is a foundational subject in physics, chemistry, biology, and engineering. A treatment based on the postulates of thermodynamics is given by *Callen* (1985). *Reif* (1965) offers an elementary and very accessible connection between thermodynamics and statistical mechanics, whereas *Huang* (1987) and *Reichl* (1987) present more advanced topics.

22.1 Conceptual basics **370**
22.1.1 Exchanges between thermodynamic systems 370
22.1.2 Quasi-static, reversible, and irreversible processes 371
22.2 First Law for materially closed systems **372**
22.2.1 Basic expression of the First Law 372
22.2.2 Path dependent (inexact) nature of working and heating 372
22.2.3 Mechanical work 373
22.2.4 Transfer of thermal energy and the Second Law 373
22.2.5 Extensive and intensive properties 373
22.2.6 Fundamental thermodyanamic relation 374
22.3 First Law for materially open systems **374**
22.3.1 Constituent masses and matter concentrations 374
22.3.2 Scaling for homogeneous functions 375
22.3.3 Chemical work and the Gibbs-Duhem relation 376
22.3.4 Partial derivatives and equations of state 376
22.3.5 Fundamental thermodynamic relation per unit mass 377
22.3.6 Seawater as a binary fluid 377
22.3.7 Further study 378

22.4 Thermodynamic potentials **378**
22.4.1 Internal energy 378
22.4.2 Entropy 379
22.4.3 Enthalpy 379
22.4.4 Helmholz free energy 380
22.4.5 Gibbs potential 381
22.5 Response functions **382**
22.5.1 Heat capacities 382
22.5.2 Thermal expansion, haline contraction, and sound speed 382
22.6 Maxwell relations for homogeneous fluids **383**
22.6.1 Maxwell relation from internal energy 383
22.6.2 Summary of the Maxwell relations 384
22.7 Adiabatic lapse rate **384**
22.7.1 Isentropic rearrangement 384
22.7.2 Thermodynamic formulation 385
22.7.3 Adiabatic lapse rate for pressure changes 385
22.7.4 Adiabatic lapse rate for height changes 385
22.7.5 Further study 386
22.8 A simple ideal gas atmosphere **386**
22.8.1 Equation of state 386
22.8.2 Internal energy 387
22.8.3 Heat capacity from statistical mechanics 387
22.8.4 Enthalpy 388
22.8.5 Thermal expansion coefficient 388
22.8.6 Fundamental thermodynamic relations 389
22.8.7 Adiabatic lapse rate 389
22.9 Potential temperature **389**
22.9.1 Adiabatic temperature changes 390
22.9.2 Defining the potential temperature 390
22.9.3 Potential temperature and specific entropy 391
22.9.4 Potential temperature for an ideal gas 392
22.10 Exercises **392**

22.1 Conceptual basics

In the study of thermodynamics it is useful to characterize how a physical system interacts with its surrounding environment through mechanical, thermal, and material exchanges. Tiny (virtually infinitesimal) fluid elements, and their accumulation into finite fluid regions, constitute the physical systems we are concerned with in this book. Furthermore, we are only concerned with systems that interact mechanically with their surroundings through macroscopic forces, such as pressure, so that the systems are mechanically open. However, we find it useful to distinguish how a physical system interacts thermally and materially and the time scales over which these interactions take place.

22.1.1 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 mechanially open. Such *material fluid parcels* (Section 12.2) maintain fixed matter and thermal properties yet move according to imbalances in mechanical forces. A *perfect fluid* is comprised of material fluid parcels.

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 slight corruption of terminology, "adiabatic" in fluid mechanics is often used for a system that is *both* thermally *and* materially closed. However, we maintain the distinction in our treatment.

22.1.2 Quasi-static, reversible, and irreversible processes

In our study of thermodynamics for moving fluid systems, 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 (effectively) infinite number of intermediate equilibrium states. This definition is an idealization of the actual situation in which macroscopic processes evolve via macroscopic time increments, $\Delta\tau_{\text{macro}}$, that are much much larger than the time required for equilibration of molecular degrees of freedom. We estimate the molecular equilibration time by the time between molecular collisions. As seen in Section 11.2.5, the time between molecular collisions is less than $t_{\text{collision}} \approx 10^{-9}$ s, which is incredibly tiny from a macroscopic perspective. Hence, any macroscopic process moves through time increments that are much larger than the molecular collision time

$$\Delta\tau_{\text{macro}} \gg t_{\text{collision}} \approx 10^{-9} \text{ s}. \tag{22.1}$$

This time scale separation means that the molecular motions have no problem equilibrating between each of the $\Delta\tau_{\text{macro}}$ time increments. In fluid mechanics, we assume that all fluid motions occur through a series of thermodynamically equilibrated states so that the processes evolve quasi-statically. The justification for that assumption is that the relevant time scales for macroscopic fluid motions are much larger than the molecular time scales. Quasi-static processes allow us to precisely define intensive quantities for a thermodynamic system, such as its temperature, pressure, specific volume, specific entropy, and specific internal energy.

A *reversible process* is one that can, at each stage, go either forward or backward in time; i.e., there is no special direction for time. Newton's dynamical laws in the absence of non-conservative forces (e.g., friction) are reversible. Furthermore, every reversible process is quasi-static and yet not all quasi-static processes are reversible. For example, a process that involves friction can proceed quasi-statically and yet the friction generates entropy. Consequently, frictional (dissipative) processes are irreversible. From a statistical mechanics perspective, *entropy* provides a measure of 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 of a system undergoing a reversible process. In contrast, as codified by the *Second Law of thermodynamics*, irreversible processes always increase entropy by increasing the number of accessible microstates.

22.2 First Law for materially closed systems

As discussed in Chapter 11, 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 fluid with fluid elements that are materially closed yet thermally open. In Section 22.3 we consider the case of a materially open system, whereas the case of multiple phases is beyond the scope of this book.

22.2.1 Basic expression of the First Law

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 the system, and the thermal energy transferred between the system and its surrounding environment. The First Law takes on the mathematical form

$$\mathrm{d}\mathcal{I}^{\mathrm{e}} = đ\mathcal{W} + đ\mathcal{Q} \qquad \Longleftarrow \text{materially closed.} \tag{22.2}$$

In this equation, $\mathrm{d}\mathcal{I}^{\mathrm{e}}$ is the infinitesimal increment (an exact differential) of the internal energy; $đ\mathcal{W}$ is the change in internal energy due to mechanical work applied to the system (*working*); and $đ\mathcal{Q}$ is the internal energy change due to thermal energy transferred to the system (*heating*).

22.2.2 Path dependent (inexact) nature of working and heating

Working and heating are both path-dependent thermodynamic processes that transform a system from one thermodynamic state to another. They are mathematically represented by inexact differentials using the đ symbol. Even so, the First Law in equation (22.4) reflects the mathematical property that the sum of two inexact differentials can equal to an exact differential. It follows that if the internal energy change occurs in the absence of heating then the working process must occur as a path independent or exact process. The converse holds if internal energy changes without any work applied to the system.

Working and heating denote actions (verbs) rather than properties (nouns). We raise this somewhat pedantic point since the terms "work" and "heat" are often used, which can spuriously lead one to seek information and the "work content" or "heat content" of a physical state. Rather, we can only talk about the work imparted to change a system (working) or likewise the thermal energy used to change a system (heating).

These 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 22.4.3). 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 carefully revisit assumptions of the analysis.

22.2.3 Mechanical work

There are many ways that mechanical forces can do work to a physical system. For fluid mechanics we generally focus on changes to the volume of a fluid element through the action of pressure, whereby the *pressure-work* takes the form

$$đ\mathcal{W} = -p\,\mathrm{d}V, \tag{22.3}$$

where p is the pressure applied to the boundaries of the fluid element. The negative sign in the pressure-work relation (22.3) arises since the compression of a fluid element into a smaller volume ($\mathrm{d}V < 0$) requires positive mechanical work be applied to the fluid ($đ\mathcal{W} > 0$). Our focus on pressure-work ignores surface tension effects, with surface tension of little concern for geophysical fluid motions on scales larger than a few centimeters (see Chapter 44 for a discussion of surface tension.)

Pressure is an intensive variable that measures the *intensity* of a force that is conjugate to the extensive variable V. Pressure is also the integrating factor connecting the inexact (path dependent) differential $đ\mathcal{W}$ to the exact differential $\mathrm{d}V$ (see Section 2.8 for the mathematics of inexact differentials).

22.2.4 Transfer of thermal energy and the Second Law

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 closed system whose thermal properties change in a quasi-static manner, the thermal energy transfer is related to entropy changes via

$$đ\mathcal{Q} = T\,\mathrm{d}\mathcal{S}^{\mathrm{e}} \qquad \text{materially closed system.} \tag{22.4}$$

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 $\mathcal{S}^{\mathrm{e}}$ is the extensive form of entropy. Entropy is an extensive state function so that T provides the integrating factor connecting the inexact differential $đ\mathcal{Q}$ to the exact differential $\mathrm{d}\mathcal{S}^{\mathrm{e}}$. A nonzero $đ\mathcal{Q}$ 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 mixing.

22.2.5 Extensive and intensive properties

The internal energy and entropy are both proportional to the size of the system, with systems having more volume and mass having more internal energy and entropy. As discussed in Section 15.3.1, properties whose value change when the system size changes are termed *extensive*. Extensive properties are labeled with a superscript e (except for the mass and volume), with this label *not* a

tensor index. Extensive properties are contracted to *intensive* properties such as temperature and pressure whose value is not directly related to the size of a physical system.

Thus far in our development we have found that thermodynamic relations pair intensive properties with extensive properties. However, in Section 22.3 we will introduce the internal energy per mass and entropy per mass, converting the extensive properties to their intensive form. It is the intensive form that provides a more straightforward transfer to the study of moving fluid elements.

22.2.6 Fundamental thermodyanamic relation

We summarize the discussion of this section by writing the First Law for quasi-static materially closed processes

$$\mathrm{d}\mathcal{I}^{\mathrm{e}} = -p\,\mathrm{d}V + T\,\mathrm{d}\mathcal{S}^{\mathrm{e}} \qquad \text{quasi-static materially closed processes.} \tag{22.5}$$

This relation is known as the *fundamental thermodyanamic relation* for quasi-static materially closed processes. It is notable that all differentials are exact, which is a result of assuming the processes are quasi-static. Although derived for quasi-static processes from the First Law using connections to work and heat, equation (22.5) holds for arbitrary infinitesimal changes and it offers great utility in the formalism of thermodynamics. Its extension to materially open systems is given in Section 22.3. Since we are only concerned with quasi-static changes to fluid elements in this book, these relations provide the starting point for all of our thermodynamic formulations.

It follows from equation (22.5) that the internal energy is a natural function of volume and entropy

$$\mathrm{d}\mathcal{I}^{\mathrm{e}} = -p\,\mathrm{d}V + T\,\mathrm{d}\mathcal{S}^{\mathrm{e}} \Longrightarrow \mathcal{I}^{\mathrm{e}} = \mathcal{I}^{\mathrm{e}}(V, \mathcal{S}^{\mathrm{e}}). \tag{22.6}$$

Conversely, the entropy for a materially closed system is naturally a function of volume and internal energy

$$T\,\mathrm{d}\mathcal{S}^{\mathrm{e}} = \mathrm{d}\mathcal{I}^{\mathrm{e}} + p\,\mathrm{d}V \Longrightarrow \mathcal{S}^{\mathrm{e}} = \mathcal{S}^{\mathrm{e}}(V, \mathcal{I}^{\mathrm{e}}). \tag{22.7}$$

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.

22.3 First Law for materially open systems

We now extend the discussion of Section 22.2 to materially open physical systems whose changes occur through quasi-static processes. These considerations have applications to real fluid elements whose matter content is modified by mixing with its surrounding environment.

22.3.1 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

$$M = \sum_{n=1}^{N} M_n. \tag{22.8}$$

As we are interested in fluid elements, we restrict our attention on constant mass physical systems, even as the system exchanges matter content with its surrounding environment. The corresponding

matter concentration is written

$$C_n = M_m/M \Longrightarrow \sum_{n=1}^{N} C_n = 1, \tag{22.9}$$

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 15.1.

22.3.2 Scaling for homogeneous functions

Both the internal energy, $\mathcal{I}^{\mathrm{e}}$, and entropy, $\mathcal{S}^{\mathrm{e}}$, 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 (22.6) and (22.7) extended to include the matter content

$$\mathcal{I}^{\mathrm{e}} = \mathcal{I}^{\mathrm{e}}(V, \mathcal{S}^{\mathrm{e}}, M_n) \qquad \text{and} \qquad \mathcal{S}^{\mathrm{e}} = \mathcal{S}^{\mathrm{e}}(V, \mathcal{I}^{\mathrm{e}}, M_n), \tag{22.10}$$

where the M_n argument is shorthand for $M_1, M_2 ... M_N$ for N matter constituents.

Now scale the system by an arbitrary parameter λ. Under this scale operation, the extensive variables, $\mathcal{I}^{\mathrm{e}}$, $\mathcal{S}^{\mathrm{e}}$, as well as the volume, V, and masses, M_n, scale by the same scale factor. The relation (22.10) thus leads to

$$\mathcal{I}^{\mathrm{e}}(\lambda\, V, \lambda\, \mathcal{S}^{\mathrm{e}}, \lambda\, M_n) = \lambda\, \mathcal{I}^{\mathrm{e}}(V, \mathcal{S}^{\mathrm{e}}, M_n) \qquad \text{and} \qquad \mathcal{S}^{\mathrm{e}}(\lambda\, V, \lambda\, \mathcal{I}^{\mathrm{e}}, \lambda\, M_n) = \lambda\, \mathcal{S}^{\mathrm{e}}(V, \mathcal{I}^{\mathrm{e}}, M_n) \tag{22.11}$$

A function that scales in this way is termed a *homogeneous function of degree one*. Such functions possess the following special property found by differentiating both sides of equation (22.11) with respect to λ and then setting $\lambda = 1$

$$\mathcal{I}^{\mathrm{e}}(V, \mathcal{S}^{\mathrm{e}}, M_n) = V \left[\frac{\partial \mathcal{I}^{\mathrm{e}}}{\partial V}\right]_{\mathcal{S}^{\mathrm{e}}, M_n} + \mathcal{S}^{\mathrm{e}} \left[\frac{\partial \mathcal{I}^{\mathrm{e}}}{\partial \mathcal{S}^{\mathrm{e}}}\right]_{V, M_n} + \sum_{n=1}^{N} M_n \left[\frac{\partial \mathcal{I}^{\mathrm{e}}}{\partial M_n}\right]_{V, \mathcal{S}^{\mathrm{e}}, M_{m\neq n}} \tag{22.12}$$

$$\mathcal{S}^{\mathrm{e}}(V, \mathcal{I}^{\mathrm{e}}, M_n) = V \left[\frac{\partial \mathcal{S}^{\mathrm{e}}}{\partial V}\right]_{\mathcal{I}^{\mathrm{e}}, M_n} + \mathcal{I}^{\mathrm{e}} \left[\frac{\partial \mathcal{S}^{\mathrm{e}}}{\partial \mathcal{I}^{\mathrm{e}}}\right]_{V, M_n} + \sum_{n=1}^{N} M_n \left[\frac{\partial \mathcal{S}^{\mathrm{e}}}{\partial M_n}\right]_{V, \mathcal{I}^{\mathrm{e}}, M_{m\neq n}}. \tag{22.13}$$

Making use of the fundamental thermodynamic relations (22.6) and (22.7) leads to

$$\mathcal{I}^{\mathrm{e}} = -V\, p + \mathcal{S}^{\mathrm{e}}\, T + \sum_{n=1}^{N} M_n \left[\frac{\partial \mathcal{I}^{\mathrm{e}}}{\partial M_n}\right]_{V, \mathcal{S}^{\mathrm{e}}, M_{m\neq n}} \tag{22.14}$$

$$T\, \mathcal{S}^{\mathrm{e}} = V\, p + \mathcal{I}^{\mathrm{e}} + \sum_{n=1}^{N} M_n \left[\frac{\partial \mathcal{S}^{\mathrm{e}}}{\partial M_n}\right]_{V, \mathcal{I}^{\mathrm{e}}, M_{m\neq n}}. \tag{22.15}$$

Self-consistency means that

$$\sum_{n=1}^{N} M_n \left[\frac{\partial \mathcal{I}^{\mathrm{e}}}{\partial M_n}\right]_{V, \mathcal{S}^{\mathrm{e}}} = -\sum_{n=1}^{N} M_n \left[\frac{\partial \mathcal{S}^{\mathrm{e}}}{\partial M_n}\right]_{V, \mathcal{I}^{\mathrm{e}}, M_{m\neq n}}, \tag{22.16}$$

which motivates defining the *chemical potential*

$$\mu_n \equiv \left[\frac{\partial \mathcal{I}^{\mathrm{e}}}{\partial M_n}\right]_{V, \mathcal{S}^{\mathrm{e}}, M_{m\neq n}} = -\left[\frac{\partial \mathcal{S}^{\mathrm{e}}}{\partial M_n}\right]_{V, \mathcal{I}^{\mathrm{e}}, M_{m\neq n}} \tag{22.17}$$

thus leading to

$$\mathcal{I}^{\mathrm{e}} = T\, \mathcal{S}^{\mathrm{e}} - p\, V + \sum_{n=1}^{N} \mu_n\, M_n \Longleftrightarrow T\, \mathcal{S}^{\mathrm{e}} = \mathcal{I}^{\mathrm{e}} + p\, V - \sum_{n=1}^{N} \mu_n\, M_n. \tag{22.18}$$

22.3.3 Chemical work and the Gibbs-Duhem relation

Changes to the matter composition of a system changes the internal energy through *chemical work*, written as $đ\mathcal{C}$. If the matter composition changes occur quasi-statically then the chemical work is written

$$đ\mathcal{C} = \sum_{n=1}^{N} \mu_n \, \mathrm{d}M_n, \tag{22.19}$$

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

$$\mathrm{d}\mathcal{I}^{\mathrm{e}} = đ\mathcal{W} + đ\mathcal{Q} + đ\mathcal{C} \qquad \text{materially open} \tag{22.20}$$

$$\mathrm{d}\mathcal{I}^{\mathrm{e}} = -p \, \mathrm{d}V + T \, \mathrm{d}\mathcal{S}^{\mathrm{e}} + \sum_{n=1}^{N} \mu_n \, \mathrm{d}M_n \qquad \text{quasi-static and materially open.} \tag{22.21}$$

Use of the quasi-static form of the First Law (22.21) along with the differential of the result (22.18) leads to the *Gibbs-Duhem* relation

$$\mathcal{S}^{\mathrm{e}} \, \mathrm{d}T - V \, \mathrm{d}p + \sum_{n} M_n \, \mathrm{d}\mu_n = 0. \tag{22.22}$$

In Exercise 22.1 we work through the derivation of Gibbs-Duhem in a bit more detail.

22.3.4 Partial derivatives and equations of state

The fundamental thermodynamic relation (22.21) appears in terms of internal energy, with the corresponding partial derivative relations derived earlier

$$\left[\frac{\partial \mathcal{I}^{\mathrm{e}}}{\partial \mathcal{S}^{\mathrm{e}}}\right]_{V,M_n} = T \tag{22.23}$$

$$\left[\frac{\partial \mathcal{I}^{\mathrm{e}}}{\partial V}\right]_{\mathcal{S}^{\mathrm{e}},M_n} = -p \tag{22.24}$$

$$\left[\frac{\partial \mathcal{I}^{\mathrm{e}}}{\partial M_n}\right]_{\mathcal{S}^{\mathrm{e}},V,M_{m\neq n}} = \mu_n. \tag{22.25}$$

Each equation relates an intensive property (right hand side) to the partial derivative of internal energy with respect to an extensive property. In thermodynamics, these equations are known as *equations of state*. However, in the geophysical fluid literature, the equation of state generally refers to the *thermal equation of state*, which expresses the mass density in terms of fluid properties (see Section 21.3).

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 the equations of state (22.23)-(22.25) 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.

22.3.5 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 23). For this purpose, one scales away the mass of the system by setting the scale factor $\lambda = M^{-1}$ and introducing the *specific* quantities

$$\mathcal{I}^{\mathrm{e}} = M\,\mathcal{I} \tag{22.26}$$

$$\mathcal{S}^{\mathrm{e}} = M\,\mathcal{S} \tag{22.27}$$

$$V = M\,\nu_{\mathrm{s}} \tag{22.28}$$

$$M_n = M\,C_n, \tag{22.29}$$

where

$$\nu_{\mathrm{s}} = 1/\rho \tag{22.30}$$

is the specific volume and the total mass, M, is held fixed. In the equality (22.29), C_n is the mass fraction or concentration of species n in the fluid (Section 22.3.1). Substituting the specific quantities (22.26)-(22.29) into the fundamental thermodynamic relation (22.21) leads to the fundamental thermodynamic relation in terms of specific thermodynamic quantities

$$\mathrm{d}\mathcal{I} = T\,\mathrm{d}\mathcal{S} - p\,\mathrm{d}\nu_{\mathrm{s}} + \sum_n \mu_n\,\mathrm{d}C_n. \tag{22.31}$$

This is the form of the fundamental thermodynamic relation most commonly used in this book.

22.3.6 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. Hence, for the atmosphere we are only concerned with homogeneous thermodynamic relations applying to a dry atmosphere.

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 satify

$$C_{\mathrm{salt}} + C_{\mathrm{water}} = 1 \Longrightarrow \mathrm{d}C_{\mathrm{water}} = -\mathrm{d}C_{\mathrm{salt}}. \tag{22.32}$$

Introducing this constraint on tracer concentrations into the first Gibbs relation (22.31) leads to

$$\mathrm{d}\mathcal{I} = T\,\mathrm{d}\mathcal{S} - p\,\mathrm{d}\nu_{\mathrm{s}} + \mu\,\mathrm{d}C, \tag{22.33}$$

where

$$C = C_{\mathrm{salt}} \tag{22.34}$$

is the concentration of salt, and

$$\mu = \mu_{\mathrm{salt}} - \mu_{\mathrm{water}} \tag{22.35}$$

is the relative chemical potential. The *absolute salinity* S, with units parts per thousand, is related to C_{salt} via

$$S = 1000\,C_{\mathrm{salt}}. \tag{22.36}$$

The range of salinity in the ocean (roughly $0 \le S \le 40$) is more convenient than the range of C_{salt}, making salinity more commonly used in oceanography.

22.3.7 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 (22.33) when considering a two-component fluid.

22.4 Thermodynamic potentials

The state functions internal energy and entropy are also known as *thermodynamic potentials* related by equation (22.18), written here in its specific form

$$\mathcal{I} = T\,\mathcal{S} - p\,\nu_{\mathrm{s}} + \sum_{n=1}^{N} \mu_n\, C_n \Longleftrightarrow T\,\mathcal{S} = \mathcal{I} + p\,\nu_{\mathrm{s}} - \sum_{n=1}^{N} \mu_n\, C_n. \tag{22.37}$$

Each thermodynamic potential is a natural function of certain other thermodynamic properties, as determined by the fundamental thermodynamic relation. When written in terms of their natural functional dependencies, the expressions for the thermodynamic potentials are known as *fundamental equations of state*.

It is useful to have access to a suite of thermodynamic potentials (internal energy, entropy, enthalpy, Gibbs potential, Helmbolz 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 basic functional relations.

22.4.1 Internal energy

Recall the fundamental thermodynamic relation (22.33) written for a binary fluid

$$\mathrm{d}\mathcal{I} = T\,\mathrm{d}\mathcal{S} - p\,\mathrm{d}\nu_{\mathrm{s}} + \mu\,\mathrm{d}C. \tag{22.38}$$

Equation (22.38) identifies the specific internal energy, $\mathcal{I}$, as a function of specific entropy, $\mathcal{S}$, specific volume, ν_{s}, and matter concentration, C

$$\mathcal{I} = \mathcal{I}(\mathcal{S}, \nu_{\mathrm{s}}, C). \tag{22.39}$$

Knowledge of the fundamental thermodynamic relation (22.38) allows us to derive the plethora of thermodynamic relations. For example, we can identify the partial derivatives

$$\left[\frac{\partial \mathcal{I}}{\partial \mathcal{S}}\right]_{\nu_{\mathrm{s}},C} = T \qquad \left[\frac{\partial \mathcal{I}}{\partial \nu_{\mathrm{s}}}\right]_{\mathcal{S},C} = -p \qquad \left[\frac{\partial \mathcal{I}}{\partial C}\right]_{\mathcal{S},\nu_{\mathrm{s}}} = \mu. \tag{22.40}$$

These equations are the intensive form of the extensive equations of state (22.23)-(22.25)

Equations (22.40) provide a relation between T, p, μ as derivatives of a function, the internal energy, which is itself a function $\mathcal{I}(\mathcal{S}, \nu_{\mathrm{s}}, C)$. Hence, we may consider T, p, μ each as a function of

$(\mathcal{S}, \nu_{\rm s}, C)$ so that their exact differentials are written

$$T = T(\mathcal{S}, \nu_{\rm s}, C) \Longrightarrow \mathrm{d}T = \left[\frac{\partial T}{\partial \mathcal{S}}\right]_{\nu_{\rm s},C} \mathrm{d}\mathcal{S} + \left[\frac{\partial T}{\partial \nu_{\rm s}}\right]_{C,\mathcal{S}} \mathrm{d}\nu_{\rm s} + \left[\frac{\partial T}{\partial C}\right]_{\mathcal{S},\nu_{\rm s}} \mathrm{d}C \tag{22.41}$$

$$p = p(\mathcal{S}, \nu_{\rm s}, C) \Longrightarrow \mathrm{d}p = \left[\frac{\partial p}{\partial \mathcal{S}}\right]_{\nu_{\rm s},C} \mathrm{d}\mathcal{S} + \left[\frac{\partial p}{\partial \nu_{\rm s}}\right]_{C,\mathcal{S}} \mathrm{d}\nu_{\rm s} + \left[\frac{\partial p}{\partial C}\right]_{\mathcal{S},\nu_{\rm s}} \mathrm{d}C \tag{22.42}$$

$$\mu = \mu(\mathcal{S}, \nu_{\rm s}, C) \Longrightarrow \mathrm{d}\mu = \left[\frac{\partial \mu}{\partial \mathcal{S}}\right]_{\nu_{\rm s},C} \mathrm{d}\mathcal{S} + \left[\frac{\partial \mu}{\partial \nu_{\rm s}}\right]_{C,\mathcal{S}} \mathrm{d}\nu_{\rm s} + \left[\frac{\partial \mu}{\partial C}\right]_{\mathcal{S},\nu_{\rm s}} \mathrm{d}C. \tag{22.43}$$

These relations prove of use for deriving certain thermodynamic identities.

22.4.2 Entropy

Rearrangement of the fundamental thermodynamic relation (22.38) leads to the exact differential for specific entropy

$$\mathrm{d}\mathcal{S} = \frac{1}{T}\,\mathrm{d}\mathcal{I} + \frac{p}{T}\,\mathrm{d}\nu_{\rm s} - \frac{\mu}{T}\,\mathrm{d}C. \tag{22.44}$$

In this form, specific entropy has the functional dependence

$$\mathcal{S} = \mathcal{S}(\mathcal{I}, \nu_{\rm s}, C), \tag{22.45}$$

whose knowledge provides yet another form of the fundamental equation of state. This functional dependence, along with equation (22.44), lead to the following thermodynamic equations of state

$$\left[\frac{\partial \mathcal{S}}{\partial \mathcal{I}}\right]_{\nu_{\rm s},C} = \frac{1}{T} \qquad \left[\frac{\partial \mathcal{S}}{\partial \nu_{\rm s}}\right]_{\mathcal{I},C} = \frac{p}{T} \qquad \left[\frac{\partial \mathcal{S}}{\partial C}\right]_{\mathcal{I},\nu_{\rm s}} = -\frac{\mu}{T}. \tag{22.46}$$

As for internal energy in Section 22.4.1, equation (22.46) provides a relation between T, p, μ as derivatives of a function, the entropy, which is itself a function $\mathcal{S}(\mathcal{I}, \nu_{\rm s}, C)$. Hence, we may consider T, p, μ as each a function of $(\mathcal{I}, \nu_{\rm s}, C)$

$$T = T(\mathcal{I}, \nu_{\rm s}, C) \qquad p = p(\mathcal{I}, \nu_{\rm s}, C) \qquad \mu = \mu(\mathcal{I}, \nu_{\rm s}, C). \tag{22.47}$$

22.4.3 Enthalpy

Thus far we have worked only with the fundamental thermodynamic relation (22.38). We now introduce the specific enthalpy

$$\mathcal{H} = \mathcal{I} + p\,\nu_{\rm s} = T\,\mathcal{S} + \sum_{n=1}^{N} \mu_n\, C_n, \tag{22.48}$$

where the second equality made use of equation (22.37). Use of the fundamental thermodynamic relation (22.38) leads to the exact differential for enthalpy

$$\mathrm{d}\mathcal{H} = \mathrm{d}\mathcal{I} + \mathrm{d}(p\,\nu_{\rm s}) \tag{22.49a}$$

$$= T\,\mathrm{d}\mathcal{S} - p\,\mathrm{d}\nu_{\rm s} + \mu\,\mathrm{d}C + p\,\mathrm{d}\nu_{\rm s} + \nu_{\rm s}\,\mathrm{d}p \tag{22.49b}$$

$$= T\,\mathrm{d}\mathcal{S} + \nu_{\rm s}\,\mathrm{d}p + \mu\,\mathrm{d}C. \tag{22.49c}$$

Recalling that for quasi-static processes, $T\,\mathrm{d}\mathcal{S}$ equals to the thermal energy added to a fluid element. Hence, for processes occuring at constant pressure and concentration, changes in enthalpy are determined by the thermal energy added to the fluid.

Equation (22.49c) provides the fundamental thermodynamic relation with enthalpy rather than internal energy. Consequently, the *Legendre transformation* (22.48) renders a functional dependence for enthalpy

$$\mathcal{H} = \mathcal{H}(\mathcal{S}, p, C), \tag{22.50}$$

which is yet another fundamental equation of state whose partial derivatives render

$$\left[\frac{\partial \mathcal{H}}{\partial \mathcal{S}}\right]_{p,C} = T \qquad \left[\frac{\partial \mathcal{H}}{\partial p}\right]_{\mathcal{S},C} = \nu_{\mathrm{s}} \qquad \left[\frac{\partial \mathcal{H}}{\partial C}\right]_{\mathcal{S},p} = \mu. \tag{22.51}$$

As for internal energy in Section 22.4.1, Equations (22.51) 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 $(\mathcal{S}, p, C)$

$$T = T(\mathcal{S}, p, C) \qquad \nu_{\mathrm{s}} = \nu_{\mathrm{s}}(\mathcal{S}, p, C) \qquad \mu = \mu(\mathcal{S}, p, C). \tag{22.52}$$

Enthalpy's functional dependence (22.50) is more convenient for studies of geophysical fluids than that for internal energy, $\mathcal{I}(\mathcal{S}, \nu_{\mathrm{s}}, C)$, or for entropy $\mathcal{S}(\mathcal{I}, \nu_{\mathrm{s}}, C)$. We offer the following reasons:

- 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 21.3.
- Correspondingly, the interaction between fluid elements typically occurs at near constant pressure. Hence, fluid elements exchange both their entropy and ethalpy 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.

22.4.4 Helmholz free energy

The Helmholz free energy is defined by the Legendre transformation

$$\mathcal{F} = \mathcal{I} - T\,\mathcal{S} = -p\,\nu_{\mathrm{s}} + \sum_{n=1}^{N} \mu_n\, C_n, \tag{22.53}$$

where the second equality made use of equation (22.37). The exact differential of the Helmholz free energy is given by

$$\begin{aligned} \mathrm{d}\mathcal{F} &= \mathrm{d}\mathcal{I} - \mathrm{d}(T\,\mathcal{S}) && (22.54\mathrm{a}) \\ &= \mathrm{d}\mathcal{I} - T\,\mathrm{d}\mathcal{S} - \mathcal{S}\,\mathrm{d}T && (22.54\mathrm{b}) \\ &= -\mathcal{S}\,\mathrm{d}T - p\,\mathrm{d}\nu_{\mathrm{s}} + \mu\,\mathrm{d}C, && (22.54\mathrm{c}) \end{aligned}$$

where we used the fundamental thermodynamic relation (22.33) 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 Helmholz free energy has the functional dependence

$$\mathcal{F} = \mathcal{F}(T, \nu_{\mathrm{s}}, C), \tag{22.55}$$

which is yet another fundamental equation of state whose partial derivatives render

$$\left[\frac{\partial \mathcal{F}}{\partial T}\right]_{\nu_s,C} = -\mathcal{S} \qquad \left[\frac{\partial \mathcal{F}}{\partial \nu_s}\right]_{T,C} = -p \qquad \left[\frac{\partial \mathcal{F}}{\partial C}\right]_{T,\nu_s} = \mu. \tag{22.56}$$

As for internal energy in Section 22.4.1, Equations (22.56) provide a relation between $\mathcal{S}, p, \mu$ as derivatives of a function, the Helmholz free energy, which is itself a function $\mathcal{F}(T, \nu_s, C)$. Hence, we may consider $\mathcal{S}, p, \mu$ each as a function of (T, ν_s, C)

$$\mathcal{S} = \mathcal{S}(T, \nu_s, C) \qquad p = p(T, \nu_s, C) \qquad \mu = \mu(T, \nu_s, C). \tag{22.57}$$

22.4.5 Gibbs potential

The Gibbs potential is defined by the Legendre transformation

$$\mathcal{G} = \mathcal{I} + p\,\nu_s - T\,\mathcal{S} = \mathcal{H} - T\,\mathcal{S} = \sum_{n=1}^{N} \mu_n\, C_n, \tag{22.58}$$

where the final equality made use of equation (22.37). The exact differential of the Gibbs potential is given by

$$\begin{aligned} \mathrm{d}\mathcal{G} &= \mathrm{d}\mathcal{H} - \mathrm{d}(T\,\mathcal{S}) && (22.59\text{a}) \\ &= T\,\mathrm{d}\mathcal{S} + \nu_s\,\mathrm{d}p + \mu\,\mathrm{d}C - T\,\mathrm{d}\mathcal{S} - \mathcal{S}\,\mathrm{d}T && (22.59\text{b}) \\ &= -\mathcal{S}\,\mathrm{d}T + \nu_s\,\mathrm{d}p + \mu\,\mathrm{d}C, && (22.59\text{c}) \end{aligned}$$

where we made use of the fundamental thermodynamic relation (22.49c) written in terms of enthalpy. The Gibbs potential has the functional dependence

$$\mathcal{G} = \mathcal{G}(T, p, C), \tag{22.60}$$

which is yet another fundamental equation of state whose partial derivatives render

$$\left[\frac{\partial \mathcal{G}}{\partial T}\right]_{p,C} = -\mathcal{S} \qquad \left[\frac{\partial \mathcal{G}}{\partial p}\right]_{T,C} = \nu_s \qquad \left[\frac{\partial \mathcal{G}}{\partial C}\right]_{T,p} = \mu. \tag{22.61}$$

As for internal energy in Section 22.4.1, Equations (22.61) provide a relation between $\mathcal{S}, \nu_s, \mu$ as derivatives of a function, the Gibbs potential, which is itself a function $\mathcal{G}(T, p, C)$. Hence, we may consider $\mathcal{S}, \nu_s, \mu$ each as a function of (T, p, C)

$$\mathcal{S} = \mathcal{S}(T, p, C) \qquad \nu_s = \nu_s(T, p, C) \qquad \mu = \mu(T, p, C). \tag{22.62}$$

The form of the fundamental equation of state (22.60), and the associated functional dependence (22.62), 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 (22.62) provides a convenient means to express $\mathcal{S}$, ν_s, and μ (e.g., see the adiabatic lapse rate discussion in Section 22.7).

Given its convenient functional dependence, the Gibbs potential plays a central role in developing the thermodynamics of seawater as formulated by *Feistel* (1993) and codifed by *IOC et al.* (2010). We thus endeavor to exhibit how quantities (e.g., response functions as in Section 22.5) can be computed based on knowledge of the Gibbs potential and its partial derivatives. For example, use of equation (22.61) renders the expression for the enthalpy

$$\mathcal{H} = \mathcal{G} + T\,\mathcal{S} = \mathcal{G} - T\left[\frac{\partial \mathcal{G}}{\partial T}\right]_{p,C}. \tag{22.63}$$

22.5 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.

22.5.1 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

$$c_{\mathrm{v}} \equiv \frac{1}{M}\left[\frac{đ\mathcal{Q}}{\partial T}\right]_{\nu_{\mathrm{s}},C} \qquad \text{SI units m}^2\ \text{s}^{-2}\ \text{K}^{-1}. \tag{22.64}$$

$$c_{\mathrm{p}} \equiv \frac{1}{M}\left[\frac{đ\mathcal{Q}}{\partial T}\right]_{p,C} \qquad \text{SI units m}^2\ \text{s}^{-2}\ \text{K}^{-1}. \tag{22.65}$$

If heating occurs quasi-statically, we can make use of the equation (22.4) relating heat and entropy, applied here in its specific form $M^{-1}\, đ\mathcal{Q} = T\, d\mathcal{S}$. The result is a state-function form of the specific heat capacities

$$c_{\mathrm{v}} = T\left[\frac{\partial \mathcal{S}}{\partial T}\right]_{\nu_{\mathrm{s}},C} = -T\left.\frac{\partial}{\partial T}\right]_{\nu_{\mathrm{s}},C}\left[\frac{\partial \mathcal{G}}{\partial T}\right]_{p,C} \tag{22.66}$$

$$c_{\mathrm{p}} = T\left[\frac{\partial \mathcal{S}}{\partial T}\right]_{p,C} = -T\left.\frac{\partial}{\partial T}\right]_{p,C}\left[\frac{\partial \mathcal{G}}{\partial T}\right]_{p,C} \tag{22.67}$$

where the second equalities in both of the above equations introduced the Gibbs potential according to equation (22.61). Furthermore, we can make use of the fundamental thermodynamic relation (22.31) with specific volume and matter concentration held fixed to yield

$$c_{\mathrm{v}} = T\left[\frac{\partial \mathcal{S}}{\partial T}\right]_{\nu_{\mathrm{s}},C} = \left[\frac{\partial \mathcal{I}}{\partial T}\right]_{\nu_{\mathrm{s}},C}. \tag{22.68}$$

Likewise, making use of the fundamental thermodynamic relation (22.49c) written in terms of enthalpy leads to the constant pressure heat capacity

$$c_{\mathrm{p}} = T\left[\frac{\partial \mathcal{S}}{\partial T}\right]_{p,C} = \left[\frac{\partial \mathcal{I}}{\partial T}\right]_{p,C} + p\left[\frac{\partial \nu_{\mathrm{s}}}{\partial T}\right]_{p,C} = \left[\frac{\partial \mathcal{H}}{\partial T}\right]_{p,C}. \tag{22.69}$$

22.5.2 Thermal expansion, haline contraction, and sound speed

The thermal expansion coefficient measures relative changes in density as temperature changes at constant pressure and concentration

$$\alpha_{\mathrm{T}} = -\frac{1}{\rho}\left[\frac{\partial \rho}{\partial T}\right]_{p,C} = \frac{1}{\nu_{\mathrm{s}}}\left[\frac{\partial \nu_{\mathrm{s}}}{\partial T}\right]_{p,C} = \frac{1}{(\partial \mathcal{G}/\partial p)_{T,C}}\left.\frac{\partial}{\partial T}\right]_{p,C}\left[\frac{\partial \mathcal{G}}{\partial p}\right]_{T,C} \tag{22.70}$$

where the final equality introduced the Gibbs function according to equation (22.61). The minus sign in the definition is introduced since density typically reduces when temperature increases,

so that for most substances $\alpha_{\mathrm{T}} > 0$. Freshwater near its freezing point is an important counter-example, with $\alpha_{\mathrm{T}} < 0$ allowing for solid ice to float on liquid water.

A similar response function measures changes to density arising from changes in the salt concentration (salinity) in seawater

$$\beta_{\mathrm{S}} = \frac{1}{\rho}\left[\frac{\partial \rho}{\partial S}\right]_{p,T} = -\frac{1}{\nu_{\mathrm{s}}}\left[\frac{\partial \nu_{\mathrm{s}}}{\partial S}\right]_{p,T} = -\frac{1}{(\partial \mathcal{G}/\partial p)_{T,S}}\left.\frac{\partial}{\partial S}\right]_{T,p}\left[\frac{\partial \mathcal{G}}{\partial p}\right]_{T,S} \tag{22.71}$$

where $S = 1000\, C$ is the salinity (22.36). Seawater density typically increases (fluid element volume contracts) when salinity increases, so that $\beta_{\mathrm{S}} > 0$.

Finally, changes in density with respect to pressure define the squared sound speed

$$\frac{1}{c_{\mathrm{s}}^2} = \left[\frac{\partial \rho}{\partial p}\right]_{T,S} = -\frac{1}{(\nu_{\mathrm{s}})^2}\left[\frac{\partial \nu_{\mathrm{s}}}{\partial p}\right]_{T,S} = -\frac{1}{[(\partial \mathcal{G}/\partial p)_{T,S}]^2}\left.\frac{\partial}{\partial p}\right]_{T,S}\left[\frac{\partial \mathcal{G}}{\partial p}\right]_{T,S}. \tag{22.72}$$

The sound speed is a strong function of pressure, generally increasing with higher pressure, as well as temperature, generally decreasing with lower temperature. 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 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.

22.6 Maxwell relations for homogeneous fluids

Thermodynamics makes use of basic properties of total (exact) differentials for the purpose of developing identities between partial derivatives. Maxwell relations provide one a particularly useful tool that is based on 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 22.7 for expressing the adiabatic lapse rate in terms of readily measurable thermomechanical properties.

22.6.1 Maxwell relation from internal energy

As seen from Section 22.4.1, the natural functional dependence for internal energy in a homogeneous fluid is given by its fundamental thermodynamic relation (22.38)

$$\mathrm{d}\mathcal{I} = \left[\frac{\partial \mathcal{I}}{\partial \mathcal{S}}\right]_{\nu_{\mathrm{s}}} \mathrm{d}\mathcal{S} + \left[\frac{\partial \mathcal{I}}{\partial \nu_{\mathrm{s}}}\right]_{\mathcal{S}} \mathrm{d}\nu_{\mathrm{s}} = T\,\mathrm{d}\mathcal{S} - p\,\mathrm{d}\nu_{\mathrm{s}} \Longrightarrow \mathcal{I} = \mathcal{I}(\mathcal{S}, \nu_{\mathrm{s}}). \tag{22.73}$$

The mixed second partial derivatives are equal

$$\left.\frac{\partial}{\partial \nu_{\mathrm{s}}}\right]_{\mathcal{S},C}\left.\frac{\partial}{\partial \mathcal{S}}\right]_{\nu_{\mathrm{s}},C}\mathcal{I} = \left.\frac{\partial}{\partial \mathcal{S}}\right]_{\nu_{\mathrm{s}},C}\left.\frac{\partial}{\partial \nu_{\mathrm{s}}}\right]_{\mathcal{S},C}\mathcal{I}, \tag{22.74}$$

so that, via the fundamental thermodynamic relation (22.73), we have the Maxwell relation

$$\left[\frac{\partial T}{\partial \nu_{\mathrm{s}}}\right]_{\mathcal{S}} = -\left[\frac{\partial p}{\partial \mathcal{S}}\right]_{\nu_{\mathrm{s}}}. \tag{22.75}$$

22.6.2 Summary of the Maxwell relations

The other thermodynamic potentials, and their associated fundamental thermodynamical relations, lead to further Maxwell relations as summarized here

$$\mathrm{d}\mathcal{I} = T\,\mathrm{d}\mathcal{S} - p\,\mathrm{d}\nu_{\mathrm{s}} \Longrightarrow \left[\frac{\partial T}{\partial \nu_{\mathrm{s}}}\right]_{\mathcal{S}} = -\left[\frac{\partial p}{\partial \mathcal{S}}\right]_{\nu_{\mathrm{s}}} \tag{22.76}$$

$$\mathrm{d}\mathcal{H} = T\,\mathrm{d}\mathcal{S} + \nu_{\mathrm{s}}\,\mathrm{d}p \Longrightarrow \left[\frac{\partial T}{\partial p}\right]_{\mathcal{S}} = \left[\frac{\partial \nu_{\mathrm{s}}}{\partial \mathcal{S}}\right]_{p} \tag{22.77}$$

$$\mathrm{d}\mathcal{G} = -\mathcal{S}\,\mathrm{d}T + \nu_{\mathrm{s}}\,\mathrm{d}p \Longrightarrow \left[\frac{\partial \mathcal{S}}{\partial p}\right]_{T} = -\left[\frac{\partial \nu_{\mathrm{s}}}{\partial T}\right]_{p}. \tag{22.78}$$

$$\mathrm{d}\mathcal{F} = -\mathcal{S}\,\mathrm{d}T - p\mathrm{d}\nu_{\mathrm{s}} \Longrightarrow \left[\frac{\partial \mathcal{S}}{\partial \nu_{\mathrm{s}}}\right]_{T} = \left[\frac{\partial p}{\partial T}\right]_{\nu_{\mathrm{s}}}. \tag{22.79}$$

These four Maxwell relations for homogeneous fluids involve permutations on cross derivatives of $(T, \mathcal{S})$ and (p, ν_{s}). In statistical mechanics, $(T, \mathcal{S})$ 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.

22.7 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 in from 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.

22.7.1 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 affects 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

$$\hat{\Gamma} \equiv \left[\frac{\partial T}{\partial p}\right]_{C,\mathcal{S}}, \tag{22.80}$$

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.

22.7.2 Thermodynamic formulation

When the matter concentration is held fixed, equation (22.62) allows us to consider entropy as a function of temperature and pressure so that

$$\mathrm{d}\mathcal{S} = \left[\frac{\partial \mathcal{S}}{\partial T}\right]_p \mathrm{d}T + \left[\frac{\partial \mathcal{S}}{\partial p}\right]_T \mathrm{d}p. \tag{22.81}$$

Substituting the definition of heat capacity from equation (22.67) leads to

$$T\,\mathrm{d}\mathcal{S} = c_{\mathrm{p}}\,\mathrm{d}T + T\left[\frac{\partial \mathcal{S}}{\partial p}\right]_T \mathrm{d}p. \tag{22.82}$$

It is useful to eliminate $(\partial \mathcal{S}/\partial p)_T$ in favor of a more easily measurable quantity. For that purpose we make use of the Maxwell relation (22.78) to write

$$\left[\frac{\partial \mathcal{S}}{\partial p}\right]_T = -\left[\frac{\partial \nu_{\mathrm{s}}}{\partial T}\right]_p. \tag{22.83}$$

Introducing the thermal expansion coefficient (22.70) yields an expression for changes in entropy in terms of changes in temperature and pressure

$$T\,\mathrm{d}\mathcal{S} = c_{\mathrm{p}}\,\mathrm{d}T - T\left[\frac{\partial \nu_{\mathrm{s}}}{\partial T}\right]_p \mathrm{d}p = c_{\mathrm{p}}\,\mathrm{d}T - \left[\frac{T\,\alpha_{\mathrm{T}}}{\rho}\right]\mathrm{d}p. \tag{22.84}$$

Since c_{p} and α_{T} are readily measurable response functions, the expression (22.84) is a useful means to compute infinitesimal entropy changes when matter concentration is held constant.

22.7.3 Adiabatic lapse rate for pressure changes

Equation (22.84) means that the change in temperature associated with motion through a pressure field, with $\mathrm{d}\mathcal{S} = 0$ and $\mathrm{d}C = 0$, can be written

$$\hat{\Gamma} = \left[\frac{\partial T}{\partial p}\right]_{C,\mathcal{S}} = \frac{T\,\alpha_{\mathrm{T}}}{\rho\,c_{\mathrm{p}}}. \tag{22.85}$$

Temperature indeed changes when pressure changes, even though there has been no heat exchanged with the environment.

22.7.4 Adiabatic lapse rate for height changes

A static fluid in a gravity field is in hydrostatic balance, whereby the pressure at a point equals to the weight per area above that point (Section 24.3). Hydrostatic balance in a constant gravity field maintains the following relation between the pressure differential increment and the vertical differential increment

$$\mathrm{d}p = -g\,\rho\,\mathrm{d}z. \tag{22.86}$$

Use of the chain rule within the lapse rate expression (22.85) leads to

$$\Gamma = \left[\frac{\partial T}{\partial z}\right]_{C,\mathcal{S}} \tag{22.87a}$$

$$= \left[\frac{\partial T}{\partial p}\right]_{C,\mathcal{S}} \left[\frac{\partial p}{\partial z}\right] \tag{22.87b}$$

$$= -\rho\, g \left[\frac{T\,\alpha_{\mathrm{T}}}{\rho\, c_{\mathrm{p}}}\right] \tag{22.87c}$$

$$= -\left[\frac{g\,T\,\alpha_{\mathrm{T}}}{c_{\mathrm{p}}}\right]. \tag{22.87d}$$

This form for the lapse rates measures the change in temperature (the *lapse*) within a fluid element as it is isentropically moved vertically through a hydrostatic pressure field.

22.7.5 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.

22.8 A simple ideal gas atmosphere

A dry atmosphere is well approximated by an ideal gas, so it is useful to develop various thermodynamic relations for an ideal gas. In an ideal gas we ignore intermolecular forces between molecules. Also, the molecules in an ideal gas are assumed to occupy zero volume. So the internal energy of an ideal gas is just due to translation, rotation, and vibration of molecules. We refer to a *simple ideal gas*, in which the internal energy is a linear function of temperature. A simple ideal gas offers a remarkably accurate basis for studying the thermodynamics of a dry atmosphere. It also provides with a nice means to illustrate special cases of the thermodynamic relations developed thus far.

22.8.1 Equation of state

An ideal gas satisfies the following equation (see Section 11.2.2)

$$P\,V = n\,R_{\mathrm{g}}\,T, \tag{22.88}$$

where p is the pressure, V is the volume, n is the number of moles,

$$R_{\mathrm{g}} = 8.314 \text{ J mole}^{-1} \text{ K}^{-1} = 8.314 \text{ kg m}^2 \text{ s}^{-2} \text{ mole}^{-1} \text{ K}^{-1} \tag{22.89}$$

is the *universal gas constant*, and T is the absolute temperature (see Section 11.2.2). The number of moles equals to the mass, M, of the gas divided by the mass per mole, M_{mole}

$$n = \frac{M}{M_{\mathrm{mole}}}. \tag{22.90}$$

The mass density, $\rho = M/V$, is thus given by

$$\rho = \frac{p\, M_{\text{mole}}}{T\, R_{\text{g}}} \equiv \frac{p}{T\, R^{\text{M}}}, \tag{22.91}$$

where

$$R^{\text{M}} = \frac{R_{\text{g}}}{M_{\text{mole}}} \qquad \text{SI units m}^2\ \text{s}^{-2}\ \text{K}^{-1} \tag{22.92}$$

is the *specific gas constant* as defined by the universal gas constant normalized by the molar mass for the constituent. The relation (22.91) is known as a *thermal equation of state*, or more succinctly just an equation of state (see Section 21.3 for more discussion). It shows that the mass density of an ideal gas is directly proportional to the pressure: increasing pressure in turn 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. However, a 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.

22.8.2 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 (i.e., kinetic energy of the elastic point molecules)

$$\mathcal{I} = \mathcal{I}(T) \qquad \text{ideal gas.} \tag{22.93}$$

Consequently, the exact differential of internal energy for an ideal gas is

$$\mathrm{d}\mathcal{I} = c_{\text{v}}\, \mathrm{d}T. \tag{22.94}$$

The appearance of c_{v}, the constant volume specific heat capacity discussed in Section 22.5.1, arises in order for the ideal gas internal energy to satisfy the general equation (22.68). 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

$$\mathcal{I} = c_{\text{v}}\, T + \text{constant} \qquad \text{simple ideal gas.} \tag{22.95}$$

The arbitrary constant of integration is generally set to zero so that the internal energy vanishes at absolute zero.

22.8.3 Heat capacity from statistical mechanics

Recall again that for an ideal gas, the internal energy is just a function of temperature, $\mathcal{I} = c_{\text{v}}\, T$. A *simple* ideal gas is an ideal gas for which the heat capacity is a constant (equation 22.95). Results from statistical mechanics (outside our scope) show that the thermal/internal energy per molecule equals to $k_{\text{B}}\, T/2$ per excited molecular degree of freedom, where

$$k_{\text{B}} = 1.3806 \times 10^{-23}\ \text{m}^2\ \text{kg}\ \text{s}^{-2}\ \text{K}^{-1} \tag{22.96}$$

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,[1] so that $\mathcal{I}_{\text{molecule}} = 5\, k_{\text{B}}\, T/2$.

We convert this energy per molecule to an energy per mole of diatomic molecules by multiplying by Avogadro's number (equation (11.10))

$$\mathcal{I}_{\text{mole diatomic}} = \frac{5\, A_v\, k_{\text{B}}\, T}{2} = \frac{5\, R_{\text{g}}\, T}{2}, \tag{22.97}$$

where the gas constant is given by

$$R_{\text{g}} = A_v\, k_{\text{B}} \tag{22.98a}$$
$$= \left(6.022 \times 10^{23}\ \text{mole}^{-1}\right)\left(1.3806 \times 10^{-23}\ \text{m}^2\ \text{kg s}^{-2}\ \text{K}^{-1}\right) \tag{22.98b}$$
$$= 8.314\ \text{kg m}^2\ \text{s}^{-2}\ \text{mole}^{-1}\ \text{K}^{-1}. \tag{22.98c}$$

Finally, dividing by the molar mass for dry air (equation (11.11))

$$M_{\text{air}} = 0.028\ \text{kg mole}^{-1} \tag{22.99}$$

leads to the *simple ideal gas* approximation to the dry air heat capacity

$$c_{\text{v}} = \frac{5\, R_{\text{g}}}{2\, M_{\text{air}}} = 742\ \text{m}^2\ \text{s}^{-2}\ \text{K}^{-1}. \tag{22.100}$$

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.

22.8.4 Enthalpy

The enthalpy is generally given by equation (22.48), which for a *simple ideal gas* takes the form

$$\mathcal{H} = \mathcal{I} + p/\rho = c_{\text{v}}\, T + \frac{T\, R_{\text{g}}}{M_{\text{mole}}} = T\left[c_{\text{v}} + R^{\text{M}}\right] \tag{22.101}$$

where $R^{\text{M}} = R_{\text{g}}/M_{\text{mole}}$ (equation (22.92)) is the universal gas constant divided by the molar mass for the gas. Recall that the constant pressure heat capacity is given by equation (22.69)

$$c_{\text{p}} = T\left[\frac{\partial \mathcal{S}}{\partial T}\right]_{p,C} = \left[\frac{\partial \mathcal{H}}{\partial T}\right]_{p,C}. \tag{22.102}$$

Consequently, for a *simple ideal gas* we have

$$c_{\text{p}} = c_{\text{v}} + R^{\text{M}} \qquad \text{and} \qquad \mathcal{H} = c_{\text{p}}\, T. \tag{22.103}$$

22.8.5 Thermal expansion coefficient

The thermal expansion coefficient for an ideal gas is given by

$$\alpha_{\text{T}} = \frac{1}{T}, \tag{22.104}$$

so that as temperature increases the thermal expansion decreases.

[1] At high temperatures, two vibrational degrees of freedom are also excited so that $\mathcal{I}_{\text{molecule}} = 7\, k_{\text{B}}\, T/2$.

22.8.6 Fundamental thermodynamic relations

The fundamental thermodynamic relation in terms of internal energy (equation (22.33)) and enthalpy (equation (22.48)) are given by

$$\mathrm{d}\mathcal{I} = T\,\mathrm{d}\mathcal{S} - p\,\mathrm{d}\nu_{\mathrm{s}} + \mu\,\mathrm{d}C \tag{22.105}$$

$$\mathrm{d}\mathcal{H} = T\,\mathrm{d}\mathcal{S} + \nu_{\mathrm{s}}\,\mathrm{d}p + \mu\,\mathrm{d}C. \tag{22.106}$$

For a simple ideal gas these relations take the form

$$c_{\mathrm{v}}\,\mathrm{d}T = T\,\mathrm{d}\mathcal{S} - p\,\mathrm{d}\nu_{\mathrm{s}} + \mu\,\mathrm{d}C \tag{22.107}$$

$$c_{\mathrm{p}}\,\mathrm{d}T = T\,\mathrm{d}\mathcal{S} + \nu_{\mathrm{s}}\,\mathrm{d}p + \mu\,\mathrm{d}C. \tag{22.108}$$

22.8.7 Adiabatic lapse rate

For an ideal gas, the thermal expansion coefficient is given by (equation (22.104)) $\alpha_{\mathrm{T}} = T^{-1}$ so that the lapse rates are

$$\hat{\Gamma} = \frac{1}{\rho\, c_{\mathrm{p}}} \qquad \text{and} \qquad \Gamma = -\frac{g}{c_{\mathrm{p}}}. \tag{22.109}$$

The measured specific heat capacity for a dry atmosphere at standard temperature (300 K) is

$$c_{\mathrm{p}} = 1005\ \mathrm{m}^2\ \mathrm{s}^{-2}\ \mathrm{K}^{-1} \tag{22.110}$$

so that the adiabatic lapse rate for a dry atmosphere is roughly

$$\Gamma_d = -9.8\ \mathrm{K}/(1000\ \mathrm{m}). \tag{22.111}$$

Hence, temperature decreases by nearly 10 K when rising 1000 m in a dry and ideal gas atmosphere.

22.9 Potential temperature

Heating and cooling of the ocean predominantly occur 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. Oceanographers thus prefer to mark or label water masses using properties that maintain their 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 15.1). However, it is *not* a property of the *in situ* temperature, T, which changes even in the absence of mixing due to pressure effects (see the adiabatic lapse rate discussion in Section 22.7).

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 23.5, we find that an even better measure of heat transport is provided by potential Conservative Temperature, Θ. Nonetheless, potential temperature remains a common field whose use remains ubiquitous in atmospheric and oceanic studies.

22.9.1 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 22.7)

$$\mathrm{d}T = \hat{\Gamma}\,\mathrm{d}p. \tag{22.112}$$

Consequently, *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 reason to introduce potential temperature.

Figure 22.1: *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.*

22.9.2 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 to a reference pressure p_{R}, with the reference pressure typically taken at the ocean/land surface (see Figure 22.1).

Mathematically, the potential temperature, θ, is the reference temperature obtained via integration of $\mathrm{d}T = \hat{\Gamma}\,\mathrm{d}p$ for an isentropic *in situ* temperature change with respect to pressure

$$\int_{\theta}^{T} \mathrm{d}T' = \int_{p_{\mathrm{R}}}^{p} \hat{\Gamma}(T, p', C)\,\mathrm{d}p' \Longrightarrow T = \theta(T, p_{\mathrm{R}}, C) + \int_{p_{\mathrm{R}}}^{p} \hat{\Gamma}(T, p', C)\,\mathrm{d}p', \tag{22.113}$$

with $\hat{\Gamma}$ the lapse rate defined in terms of pressure changes (equation (22.85)). By definition, the *in situ* temperature T equals the potential temperature θ at the reference pressure $p = p_{\mathrm{R}}$. Elsewhere, they differ by an amount determined by the adiabatic lapse rate. Furthermore, we see that

$$\left[\frac{\partial T}{\partial p}\right]_{C,\mathcal{S}} = \left[\frac{\partial \theta}{\partial p}\right]_{C,\mathcal{S}} + \hat{\Gamma}. \tag{22.114}$$

However, by definition

$$\left[\frac{\partial T}{\partial p}\right]_{C,\mathcal{S}} = \hat{\Gamma}, \tag{22.115}$$

so that

$$\left[\frac{\partial \theta}{\partial p}\right]_{C,\mathcal{S}} = 0. \tag{22.116}$$

That is, by definition, the potential temperature has a parametric dependence on the reference pressure (is depends explicity on the concentration, C, and *in situ* temperature, T) yet it has no explicit dependence on the *in situ* pressure when holding concentration and entropy fixed. Finally, note that potential temperature is a function of concentration, C. Hence, the potential temperature generally changes if the material concentration changes. For example, potential temperature in the ocean changes if the salinity changes.

22.9.3 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 (22.62)

$$\mathcal{S} = \mathcal{S}(T, p, C) \tag{22.117}$$

leads to the defining identity for potential temperature

$$\mathcal{S}(T, p, C) = \mathcal{S}(\theta, p_{\mathrm{R}}, C). \tag{22.118}$$

This relation directly connects changes in entropy to changes in potential temperature

$$\mathrm{d}\mathcal{S} = \left[\frac{\partial \mathcal{S}(\theta, p_{\mathrm{R}}, C)}{\partial \theta}\right]_C \mathrm{d}\theta. \tag{22.119}$$

Consequently, the reversible transport of a fluid element with constant matter concentration ($\mathrm{d}C = 0$) occurs with both a constant entropy and constant potential temperature.

We can go even further than the relation (22.119) by recalling that equation (22.84) relates increments in specific entropy to temperature and pressure

$$T\,\mathrm{d}\mathcal{S} = c_{\mathrm{p}}\,(\mathrm{d}T - \hat{\Gamma}\,\mathrm{d}p), \tag{22.120}$$

where $\hat{\Gamma}$ is the adiabatic lapse rate defined in terms of pressure changes (equation (22.85)), and we set $\mathrm{d}C = 0$. Taking the infinitesimal increment (i.e., the differential) of the potential temperature (22.113) leads to

$$\mathrm{d}T = \mathrm{d}\theta + \hat{\Gamma}(T, p, C)\,\mathrm{d}p + \int_{p_{\mathrm{R}}}^{p} [\mathrm{d}\hat{\Gamma}(T, p', C)]\,\mathrm{d}p'. \tag{22.121}$$

Evaluate this increment at the reference pressure, $p = p_{\mathrm{R}}$, so that the integral vanishes, thus leaving

$$\mathrm{d}\theta = \mathrm{d}T - \hat{\Gamma}(T, p_{\mathrm{R}}, C)\,\mathrm{d}p. \tag{22.122}$$

We make use of this relation in equation (22.120) to render an expression for the entropy increment in terms of the potential temperature increment

$$\mathrm{d}\mathcal{S} = c_{\mathrm{p}}\,\theta^{-1}\,\mathrm{d}\theta \qquad p = p_{\mathrm{R}} \text{ and } \mathrm{d}C = 0. \tag{22.123}$$

As part of exercise 23.2, we show that this relation holds for an ideal gas at all pressures, and as part of exercise 23.3, we see that the relation also holds for all pressures in certain liquids.

22.9.4 Potential temperature for an ideal gas

The fundamental thermodynamic relation for a *simple ideal gas* (22.108) takes on the following form for an isentropic change

$$c_{\mathrm{p}}\,\mathrm{d}T = \nu_{\mathrm{s}}\,\mathrm{d}p. \tag{22.124}$$

Dividing both sides by temperature and using the ideal gas relation

$$\frac{\nu_{\mathrm{s}}}{T} = \frac{R^{\mathrm{M}}}{p} \tag{22.125}$$

leads to

$$c_{\mathrm{p}}\,\mathrm{d}(\ln T) = R^{\mathrm{M}}\,\mathrm{d}(\ln p). \tag{22.126}$$

Since c_{p} and R^{M} are constants, we can integrate this relation from the reference pressure to an arbitrary pressure

$$c_{\mathrm{p}}\int_{\theta}^{T}\mathrm{d}(\ln T) = R^{\mathrm{M}}\int_{p_{\mathrm{R}}}^{p}\mathrm{d}(\ln p), \tag{22.127}$$

which renders the explicit expression for the simple ideal gas potential temperature

$$\theta = T\left[\frac{p_{\mathrm{R}}}{p}\right]^{R^{\mathrm{M}}/c_{\mathrm{p}}} \qquad \text{where} \qquad c_{\mathrm{p}} = \frac{7\,R^{\mathrm{M}}}{2}, \tag{22.128}$$

with c_{p} the constant pressure heat capacity of a simple ideal gas of diatomic molecules (Section 22.8.3). In Exercise 22.4 we show that $\partial\theta/\partial p = 0$ for the ideal gas, thus exemplifying the removal of explicit pressure effects from the potential temperature.

22.10 Exercises

EXERCISE 22.1: DERIVATION OF THE GIBBS-DUHEM RELATION
Show all of the steps leading to the Gibbs-Duhem relation (22.22).

EXERCISE 22.2: 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

$$p\,\nu_{\mathrm{s}}^{c_{\mathrm{p}}/c_{\mathrm{v}}} = \text{constant}, \tag{22.129}$$

where $\nu_{\mathrm{s}} = \rho^{-1}$ is the specific volume.

EXERCISE 22.3: GEOPOTENTIAL HEIGHT
The *geopotential height* is the height above the earth of a chosen pressure surface.

(a) Show that an ideal gas atmosphere in hydrostatic balance with a uniform lapse rate

$$\frac{\partial T}{\partial z} = -|\Gamma| = \text{constant} \tag{22.130}$$

has a geopotential height at a pressure p given by

$$z = \frac{T_0}{|\Gamma|}\left[1 - \left[\frac{p_0}{p}\right]^{-R^{\mathrm{M}}\,|\Gamma|/g}\right], \tag{22.131}$$

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 (22.131) in the appropriate limit.

EXERCISE 22.4: POTENTIAL TEMPERATURE FOR AN IDEAL GAS

Show that $\partial\theta/\partial p = 0$ for the potential temperature of an ideal gas given by equation (22.128)

$$\theta = T\left[\frac{p_{\mathrm{R}}}{p}\right]^{R^{\mathrm{M}}/c_{\mathrm{p}}}. \tag{22.132}$$

Hint: remember that $\partial T/\partial p \neq 0$.

23

Energy dynamics

In this chapter, we develop the dynamical equations for mechanical energy (kinetic plus gravitational potential) and total energy (mechanical plus internal) relevant to a geophysical fluid. For this purpose we extend the equilibrium thermodynamics of Chapter 22 to time dependent phenomena. To make this transition requires us to assume fluid elements to be in *local thermodynamic equilibrium*. This assumption is based on the quasi-static nature of macroscopic motions (Section 22.1.2) whereby microscopic motions have a much much shorter equilibration time as compared to the relatively slow macroscopic processes of interest for fluid mechanics such as advection, waves, turbulence, and turbulent mixing. Hence, we make use of phenomenological thermodynamic laws to develop evolution equations for thermodynamic properties of continuum fluid elements. Consistent with the huge time scale separation between microscopic and macroscopic processes, we can assume that the macroscopic motion of a fluid element does not alter its entropy. That is, advective transport is a reversible process. In contrast, mixing of properties between fluid elements, whether molecular or turbulent, is irreversible and thus increases entropy.

READER'S GUIDE FOR THIS CHAPTER

This chapter builds from the momentum dynamics of Chapter 18 and the thermodynamics of Chapter 22. As part of our development of energetic analysis, we develop the heat equation as manifested by the equation for potential enthalpy. Variations of the heat equation are considered throughout the remainder of this book.

The formalism for making use of thermodynamics in a moving fluid is known as *quasi-equilibrium thermodynamics* or *linear irreversible thermodynamics*, with the fundamentals thoroughly developed and applied in *DeGroot and Mazur* (1984) and *Landau and Lifshitz* (1987). The term "linear" in the name linear irreversible thermodynamics refers to an assumption that the system is close to thermodynamic equilibrium. This assumption is very accurate for atmospheric and oceanic flows, whereby local thermodynamic equilibrium is maintained for fluid elements even as those elements move through imbalances in macroscopic forces. Thermodynamic fluxes are thus linear functions of the gradients of the thermodynamic state variables.

23.1 Loose threads **396**
23.2 Thermodynamics of a moving fluid **396**
23.2.1 Concerning the transition to a moving fluid 397
23.2.2 Thermodynamic variables versus space-time fields 397

23.2.3 Thermodynamic relations for steady state flow . . . 398
23.2.4 Materially constant specific entropy for a perfect fluid . . . 398
23.2.5 First Law for a moving fluid element . . . 398
23.2.6 Enthalpy budget . . . 399
23.2.7 First Law for homogeneous fluids in terms of potential temperature . 400
23.2.8 Further study . . . 400
23.3 Mechanical energy budget . . . **400**
23.3.1 Gravitational potential energy . . . 400
23.3.2 Kinetic energy . . . 401
23.3.3 Kinetic energy dissipation . . . 402
23.3.4 Mechanical energy . . . 403
23.3.5 Further study . . . 403
23.4 Total energy budget . . . **403**
23.4.1 Equations for total energy, internal energy, and enthalpy . . . 403
23.4.2 Joule heating from friction . . . 404
23.4.3 Bernoulli potential and mechanical injection work . . . 405
23.4.4 Bernoulli's theorem for a steady perfect homogeneous fluid . . . 406
23.4.5 Comments on gauge symmetry . . . 407
23.4.6 Further study . . . 408
23.5 Evolution of temperature . . . **408**
23.5.1 Enthalpy equation again . . . 408
23.5.2 Evolution of *in situ* temperature . . . 408
23.5.3 Evolution of potential temperature . . . 409
23.5.4 Evolution of Conservative Temperature . . . 409
23.5.5 Further study . . . 411
23.6 Summary of the ocean-atmosphere equations . . . **411**
23.7 Moist static energy . . . **412**
23.7.1 Further study . . . 413
23.8 Exercises . . . **413**

23.1 Loose threads

- Moist static energy as per David Romps' notes
- Available internal energy
- Potential energy as per *Landau and Lifshitz* (1987) and possible extenstion to APE.

23.2 Thermodynamics of a moving fluid

Recall the fundamental thermodynamic relation (22.33) for a two-component fluid such as seawater

$$\mathrm{d}\mathcal{I} = T\,\mathrm{d}\mathcal{S} - p\,\mathrm{d}\nu + \mu\,\mathrm{d}C. \tag{23.1}$$

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 is 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. Changes in the microscopic degrees of freedom also occur on a tiny time scale compare to those for changes in the macroscopic forces. We discussed this extremely large time scale separation in Section 22.1.2 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.

23.2.1 Concerning the transition to a moving fluid

For a moving continuous fluid, each of the thermodynamic properties in the equilibrium First Law expression (23.1) are continuous functions of space and time. Furthermore, equation (23.1) 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 12.5.1 to write

$$\mathrm{d}\Psi = \Psi(\boldsymbol{x} + \mathrm{d}\boldsymbol{x}, t + \mathrm{d}t) - \Psi(\boldsymbol{x}, t) = \mathrm{d}t\, \partial_t \Psi + \mathrm{d}\boldsymbol{x} \cdot \nabla \Psi, \tag{23.2}$$

where Ψ is one of the thermodynamic properties, $\mathrm{d}t$ is the infinitesimal time increment, and $\mathrm{d}\boldsymbol{x}$ is the vector of infinitesimal space increments. Following the discussion in Section 12.5.2, we are thus led to the total time derivative for a property following a trajectory $\boldsymbol{x} = \boldsymbol{x}(t)$

$$\frac{\mathrm{d}\Psi}{\mathrm{d}t} = \frac{\partial \Psi}{\partial t} + \frac{\mathrm{d}\boldsymbol{x}}{\mathrm{d}t} \cdot \nabla \Psi. \tag{23.3}$$

Restricting the trajectory to that defined by a fluid particle, so that $\boldsymbol{v} = \mathrm{d}\boldsymbol{x}/\mathrm{d}t$, then renders the material time derivative as in Section 12.5.4

$$\frac{\mathrm{D}\Psi}{\mathrm{D}t} = \frac{\partial \Psi}{\partial t} + \boldsymbol{v} \cdot \nabla \Psi. \tag{23.4}$$

We make use of this result in Section 23.2.5 to transition the quasi-static relation (23.1) to a moving fluid.

23.2.2 Thermodynamic variables versus space-time fields

There is room for confusion when allowing thermodynamic variables 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 (23.1), or whether it expresses a relation involving thermodynamic space-time fields. 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 (22.61), which says that the partial derivative of the Gibbs potential with respect to pressure, holding temperature and concentration fixed, equals to the specific volume

$$\left[\frac{\partial \mathcal{G}}{\partial p}\right]_{T,C} = \rho^{-1}. \tag{23.5}$$

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

$$\left[\frac{\partial \mathcal{G}}{\partial p}\right]_{x,y,t} \neq \left[\frac{\partial \mathcal{G}}{\partial p}\right]_{T,C}. \tag{23.6}$$

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 25.4.

23.2.3 Thermodynamic relations for steady state flow

Before considering time evolution, let us establish some steady (stationary) state relations defined for flows that have no time Eulerian dependence. In this case, the exact differential for a thermodynamic potential arises just from spatial variations

$$\mathrm{d}\Psi = \Psi(\boldsymbol{x} + \mathrm{d}\boldsymbol{x}) - \Psi(\boldsymbol{x}) = \mathrm{d}\boldsymbol{x} \cdot \nabla\Psi \qquad \text{steady flows.} \tag{23.7}$$

Applying this result to the fundamental thermodynamic expression (23.1) renders the steady state relation

$$\mathrm{d}\boldsymbol{x} \cdot (\nabla\mathcal{I} - T\,\nabla\mathcal{S} + p\,\nabla\nu - \mu\,\nabla C) = 0. \tag{23.8}$$

Since the spatial increment $\mathrm{d}\boldsymbol{x}$ is arbitrary, the bracketed term generally vanishes

$$\nabla\mathcal{I} = T\,\nabla\mathcal{S} - p\,\nabla\nu + \mu\,\nabla C \qquad \text{steady state.} \tag{23.9}$$

Analogous identities hold for steady state thermodynamic potentials following from the fundamental thermodynamic relations derived in Section 22.4

STEADY STATE FLOW

$$\nabla\mathcal{I} = T\,\nabla\mathcal{S} - p\,\nabla\nu + \mu\,\nabla C \tag{23.10}$$

$$T\,\nabla\mathcal{S} = \nabla\mathcal{I} + p\,\nabla\nu - \mu\,\nabla C \tag{23.11}$$

$$\nabla\mathcal{H} = T\,\nabla\mathcal{S} + \nu\,\nabla p + \mu\,\nabla C \tag{23.12}$$

$$\nabla\mathcal{G} = -\mathcal{S}\,\nabla T + \nu\,\nabla p + \mu\,\nabla C. \tag{23.13}$$

23.2.4 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. Consequently, specific entropy for each fluid parcel is reversibly stirred through advection

$$\frac{\mathrm{D}\mathcal{S}}{\mathrm{D}t} = 0, \tag{23.14}$$

which takes on the steady state form

$$\boldsymbol{v} \cdot \nabla\mathcal{S} = 0. \tag{23.15}$$

Notably, a perfect fluid generally maintains macroscopic gradients of specific entropy even as each fluid element moves without altering its local value of specific entropy.

23.2.5 First Law for a moving fluid element

Section 23.2.1 provides the key operational means to 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 (23.1), which is the First Law for a quasi-static process, takes the form for a moving fluid element

$$\frac{\mathrm{D}\mathcal{I}}{\mathrm{D}t} = T\,\frac{\mathrm{D}\mathcal{S}}{\mathrm{D}t} - p\,\frac{\mathrm{D}\nu}{\mathrm{D}t} + \mu\,\frac{\mathrm{D}C}{\mathrm{D}t}. \tag{23.16}$$

Exposing the Eulerian space and time derivatives renders

$$\partial_t \mathcal{I} - T\,\partial_t \mathcal{S} + p\,\partial_t \nu - \mu\,\partial_t C + \boldsymbol{v}\cdot(\nabla\mathcal{I} - T\,\nabla\mathcal{S} + p\,\nabla\nu - \mu\,\nabla C) = 0. \tag{23.17}$$

As discussed in Section 23.2.3, for a steady state flow all Eulerian time derivatives vanish and the spatial gradients reduce to the steady state relation (23.10). 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 (23.16).

We can massage the result (23.16) to further reveal its connection to the First Law written in the form (23.1). For this purpose, recall that mass conservation as discussed in Section 14.1 means that changes in the volume of a fluid element are related to density changes via

$$\frac{1}{\delta V}\frac{\mathrm{D}\delta V}{\mathrm{D}t} = \frac{1}{\nu}\frac{\mathrm{D}\nu}{\mathrm{D}t}, \tag{23.18}$$

where again $\nu = \rho^{-1}$ is the specific volume. Hence, equation (23.16) can be written

$$\delta M\,\frac{\mathrm{D}\mathcal{I}}{\mathrm{D}t} = T\,\delta M\,\frac{\mathrm{D}\mathcal{S}}{\mathrm{D}t} - p\,\frac{\mathrm{D}\delta V}{\mathrm{D}t} + \mu\,\delta M\,\frac{\mathrm{D}C}{\mathrm{D}t}, \tag{23.19}$$

where $\delta M = \rho\,\delta V$ is the mass of the fluid element. Since the mass of the fluid element is constant, equation (23.19) is the fluid element extension of the First Law given by equation (23.1). Alternatively, we can use the further result from mass conservation (equation (14.23))

$$\frac{1}{\delta V}\frac{\mathrm{D}\delta V}{\mathrm{D}t} = \nabla\cdot\boldsymbol{v} \tag{23.20}$$

to write

$$\frac{\mathrm{D}\mathcal{I}}{\mathrm{D}t} = -p\,\nu\,\nabla\cdot\boldsymbol{v} + T\,\frac{\mathrm{D}\mathcal{S}}{\mathrm{D}t} + \mu\,\frac{\mathrm{D}C}{\mathrm{D}t}. \tag{23.21}$$

Processes affecting internal energy that appear on the right hand side are (i) entropy production, whose form is developed in the following, (ii) mechanical work from pressure modifying the volume of the fluid element, and (iii) mixing (chemical work) through the exchange of matter constituents between fluid elements.

23.2.6 Enthalpy budget

Rather than work with internal energy, it is often more convenient to consider the specific enthalpy (Section 22.4.3),

$$\mathcal{H} = \mathcal{I} + p/\rho = \mathcal{I} + p\,\nu. \tag{23.22}$$

The mass continuity equation (14.8c) and the internal energy equation (23.21) yield

$$\frac{\mathrm{D}\mathcal{H}}{\mathrm{D}t} = \frac{\mathrm{D}\mathcal{I}}{\mathrm{D}t} + \frac{1}{\rho}\frac{\mathrm{D}p}{\mathrm{D}t} - \frac{p}{\rho^2}\frac{\mathrm{D}\rho}{\mathrm{D}t} \Longrightarrow \rho\,\frac{\mathrm{D}\mathcal{H}}{\mathrm{D}t} = \frac{\mathrm{D}p}{\mathrm{D}t} + \rho\left[T\,\frac{\mathrm{D}\mathcal{S}}{\mathrm{D}t} + \mu\,\frac{\mathrm{D}C}{\mathrm{D}t}\right]. \tag{23.23}$$

This equation is quite useful as it means that for constant pressure processes, changes to the specific enthalpy of a fluid element arise just from irreversible processes

$$\frac{\mathrm{D}p}{\mathrm{D}t} = 0 \Longrightarrow \frac{\mathrm{D}\mathcal{H}}{\mathrm{D}t} = T\,\frac{\mathrm{D}\mathcal{S}}{\mathrm{D}t} + \mu\,\frac{\mathrm{D}C}{\mathrm{D}t}. \tag{23.24}$$

Since many boundary processes occur 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.

23.2.7 First Law for homogeneous fluids in terms of potential temperature

Equation (22.123) says that the change in entropy for an element moving with constant matter concentration and at the reference pressure is given in terms of the potential temperature

$$\frac{\mathrm{D}\mathcal{S}}{\mathrm{D}t} = \frac{c_p}{\theta}\frac{\mathrm{D}\theta}{\mathrm{D}t} \tag{23.25}$$

Since the potential temperature equals to the temperature when $p = p_{\mathrm{R}}$, we have

$$c_p \frac{\mathrm{D}\theta}{\mathrm{D}t} = \theta \frac{\mathrm{D}\mathcal{S}}{\mathrm{D}t} = \dot{Q} \qquad \text{at } p = p_{\mathrm{R}} \text{ and } \mathrm{d}C = 0. \tag{23.26}$$

In Exercise 23.2 we consider these relations for an ideal gas.

The *in situ* temperature, T, does not make for a convenient prognostic variable for numerical models of 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 adiabatic lapse rate discussion in Section 22.7), the *in situ* temperature T is modified even when the flow is adiabatic. In contrast, modelers prefer to work with scalar fields that remain materially constant in the absence of mixing. Hence, potential temperature has historically been used for both atmosphere and ocean models rather than *in situ* temperature. That is, in numerical and theoretical models, potential temperature satisfies an advection/diffusion equation analogous to material tracers such as salinity (see Chapter 47). We make use of potential temperature as the primary field measuring the buoyancy of a fluid in various theoretical models (Chapter 21). Finally, we note that further advances (*McDougall*, 2003; *IOC et al.*, 2010) show that Conservative Temperature is an even more suitable thermodynamic tracer for the ocean than potential temperature.

23.2.8 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).

23.3 Mechanical energy budget

Momentum is but one dynamical property of a fluid. Mechanical energy is another and it is 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 mechanical energy of a fluid on a rotating sphere, making use of the momentum equation written in the general form (23.86).

23.3.1 Gravitational potential energy

The gravitational potential energy per mass; i.e., the geopotential, for a fluid element in a constant effective gravitational field is given by

$$\Phi = g\, z \tag{23.27}$$

so that its material evolution is

$$\rho \frac{\mathrm{D}\Phi}{\mathrm{D}t} = w\, g\, \rho. \tag{23.28}$$

As already noted above in our discussion of kinetic energy, we see that just as for the point particle, there is an exchange of mechanical energy between the kinetic energy and potential energy conveyed through vertical motion in the gravitational field. Vertically upward motion ($w > 0$) increases potential energy whereas downward motion ($w < 0$) decreases it. As seen in Section 23.3.2, these potential energy changes are exactly balanced by kinetic energy changes. For a more general geopotential we have the kinematic expression for the material evolution

$$\frac{\mathrm{D}\Phi}{\mathrm{D}t} = \frac{\partial \Phi}{\partial t} + \boldsymbol{v} \cdot \nabla \Phi. \tag{23.29}$$

Local time dependence to the geopotential arises from astronomical tide forcing, with details provided in Chapter 43.

23.3.2 Kinetic energy

Following from our discussion of mechanical energy of a point particle in Section 10.5, we define the kinetic energy per mass as

$$\mathscr{K} = \frac{1}{2}\, \boldsymbol{v} \cdot \boldsymbol{v}. \tag{23.30}$$

Making use of the momentum equation (23.86) reveals that the material evolution of the kinetic energy per mass is given by

$$\rho \frac{\mathrm{D}\mathscr{K}}{\mathrm{D}t} = -\boldsymbol{v} \cdot \nabla p - \rho\, \boldsymbol{v} \cdot \nabla \Phi + \rho\, \boldsymbol{v} \cdot \boldsymbol{F}, \tag{23.31}$$

where we wrote

$$\boldsymbol{F} = \rho^{-1}\, \nabla \cdot \mathbb{T} \tag{23.32}$$

for the friction acceleration. Equation (23.31) says that the kinetic energy of a fluid element is affected by three physical processes as described below.

Contribution from the pressure force

Kinetic energy increases in regions where the velocity projects down the pressure gradient,

$$\boldsymbol{v} \cdot \nabla p < 0 \Longrightarrow \text{increase kinetic energy}, \tag{23.33}$$

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 26.3) with

$$\boldsymbol{v}_{\mathrm{g}} = \frac{\hat{\boldsymbol{z}} \wedge \nabla p}{f \rho} \tag{23.34}$$

have the velocity oriented perpendicular to the horizontal pressure gradient ($\boldsymbol{v}_{\mathrm{g}} \cdot \nabla p = 0$). Consequently, pressure has no impact on the horizontal kinetic energy of a geostrophic fluid.

Contribution from the geopotential

For a simple geopotential with $\Phi = g\,z$, we readily see that kinetic energy of a fluid element decreases in regions where the vertical velocity is positive,

$$w > 0 \Longrightarrow -w\,g\,\rho < 0 \leftrightarrow \text{upward motion decreases } \mathcal{K}. \tag{23.35}$$

Hence, motion against the gravitational field (upward) reduces kinetic energy. As seen in Section 23.3.1, this decrease in kinetic energy due to vertical motion is exactly balanced by an increase in gravitational potential energy: raising fluid elements increases their gravitational potential energy. The interpretation for a general geopotential holds similarly.

Contribution from the friction vector

Kinetic energy is reduced in regions where the velocity has a negative projection onto the direction of the friction vector, $\rho\,\boldsymbol{v}\cdot\boldsymbol{F} < 0$. As detailed in Section 23.3.3, 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.

23.3.3 Kinetic energy dissipation

We here examine the role of the friction term

$$\text{friction power per volume} = \rho\,\boldsymbol{v}\cdot\boldsymbol{F}, \tag{23.36}$$

with this term interpreted as a power (energy per time) per volume acting to alter the kinetic energy per volume of a fluid element. To get started, consider the special case of Rayleigh drag,

$$\boldsymbol{F} = \boldsymbol{F}_{\text{Rayleigh}} = -\gamma\,\boldsymbol{v}, \tag{23.37}$$

in which case the kinetic energy is dissipated locally at each point

$$\rho\,\boldsymbol{v}\cdot\boldsymbol{F}_{\text{Rayleigh}} = -\gamma\,\rho\,\boldsymbol{v}\cdot\boldsymbol{v} \leq 0. \tag{23.38}$$

What happens with a more general stress-rate of strain relation (19.33) applicable to a Newtonian fluid? To answer this question, expose Cartesian tensor labels to render

$$\begin{aligned}
\rho\,\boldsymbol{v}\cdot\boldsymbol{F} &= v_m\,\rho\,F_m && \text{(23.39a)}\\
&= 2\,v_m\,\partial_n(\rho\,\nu\,\mathbb{S}_{mn}) && \text{(23.39b)}\\
&= 2\,\partial_n\,(v_m\,\rho\,\nu\,\mathbb{S}_{mn}) - 2\,\partial_n\,v_m\,\rho\,\nu\,\mathbb{S}_{mn} && \text{(23.39c)}\\
&= 2\,\nabla\cdot(\rho\,\nu\,\boldsymbol{v}\cdot\mathbb{S}) - 2\,\rho\,\nu\,\mathbb{S}\cdot\mathbb{S}, && \text{(23.39d)}
\end{aligned}$$

where

$$\mathbb{S}\cdot\mathbb{S} = \mathbb{S}_{mn}\,\mathbb{S}_{mn} \geq 0 \tag{23.40}$$

is the contraction of the rate-of-strain tensor with itself. We interpret the two contributions in equation (23.39d) as

$$\rho\,\boldsymbol{v}\cdot\boldsymbol{F} = \text{viscous flux divergence} - \text{kinetic energy dissipation}. \tag{23.41}$$

When integrated over the full domain, the flux divergence reduces to a contribution of boundary stresses, which are typically quite turbulent. In contrast, the dissipation term provides a sign-definite sink to the kinetic energy at each point in the fluid interior, with this term commonly written

$$\epsilon \equiv 2\,\nu\,\mathbb{S}\cdot\mathbb{S} \geq 0 \qquad \text{mechanical energy dissipation.} \tag{23.42}$$

The dimensions of ϵ are $\mathrm{L}^2\ \mathrm{T}^{-3}$, which in SI units are $\mathrm{m}^2\ \mathrm{s}^{-3} = \mathrm{W\ kg}^{-1}$. We thus commonly refer to ϵ as the mechanical energy dissipation per mass due to viscous effects; i.e., the dissipated mechanical power per mass.

23.3.4 Mechanical energy

Adding the material time evolution equations for kinetic energy per mass and potential energy per mass leads to the material form of the mechanical energy per mass

$$\rho\,\frac{\mathrm{D}\mathcal{M}}{\mathrm{D}t} = -\boldsymbol{v}\cdot\nabla p + 2\,\nabla\cdot(\rho\,\nu\,\boldsymbol{v}\cdot\mathbb{S}) + \rho\,(-\epsilon + \partial_t\Phi), \tag{23.43}$$

where

$$\mathcal{M} = \mathcal{K} + \Phi \tag{23.44}$$

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. In the absence of pressure and friction, and for a time-independent geopotential, equation (23.43) reduces to the mechanical energy budget for a point particle detailed in Section 10.5. This trivial limiting case is expected since pressure and friction arise from the continuum nature of a fluid and are absent from the dynamics of a point particle.

23.3.5 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.

23.4 Total energy budget

Recall from Section 10.5 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 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 heat (Section 23.4.2). In this section we combine the mechanical energy budget from Section 23.3 to the internal energy budget from Section 23.2, thus rendering the budget for total energy of a fluid element.

23.4.1 Equations for total energy, internal energy, and enthalpy

The total energy per mass of a fluid element is the sum of the internal energy plus the mechanical energy,

$$\mathcal{E} = \mathcal{I} + \mathcal{M} = \mathcal{I} + \mathcal{K} + \Phi. \tag{23.45}$$

Adding the internal energy equation in the form (23.21) to the mechanical energy equation (23.43) leads to the material evolution equation for the total energy

$$\rho \frac{\mathrm{D}\mathcal{E}}{\mathrm{D}t} = -\nabla \cdot (p\,\boldsymbol{v} - 2\,\rho\,\nu\,\boldsymbol{v} \cdot \mathbb{S}) + \rho \left[-\epsilon + T\,\frac{\mathrm{D}\mathcal{S}}{\mathrm{D}t} + \mu\,\frac{\mathrm{D}C}{\mathrm{D}t} \right] + \rho\,\partial_t \Phi. \tag{23.46}$$

We now assume that the tracer concentration and specific entropy satisfy the following equations

$$\rho \frac{\mathrm{D}C}{\mathrm{D}t} = -\nabla \cdot \boldsymbol{J}_{\mathrm{C}} \qquad \text{and} \qquad \rho \frac{\mathrm{D}\mathcal{S}}{\mathrm{D}t} = -\nabla \cdot \boldsymbol{J}_{\mathcal{S}} + \Sigma_{\mathcal{S}}, \tag{23.47}$$

where $\boldsymbol{J}_{\mathrm{C}}$ is the non-advective tracer flux (Section 15.1.4), $\boldsymbol{J}_{\mathcal{S}}$ is the non-advective specific entropy flux, and $\Sigma_{\mathcal{S}}$ is an entropy source. We could readily add a tracer source; however, sources for salt are generally absent for the ocean and so for brevity we do not include one. These expressions thus lead to

$$\rho \left[T\,\frac{\mathrm{D}\mathcal{S}}{\mathrm{D}t} + \mu\,\frac{\mathrm{D}C}{\mathrm{D}t} \right] = -T\,\nabla \cdot \boldsymbol{J}_{\mathcal{S}} + T\,\Sigma_{\mathcal{S}} - \mu\,\nabla \cdot \boldsymbol{J}_{\mathrm{C}} \tag{23.48a}$$

$$= -\nabla \cdot (T\,\boldsymbol{J}_{\mathcal{S}} + \mu\,\boldsymbol{J}_{\mathrm{C}}) + \nabla T \cdot \boldsymbol{J}_{\mathcal{S}} + \nabla \mu \cdot \boldsymbol{J}_{\mathrm{C}} + T\,\Sigma_{\mathcal{S}}, \tag{23.48b}$$

which then brings the total energy equation (23.46) into the form

$$\rho \frac{\mathrm{D}\mathcal{E}}{\mathrm{D}t} = -\nabla \cdot (p\,\boldsymbol{v} - 2\,\rho\,\nu\,\boldsymbol{v} \cdot \mathbb{S} + T\,\boldsymbol{J}_{\mathcal{S}} + \mu\,\boldsymbol{J}_{\mathrm{C}}) + [-\rho\,\epsilon + \nabla T \cdot \boldsymbol{J}_{\mathcal{S}} + \nabla \mu \cdot \boldsymbol{J}_{\mathrm{C}} + T\,\Sigma_{\mathcal{S}}] + \rho\,\partial_t \Phi. \tag{23.49}$$

Next we postulate that the total energy is affected only by boundary terms when integrated over the global fluid domain in the absence of astronomical forces; i.e., when the geopotential is time-independent. 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

$$T\,\Sigma_{\mathcal{S}} = \rho\,\epsilon - \nabla T \cdot \boldsymbol{J}_{\mathcal{S}} - \nabla \mu \cdot \boldsymbol{J}_{\mathrm{C}} \tag{23.50}$$

so that the total energy budget is given by

$$\rho \frac{\mathrm{D}\mathcal{E}}{\mathrm{D}t} = -\nabla \cdot (p\,\boldsymbol{v} - 2\,\rho\,\nu\,\boldsymbol{v} \cdot \mathbb{S} + T\,\boldsymbol{J}_{\mathcal{S}} + \mu\,\boldsymbol{J}_{\mathrm{C}}) + \rho\,\partial_t \Phi. \tag{23.51}$$

The modified form of the internal energy budget (23.21) can be found by subtracting the mechanical energy budget (23.43) from the total energy budget (23.51)

$$\rho \frac{\mathrm{D}\mathcal{I}}{\mathrm{D}t} = \rho \frac{\mathrm{D}(\mathcal{E} - \mathcal{M})}{\mathrm{D}t} = -p\,\nabla \cdot \boldsymbol{v} - \nabla \cdot (T\,\boldsymbol{J}_{\mathcal{S}} + \mu\,\boldsymbol{J}_{\mathrm{C}}) + \rho\,\epsilon. \tag{23.52}$$

The corresponding enthalpy budget (23.23) is given by

$$\rho \frac{\mathrm{D}\mathcal{H}}{\mathrm{D}t} = \frac{\mathrm{D}p}{\mathrm{D}t} - \nabla \cdot (T\,\boldsymbol{J}_{\mathcal{S}} + \mu\,\boldsymbol{J}_{\mathrm{C}}) + \rho\,\epsilon. \tag{23.53}$$

23.4.2 Joule heating from friction

The frictional dissipation term, $\epsilon > 0$, represents a conversion of kinetic energy into heat

$$\dot{Q}_{\mathrm{Joule}} \equiv \epsilon. \tag{23.54}$$

The Joule heating 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 $\epsilon \approx 10^{-9}$ W kg^{-1}. Dividing by $c_p = 3900$ J kg^{-1} K^{-1} leads to a heating rate of less than 10^{-3} 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.[1]

23.4.3 Bernoulli potential and mechanical injection work

Equation (23.51) 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 $p\,\boldsymbol{v}$ is fundamental to energy within the continuum. As shown here, it is the pressure work required for the fluid element to mechanically exist within the continuum; i.e., it is the mechanical *injection work*.

For that purpose, convert the material budget (23.51) into its Eulerian form

$$\frac{\partial(\rho\,\mathcal{E})}{\partial t} + \nabla\cdot[\rho\,\boldsymbol{v}\,(\mathcal{E} + p/\rho) - 2\,\rho\,\nu\,\boldsymbol{v}\cdot\mathbb{S} + T\,\boldsymbol{J}_{\mathbb{S}} + \mu\,\boldsymbol{J}_{\mathrm{C}}] = \rho\,\partial_t\Phi. \tag{23.55}$$

The left hand side indicates that total energy of a fluid element is locally modified by the advective transport of the quantity

$$\mathcal{E} + p/\rho = \mathcal{K} + \Phi + \mathcal{I} + p\,\nu \tag{23.56a}$$
$$= \mathcal{K} + \Phi + \mathcal{H} \tag{23.56b}$$
$$= \mathcal{M} + \mathcal{H} \tag{23.56c}$$
$$\equiv \mathcal{B}, \tag{23.56d}$$

where we introduced the Bernoulli potential, which is the sum of the enthalpy per mass plus the mechanical energy per mass

$$\mathcal{B} = \mathcal{H} + \mathcal{K} + \Phi = \mathcal{H} + \mathcal{M}. \tag{23.57}$$

Integration over a region with zero boundary transfer of $\boldsymbol{v}\,\mathcal{B}$ leads to the conservation of total energy for a perfect fluid in the presence of a time-independent geopotential.

Why is $\rho\,\mathcal{E}$ affected by the convergence of $\rho\,\boldsymbol{v}\,\mathcal{B}$ rather than the convergence of $\rho\,\boldsymbol{v}\,\mathcal{E}$, even for a perfect fluid. To answer this question,[2] again note that the Bernoulli potential is the sum of the total energy per mass of a fluid parcel, $\mathcal{E}$, plus the term $p/\rho = p\,\nu$. So what is $p\,\nu$? Imagine carving out a unit mass from within a continuous fluid with pressure p and specific volume ν, leaving behind a "hole". The mechanical work required to carve out this hole is precisely equal to $p\,\nu$. Correspondingly, we interpret $p\,\nu$ as the mechanical work required to inject a unit mass of fluid with specific volume ν into a region with pressure p. We thus refer to $p\,\nu$ as the *injection work*. Since specific enthalpy is the sum of the specific internal energy plus the injection work (equation (23.22)), we conclude that specific enthalpy provides a measure of the energy (internal energy plus mechanical energy) required for a fluid element to exist within a continuum fluid. Adding this energy to the mechanical energy yields the Bernoulli potential.

[1] SMG: discuss the atmospheric Joule heating and why it is not negligible. Provide some numbers.

[2] This argument follows Section 13.5.4 of *Thorne and Blandford* (2017) as well as Section 6 of *Landau and Lifshitz* (1987).

We support the above interpretation by considering the flux of total energy in a perfect fluid that penetrates a static closed surface

$$\oint_{\mathcal{S}} \rho\, \boldsymbol{v}\, \mathcal{B} \cdot \hat{\boldsymbol{n}}\, \mathrm{d}\mathcal{S} = \oint_{\mathcal{S}} \rho\, \boldsymbol{v}\, \mathcal{E} \cdot \hat{\boldsymbol{n}}\, \mathrm{d}\mathcal{S} + \oint_{\mathcal{S}} p\, \boldsymbol{v} \cdot \hat{\boldsymbol{n}}\, \mathrm{d}\mathcal{S} \tag{23.58}$$

The first term is the flux of total energy penetrating the surface, whereas the second term is the mechanical work done by pressure acting on the surface.

23.4.4 Bernoulli's theorem for a steady perfect homogeneous fluid

Consider a perfect fluid flow in steady state (vanishing Eulerian time derivatives). Steady state mass conservation means that

$$\frac{\partial \rho}{\partial t} = -\nabla \cdot (\rho\, \boldsymbol{v}) = 0. \tag{23.59}$$

This relation, along with a steady state energy in equation (23.55) (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

$$\boldsymbol{v} \cdot \nabla \mathcal{B} = 0. \tag{23.60}$$

We thus see that for the perfect fluid to be in a steady state, the Bernoulli potential is constant along streamlines. Hence, as the fluid moves along a streamline, there is an exchange between the total energy per mass, $\mathcal{E}$, and the injection work, $p\,\nu$, 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 $p + \rho\, \boldsymbol{v}^2/2$ 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) and slow in regions of high pressure (e.g., bottom of the wing). We revisit Bernoulli's Principle in Section 45.1.2 when studying inviscid and irrotational flow.

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 (18.13) into its vector-invariant form by making use of the vector identity (see Section 2.3.4)[3]

$$\boldsymbol{\omega} \wedge \boldsymbol{v} = -\mathcal{K} + (\boldsymbol{v} \cdot \nabla)\, \boldsymbol{v}. \tag{23.61}$$

This identity allows us to eliminate velocity self-advection in favor of the vorticity and kinetic energy per mass

$$\frac{\partial \boldsymbol{v}}{\partial t} + \boldsymbol{\omega}_{\mathrm{a}} \wedge \boldsymbol{v} = -\frac{1}{\rho} \nabla p - \nabla \mathcal{M}, \tag{23.62}$$

where

$$\boldsymbol{\omega}_{\mathrm{a}} = \boldsymbol{\omega} + 2\, \boldsymbol{\Omega} \tag{23.63}$$

[3]We pursue the same manipulations in Section 34.3.1 when deriving the vorticity equation.

is the absolute vorticity (see Chapter 34) and we set the friction and heating to zero since we are assuming a perfect fluid. The Eulerian time evolution for the kinetic energy per mass is therefore given by

$$\frac{\partial \mathcal{K}}{\partial t} = -\frac{1}{\rho}\, \boldsymbol{v} \cdot \nabla p - \boldsymbol{v} \cdot \nabla \mathcal{M}, \tag{23.64}$$

where we set $\boldsymbol{v} \cdot (\boldsymbol{\omega}_{\mathrm{a}} \wedge \boldsymbol{v}) = 0$.

For Bernoulli's theorem we are interested in the steady state, with a steady kinetic energy per mass realized by the balance

$$\rho^{-1}\, \boldsymbol{v} \cdot \nabla p = -\boldsymbol{v} \cdot \nabla \mathcal{M}. \tag{23.65}$$

We can connect this steady state balance to the Bernoulli potential by noting that for a steady perfect fluid, equation (23.12) relates the pressure gradient force to

$$-\,\rho^{-1}\, \nabla p = -\nabla \mathcal{H} + T\, \nabla \mathcal{S}. \tag{23.66}$$

Combining with equation (23.65) renders

$$\boldsymbol{v} \cdot (\nabla \mathcal{H} + \nabla \mathcal{M} - T\, \nabla \mathcal{S}) = \boldsymbol{v} \cdot (\nabla \mathcal{B} - T\, \nabla \mathcal{S}) = 0. \tag{23.67}$$

A perfect fluid maintains materially constant specific entropy (Section 23.2.4), which in a steady state means that

$$\boldsymbol{v} \cdot \nabla \mathcal{S} = 0 \qquad \text{steady state perfect fluid}, \tag{23.68}$$

which in turn means that

$$\boldsymbol{v} \cdot \nabla \mathcal{B} = 0 \qquad \text{steady state perfect fluid}. \tag{23.69}$$

That is, for a steady perfect fluid the velocity is aligned with isosurfaces of specific entropy and Bernoulli potential.

23.4.5 Comments on gauge symmetry

Consider again the total energy equation as written in the Eulerian form (23.55) with the Bernoulli potential

$$\frac{\partial(\rho\, \mathcal{E})}{\partial t} + \nabla \cdot [\rho\, \boldsymbol{v}\, \mathcal{B} - 2\, \rho\, \nu\, \boldsymbol{v} \cdot \mathbb{S} + T\, \boldsymbol{J}_{\mathcal{S}} + \mu\, \boldsymbol{J}_{\mathrm{C}}] = \rho\, \partial_t \Phi. \tag{23.70}$$

It is notable that the time tendency for the total energy at a point remains unchanged if we shift the energy flux by a total curl, $\nabla \wedge \boldsymbol{G}$, with $\boldsymbol{G}$ an arbitrary *gauge function*. This arbitrariness in the definition of 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 the Feynman Lectures. 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 field energy at a point in space.

We also encounter such *gauge symmetry* in the potential vorticity flux discussed in Chapter 37, as well as the vector streamfunction for an incompressible fluid in Section 16.5.2. In some cases we can exploit the symmetry to our subjective desires, such as discussed in Section 37.4.6 for potential vorticity. However, we know of no strategic use of gauge symmetry for the study of energy budgets.

23.4.6 Further study

The postulatory approach to the conservation of total energy in Section 23.4.1 follows that of *Landau and Lifshitz* (1987), *DeGroot and Mazur* (1984) (see their Section II.4), Appendix A.13 of *IOC et al.* (2010), Appendix B in Chapter 1 of *Vallis* (2017), Section 2.4 of *Olbers et al.* (2012), and Section 13.5.5 of *Thorne and Blandford* (2017). Given the phenomenological nature of fluid mechanics, there is not a more fundamental approach.

For an examination of Bernoulli's theorem for non-rotating flows, such next to laminar boundary layers, see this tutorial video from Prof. Ascher Shapiro.

23.5 Evolution of temperature

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. These properties are the natural variables for the Gibbs potential (Section 22.4.5). 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 15.1. Pressure measures the compressive stress acting on each fluid element (Section 19.6), with its specification depending on the dominant dynamical balances (see Section 23.6). 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 beyond flux convergences.

23.5.1 Enthalpy equation again

In developing the temperature equation we find it useful to start from the prognostic equation for enthalpy as developed in Sections 23.2.6 and 23.4.1. For that purpose we write the enthalpy equation (23.53) as

$$\rho \, \frac{\mathrm{D}\mathcal{H}}{\mathrm{D}t} = \frac{\mathrm{D}p}{\mathrm{D}t} - \nabla \cdot \boldsymbol{J}_{\mathcal{H}} + \rho \, \epsilon, \tag{23.71}$$

with the enthalpy flux written in terms of the entropy and tracer fluxes

$$\boldsymbol{J}_{\mathcal{H}} = T \, \boldsymbol{J}_{\mathcal{S}} + \mu \, \boldsymbol{J}_{\mathrm{C}}. \tag{23.72}$$

23.5.2 Evolution of *in situ* temperature

To reveal a prognostic equation for temperature, we write enthalpy as a function of (T, p, C) so that

$$\frac{\mathrm{D}\mathcal{H}}{\mathrm{D}t} = \left[\frac{\partial \mathcal{H}}{\partial T}\right]_{p,C} \frac{\mathrm{D}T}{\mathrm{D}t} + \left[\frac{\partial \mathcal{H}}{\partial p}\right]_{T,C} \frac{\mathrm{D}p}{\mathrm{D}t} + \left[\frac{\partial \mathcal{H}}{\partial C}\right]_{T,p} \frac{\mathrm{D}C}{\mathrm{D}t}. \tag{23.73}$$

The partial derivatives can be related to response functions via the following. First, the specific heat capacity at constant pressure is given by equation (22.69)

$$\left[\frac{\partial \mathcal{H}}{\partial T}\right]_{p,C} = c_{\mathrm{p}}. \tag{23.74}$$

Next, we make use of the Gibbs potential identities in Section 22.4.5 to write

$$\left[\frac{\partial \mathcal{H}}{\partial p}\right]_{T,C} = \left[\frac{\partial \mathcal{G}}{\partial p}\right]_{T,C} - T\left[\frac{\partial}{\partial p}\right]_{T,C}\left[\frac{\partial \mathcal{G}}{\partial T}\right]_{p,C} \tag{23.75a}$$

$$= \nu_{\mathrm{s}} - T\left[\frac{\partial}{\partial T}\right]_{p,C}\left[\frac{\partial \mathcal{G}}{\partial p}\right]_{T,C} \tag{23.75b}$$

$$= \nu_{\mathrm{s}} - T\left[\frac{\partial \nu_{\mathrm{s}}}{\partial T}\right]_{p,C} \tag{23.75c}$$

$$= \nu_{\mathrm{s}}\,(1 - T\,\alpha_{\mathrm{T}}), \tag{23.75d}$$

where α_{T} is the thermal expansion coefficient given by equation (22.70). Use of these identities in the enthalpy equation (23.73) and rearrangement leads to the *in situ* temperature equation

$$c_{\mathrm{p}}\,\rho\,\frac{\mathrm{D}T}{\mathrm{D}t} = -\nabla\cdot\boldsymbol{J}_{\mathcal{H}} + \left[\frac{\partial \mathcal{H}}{\partial C}\right]_{T,p}\nabla\cdot\boldsymbol{J}_{\mathrm{C}} + \alpha_{\mathrm{T}}\,T\,\frac{\mathrm{D}p}{\mathrm{D}t} + \rho\,\epsilon. \tag{23.76}$$

The *in situ* temperature of a fluid element thus evolves according to convergence of the enthalpy fluxes, trace matter fluxes, pressure changes, and frictional dissipation.

23.5.3 Evolution of potential temperature

We can convert the *in situ* temperature equation (23.76) into a version of the potential temperature equation by recalling the expression (22.85) for the lapse rate

$$\hat{\Gamma} = \left[\frac{\partial T}{\partial p}\right]_{C,\mathcal{S}} = \frac{T\,\alpha_{\mathrm{T}}}{\rho\,c_{\mathrm{p}}} \tag{23.77}$$

so that equation (23.76) takes the form

$$c_{\mathrm{p}}\,\rho\left[\frac{\mathrm{D}T}{\mathrm{D}t} - \hat{\Gamma}\,\frac{\mathrm{D}p}{\mathrm{D}t}\right] = -\nabla\cdot\boldsymbol{J}_{\mathcal{H}} + \left[\frac{\partial \mathcal{H}}{\partial C}\right]_{T,p}\nabla\cdot\boldsymbol{J}_{\mathrm{C}} + \rho\,\epsilon. \tag{23.78}$$

Making use of the definition (22.113) for potential temperature renders

$$c_{\mathrm{p}}\,\rho\,\frac{\mathrm{D}\theta}{\mathrm{D}t} = -\nabla\cdot\boldsymbol{J}_{\mathcal{H}} + \left[\frac{\partial \mathcal{H}}{\partial C}\right]_{T,p}\nabla\cdot\boldsymbol{J}_{\mathrm{C}} + \rho\,\epsilon. \tag{23.79}$$

As expected, pressure changes are removed from the evolution equation for potential temperature.

23.5.4 Evolution of Conservative Temperature

Rather than expressing enthalpy as a function $\mathcal{H}(T,p,C)$, we make use of its natural coordinate dependence $\mathcal{H}(\mathcal{S},p,C)$ from Section 22.4.3, which leads to the enthalpy equation in the form (23.53)

$$\rho\,\frac{\mathrm{D}\mathcal{H}}{\mathrm{D}t} = \frac{\mathrm{D}p}{\mathrm{D}t} - \nabla\cdot\boldsymbol{J}_{\mathrm{H}} + \rho\,\epsilon. \tag{23.80}$$

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

$$\mathcal{H}^{\mathrm{pot}}(\mathcal{S}, C) = \mathcal{H}(\mathcal{S}, p_{\mathrm{R}}, C). \tag{23.81}$$

As for potential temperature, 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 (22.118). By construction, the material time derivative of potential enthalpy is given by

$$\rho \frac{\mathrm{D}\mathcal{H}^{\mathrm{pot}}}{\mathrm{D}t} = \rho \left[\frac{\partial \mathcal{H}^{\mathrm{pot}}}{\partial \mathcal{S}}\right]_C \frac{\mathrm{D}\mathcal{S}}{\mathrm{D}t} + \rho \left[\frac{\partial \mathcal{H}^{\mathrm{pot}}}{\partial C}\right]_{\mathcal{S}} \frac{\mathrm{D}C}{\mathrm{D}t} \tag{23.82a}$$

$$= \theta \left(-\nabla \cdot \boldsymbol{J}_{\mathcal{S}} + \Sigma_{\mathcal{S}}\right) - \mu_{\mathrm{R}} \nabla \cdot \boldsymbol{J}_{\mathrm{C}} \tag{23.82b}$$

$$= (\theta/T) \left[\rho \, \epsilon - \nabla \cdot \boldsymbol{J}_{\mathrm{H}}\right] - \left[\mu_{\mathrm{R}} - (\theta/T) \, \mu\right] \nabla \cdot \boldsymbol{J}_{\mathrm{C}} \tag{23.82c}$$

where we set

$$\theta = \left[\frac{\partial \mathcal{H}^{\mathrm{pot}}}{\partial \mathcal{S}}\right]_C \qquad \text{and} \qquad \mu_{\mathrm{R}} = \left[\frac{\partial \mathcal{H}^{\mathrm{pot}}}{\partial C}\right]_{\mathcal{S}} \tag{23.83}$$

and used equation (23.50) for the entropy source, $\Sigma_{\mathcal{S}}$. Now define the Conservative Temperature, Θ, via

$$c_{\mathrm{p}}^{\mathrm{ref}} \, \Theta \equiv \mathcal{H}^{\mathrm{pot}}(\mathcal{S}, C) = \mathcal{H}(\mathcal{S}, p_{\mathrm{R}}, C), \tag{23.84}$$

where $c_{\mathrm{p}}^{\mathrm{ref}}$ is an arbitrary reference specific heat capacity. For the ocean, *McDougall* (2003) suggested that $c_{\mathrm{p}}^{\mathrm{ref}}$ 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 (23.82c) 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

$$\rho \, c_{\mathrm{p}}^{\mathrm{ref}} \frac{\mathrm{D}\Theta}{\mathrm{D}t} = -\nabla \cdot \boldsymbol{J}_{\mathrm{H}} \qquad \text{ocean.} \tag{23.85}$$

Key points regarding the Conservative Temperature equation

The Conservative Temperature equation (23.85) 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 (23.79).

- 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. Even so, for many practical calculations the errors arising from ignoring these source terms are far smaller than other errors that arise from field measurements.

- 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 (23.76) and potential temperature equation (23.79). 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 rather inconvenient for purposes of budget analyses.

We close by noting that the enthalpy flux, $\boldsymbol{J}_{\mathrm{H}}$, is related to the entropy flux and concentration flux as per equation (23.72). 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 $\boldsymbol{J}_{\mathrm{H}}$ is well approximated as a flux just of Θ.

23.5.5 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 23.4.2).

23.6 Summary of the ocean-atmosphere equations

We have now completed the derivation of the full suite of equations describing rotating and stratified fluids, at least in their general form. We thus find it useful to here list these equations.

$$\rho \frac{\mathrm{D}\boldsymbol{v}}{\mathrm{D}t} + 2\,\rho\,\boldsymbol{\Omega} \wedge \boldsymbol{v} = -\rho\,\nabla\Phi - \nabla p + \nabla \cdot \mathbb{T} \qquad \text{momentum} \tag{23.86}$$

$$\frac{\mathrm{D}\rho}{\mathrm{D}t} = -\rho\,\nabla \cdot \boldsymbol{v} \qquad \text{mass continuity} \tag{23.87}$$

$$\rho \frac{\mathrm{D}C}{\mathrm{D}t} = -\nabla \cdot \boldsymbol{J}(C) \qquad \text{matter conservation} \tag{23.88}$$

$$\rho \frac{\mathrm{D}\Theta}{\mathrm{D}t} = -\nabla \cdot \boldsymbol{J}(\Theta) \qquad \text{potential enthalpy conservation} \tag{23.89}$$

$$\rho = \rho(C, \Theta, p) \qquad \text{equation of state.} \tag{23.90}$$

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, $\boldsymbol{v}$. Each of the three velocity components evolves according to its respective dynamical equation (23.86). As noted at the end of Section 18.1.3, we write the momentum equation in the form (23.86) by separating the time dependence of the basis vectors into a term arising from solid-body rotation (which leads to Coriolis and 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 spherical coordintes).

- MASS CONSERVATION: Kinematics provides a constraint on the velocity field according to the needs of mass conservation for a fluid element (Section 14.1). This constraint leads to the continuity equation (23.87).

- MATERIAL TRACER CONSERVATION: Kinematic constraints from the conservation of trace matter (Section 14.1) leads to the material tracer equation (23.88). Evolution is determined by the convergence of tracer fluxes, $\boldsymbol{J}$, with this flux specified by molecular diffusion as

discussed in Section 47.2, or through other parameterized processes when describing flow using a space cutoff larger than millimetres (see Chapter 49).

- THERMODYNAMIC TRACER: The Conservative Temperature, Θ, evolution (Section 23.5.4), is also determined by the convergence of fluxes, just like a material tracer.

- DENSITY: The *in situ* density can be updated in time via mass continuity (equation (23.87)) or via knowledge of (C, Θ, p). We discussed the many forms of density for the ocean and atmosphere in Section 21.3.

- 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 (22.91) using the density and temperature.
  - For a hydrostatic fluid (Section 24.3), pressure is diagnosed at a point through knowledge of the weight per area above the point (i.e., the mass density).
  - For an incompressible liquid, pressure is diagnosed by solving a Poisson equation derived from taking the divergence of the momentum equation (see Exercise 25.2).

- GEOPOTENTIAL: The geopotential, Φ, is specified once the height above an arbitrary reference level is known, as well as the effective gravitational acceleration (Section 9.12.2). 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

$$r = R_{\mathrm{e}} + z \tag{23.91}$$

  where $R_{\mathrm{e}} = 6.371 \times 10^6$ m is the earth radius (equation (9.108)), and z is the geopotential coordinate measuring the height above sea level.

- ANGULAR ROTATION: The earth's angular velocity, $\boldsymbol{\Omega}$, is constant for geophysical fluid studies of concern here. Its value is discussed in Section 9.2.

- FRICTION: The friction vector,

$$\rho \boldsymbol{F} = \nabla \cdot \mathbb{T}, \tag{23.92}$$

  is the divergence of a symmetric and trace-free deviatoric stress tensor, $\mathbb{T}$ (Section 19.6).

- 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 14.4; for momentum in Section 19.8; and for ocean buoyancy in Section 50.3.

23.7 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 (22.108) for the special case in which entropy changes occur only through heating in a homogeneous fluid

$$T \, \mathrm{d}\mathcal{S} = đ\mathcal{Q}. \tag{23.93}$$

Now assume heating occurs through radiation plus another term whose

$$T\,\mathrm{d}\mathcal{S} = đ\mathcal{Q} \tag{23.94a}$$
$$= L\,\mathrm{d}T + đ\mathcal{Q}_{\mathrm{rad}}, \tag{23.94b}$$

where L is the latent heat of vaporization.

23.7.1 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.

23.8 Exercises

EXERCISE 23.1: CROCCO'S THEOREM
Prove that the spatial gradient of the Bernoulli potential for a single-component steady state perfect fluid can be written

$$\nabla\mathcal{B} = T\,\nabla\mathcal{S} + \boldsymbol{v}\wedge\boldsymbol{\omega}_{\mathrm{a}}. \tag{23.95}$$

This result is known as Crocco's Theorem. 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. Hint: to help formulate this proof, study the discussion in Section 23.4.4 where we showed that the Bernoulli potential is constant along a steady flow streamline in a perfect fluid. Also recall equation (23.12) valid for a steady state and applied here to a single-component fluid.

EXERCISE 23.2: THERMODYNAMIC MANIPULATIONS FOR IDEAL GASES
This question develops some manipulations with the potential temperature.

(a) Beginning with the expression (22.128) for potential temperature of an ideal gas, show that

$$\mathrm{d}\theta = \frac{\theta}{T}\left[\mathrm{d}T - \frac{\nu}{c_{\mathrm{p}}}\mathrm{d}p\right]. \tag{23.96}$$

(b) Given the result (23.96), show that an ideal gas satisfies the following relation

$$T\,\mathrm{d}\mathcal{S} = \frac{c_p\,T}{\theta}\,\mathrm{d}\theta. \tag{23.97}$$

Whereas the relation (22.123) 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

$$T\,\frac{\mathrm{D}\mathcal{S}}{\mathrm{D}t} = \frac{c_p\,T}{\theta}\frac{\mathrm{D}\theta}{\mathrm{D}t} \Rightarrow \frac{c_p\,T}{\theta}\frac{\mathrm{D}\theta}{\mathrm{D}t} = \dot{Q}. \tag{23.98}$$

EXERCISE 23.3: THERMODYNAMIC MANIPULATIONS FOR A LIQUID
Consider seawater with specific entropy given by (see Section 1.7.2 of *Vallis* (2017))

$$\mathcal{S}(S,T,p) = \mathcal{S}_0 + c_{\mathrm{p0}}\ln(T/T_o)\left[1+\beta_{\mathrm{S}}^*\,(S-S_o)\right] - \alpha_o\,p\left[\beta_{\mathrm{T}} + \beta_{\mathrm{T}}\,\gamma^*\frac{p}{2} + \beta_{\mathrm{T}}^*\,(T-T_o)\right], \tag{23.99}$$

and corresponding specific heat capacity at constant pressure

$$c_{\mathrm{p}}(S,T,p) = c_{\mathrm{p0}}\,[1 + \beta^*_{\mathrm{S}}\,(S - S_o)] - \alpha_o\,p\,\beta^*_{\mathrm{T}}\,T. \tag{23.100}$$

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

$$\theta\,\mathrm{d}\mathcal{S} = c_{\mathrm{p}}(S,\theta,p_{\mathrm{R}})\,\mathrm{d}\theta, \tag{23.101}$$

where θ is the potential temperature and p_{R} is the corresponding reference pressure. Consequently, we can write for a moving fluid element

$$\dot{Q} = \frac{c_{\mathrm{p}}\,T}{\theta}\,\frac{\mathrm{D}\theta}{\mathrm{D}t}, \tag{23.102}$$

where we evaluate the heat capacity at $c_{\mathrm{p}}(S,\theta,p_{\mathrm{R}})$. 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 23.2. Hint: Make use of the identity (22.118).

EXERCISE 23.4: FRICTIONAL DISSIPATION FROM VISCOSITY

This exercise works through a simple case of the more general considerations from Section 23.3.3. Here, assume the friction in the momentum equation takes the form

$$\rho\,\boldsymbol{F} = \nabla\cdot(\rho\,\kappa\,\nabla\boldsymbol{v}) = \partial_n\,(\rho\,\kappa\,\partial_n\boldsymbol{v}), \tag{23.103}$$

with $\kappa > 0$ a scalar kinematic viscosity (generally non-constant). Show that when integrated over the full domain

$$\int \boldsymbol{F}\cdot\boldsymbol{v}\,\rho\,\mathrm{d}V < 0, \tag{23.104}$$

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

$$\rho\,\boldsymbol{F}\cdot\boldsymbol{v} = \rho\,F_m\,v_m = \partial_n(\rho\,\kappa\,\partial_n\,v_m)\,v_m. \tag{23.105}$$

24

Filtered equations

The thermo-hydrodynamical equations (23.86)-(23.90) are suitable to explain a huge range of phenomena. Unfortunately, this generality comes at a cost. Namely, by encapsulating so many physical scales of motion and associated dynamical processes, the equations are difficult to manage when aiming to study a focused dynamical regime. They offer us a tool whose power is overwhelming. Therefore, it is common to seek systematic methods to filter the equations to remove scales of little interest, thus enabling a more telescopic view of the dynamics. In this chapter, we develop certain of the 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.

24.1 The hydrostatic primitive equations . . . **416**
- 24.1.1 Hydrostatic approximation . . . 416
- 24.1.2 Shallow fluid approximation . . . 416
- 24.1.3 Traditional approximation . . . 416
- 24.1.4 Summary of the hydrostatic primitive equations . . . 417
- 24.1.5 Comments and further study . . . 418

24.2 The f-plane and β-plane approximations . . . **418**
- 24.2.1 Tangent plane approximation . . . 418
- 24.2.2 Traditional approximation and the f-plane . . . 419
- 24.2.3 β-plane approximation . . . 419

24.3 Hydrostatic approximation . . . **419**
- 24.3.1 Preliminaries . . . 420
- 24.3.2 Scaling relations . . . 421
- 24.3.3 Computing the horizontal hydrostatic pressure gradient . . . 422
- 24.3.4 Concerning vertical motion in hydrostatic fluids . . . 423
- 24.3.5 Gopotential thickness in an ideal gas atmosphere . . . 424
- 24.3.6 Ocean dynamic topography . . . 425
- 24.3.7 Further study . . . 426

24.4 Exercises . . . **426**

24.1 The hydrostatic primitive equations

The hydrostatic *primitive equations* provide a set of filtered 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 to the unfiltered equations.

24.1.1 Hydrostatic approximation

As discussed in Section 18.3, a static fluid in a gravity field maintains the hydrostatic balance, whereby the pressure at a point equals to the weight per area of fluid above that point. As shown in Section 24.3, the hydrostatic balance is very closely maintained for the larger 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 (18.16) between the vertical pressure gradient and the effective gravitational force

$$\frac{\partial p}{\partial r} = -\rho\, g. \tag{24.1}$$

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

$$p(r) = p(r_0) + g \int_r^{r_0} \rho \, \mathrm{d}r'. \tag{24.2}$$

We emphasize that in making the hydrostatic approximation 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 contraints imposed on the motion. We have more comment on this point in Section 24.3.4.

24.1.2 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 approximation[1] builds this scale separation into the equations of motion by setting the radial coordinate equal to the earth's radius

$$r = R_{\mathrm{e}} + z \approx R_{\mathrm{e}}. \tag{24.3}$$

This approximation is made where r appears as a multiplier, but not as a derivative operator. For example, the spherical coordinate gradient operator takes the form

$$\nabla \approx \frac{\hat{\boldsymbol{\lambda}}}{R_{\mathrm{e}} \cos\phi} \frac{\partial}{\partial \lambda} + \frac{\hat{\boldsymbol{\phi}}}{R_{\mathrm{e}}} \frac{\partial}{\partial \phi} + \hat{\boldsymbol{r}} \frac{\partial}{\partial r}. \tag{24.4}$$

24.1.3 Traditional approximation

The Traditional approximation comprises three approximations that come as a package.

[1] The shallow fluid approximation is distinct from the *shallow water approximation* treated in Part V.

Coriolis acceleration

The Traditional approximation sets to zero the Coriolis terms in the horizontal momentum equations involving the vertical velocity. The earth's angular rotation vector thus takes the form discussed in Section 9.11.6

$$\boldsymbol{\Omega} = \Omega\,\hat{\boldsymbol{Z}} \tag{24.5a}$$
$$= \Omega\,(\hat{\boldsymbol{\phi}}\,\cos\phi + \hat{\boldsymbol{r}}\,\sin\phi) \tag{24.5b}$$
$$\approx \Omega\,\sin\phi\,\hat{\boldsymbol{r}} \tag{24.5c}$$
$$= \boldsymbol{f}/2, \tag{24.5d}$$

where

$$\boldsymbol{f} = (2\,\Omega\sin\phi)\,\hat{\boldsymbol{r}} \tag{24.6}$$

is the Coriolis parameter. Hence, the Traditional approximation is concerned only with the local vertical component of the Earth's angular velocity.

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.

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. Taking one but not the other leads to an inconsistent set of equations (see Exercise 24.1).

24.1.4 Summary of the hydrostatic primitive equations

The above approximations lead to the primitive equations written in spherical coordinates

$$\frac{\mathrm{D}u}{\mathrm{D}t} - \frac{u\,v\tan\phi}{R} - f\,v = -\frac{1}{\rho\,R\cos\phi}\,\frac{\partial p}{\partial\lambda} + F^{\lambda} \tag{24.7}$$
$$\frac{\mathrm{D}v}{\mathrm{D}t} + \frac{u^2\tan\phi}{R} + f\,u = -\frac{1}{\rho\,R}\,\frac{\partial p}{\partial\phi} + F^{\phi} \tag{24.8}$$
$$\frac{\partial p}{\partial z} = -g\,\rho, \tag{24.9}$$

where the gradient operator is given by

$$\nabla = \frac{\hat{\boldsymbol{\lambda}}}{R\cos\phi}\,\frac{\partial}{\partial\lambda} + \frac{\hat{\boldsymbol{\phi}}}{R}\,\frac{\partial}{\partial\phi} + \hat{\boldsymbol{z}}\frac{\partial}{\partial z}. \tag{24.10}$$

We can write these equations in the succinct form

$$\left(\frac{\mathrm{D}}{\mathrm{D}t} + (f + u\tan\phi/R)\,\hat{\boldsymbol{z}}\wedge\right)\boldsymbol{u} = -\rho\,\nabla\Phi - \nabla p + \boldsymbol{F}, \tag{24.11}$$

where

$$\boldsymbol{F} = \hat{\boldsymbol{\lambda}}\,F^{\lambda} + \hat{\boldsymbol{\phi}}\,F^{\phi} \tag{24.12}$$

is the horizontal friction vector. Furthermore, the material time derivative in this equation signifies the relative acceleration

$$\frac{\mathrm{D}\boldsymbol{u}}{\mathrm{D}t} = \hat{\boldsymbol{\lambda}}\,\frac{\mathrm{D}u}{\mathrm{D}t} + \hat{\boldsymbol{\phi}}\,\frac{\mathrm{D}v}{\mathrm{D}t}. \tag{24.13}$$

24.1.5 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 among the earliest proponents of the hydrostatic primitive equations for use in studying the large-scale ocean and atmospheric circulation. These equations form the basis for many of the general circulation models. However, it is notable that even finer resolution simulations are making use of the non-hydrostatic equations, particularly when studying clouds in the atmosphere and convection and mixing in the ocean. These models sometimes also time step the linear momentum equations, and as such are referred to as *non-hydrostatic primitive equation* models.

24.2 The f-plane and β-plane approximations

Spherical coordinates are ideally 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 the full geophysical system. This motivation leads to the f-plane and β-plane approximations.

24.2.1 Tangent plane approximation

Consider a tangent plane located at latitude $\phi = \phi_0$ and introduce a Cartesian set of coordinates according to

$$(x, y, z) = (R\,\lambda\cos\phi_0, R\,(\phi - \phi_0), z) \tag{24.14}$$

$$(\hat{\boldsymbol{x}}, \hat{\boldsymbol{y}}, \hat{\boldsymbol{z}}) = (\hat{\boldsymbol{\lambda}}, \hat{\boldsymbol{\phi}}, \hat{\boldsymbol{r}}). \tag{24.15}$$

Use of these Cartesian coordinates leads to the following inviscid (i.e., no friction) equations of motion

$$\frac{\mathrm{D}u}{\mathrm{D}t} + 2\,(\Omega^y\,w - \Omega^z\,v) = -\frac{1}{\rho}\,\frac{\partial p}{\partial x} \tag{24.16}$$

$$\frac{\mathrm{D}v}{\mathrm{D}t} + 2\,(\Omega^z\,u - \Omega^x\,w) = -\frac{1}{\rho}\,\frac{\partial p}{\partial y} \tag{24.17}$$

$$\frac{\mathrm{D}w}{\mathrm{D}t} + 2\,(\Omega^x\,v - \Omega^y\,u) = -\frac{1}{\rho}\,\frac{\partial p}{\partial z} - g, \tag{24.18}$$

with rotational vector components

$$\boldsymbol{\Omega} = \Omega\,(\cos\phi_0\,\hat{\boldsymbol{y}} + \sin\phi_0\,\hat{\boldsymbol{z}}). \tag{24.19}$$

Note the absence of metric terms is due to the use of Cartesian coordinates on a flat planar geometry.

It is important to remind ourselves that the tangent plane approximation originates from the geopotential vertical coordinate system used for the sphere. In that coordinate system, the effective gravitational acceleration (gravity plus centrifugal) is aligned with the local vertical direction.

Correspondingly, the resulting tangent plane equations have the effective gravitational force aligned just in the $\hat{\boldsymbol{z}}$ direction. These equations are thus slightly different from those describing a fluid in a rotating tank, in which the effective gravity is not aligned with the vertical (see Exercise 24.4 and Section 29.3).

24.2.2 Traditional approximation and the f-plane

The Traditional approximation retains only the local vertical component of the rotation vector, resulting in

$$\frac{\mathrm{D}u}{\mathrm{D}t} - f\,v = -\frac{1}{\rho}\frac{\partial p}{\partial x} \tag{24.20}$$

$$\frac{\mathrm{D}v}{\mathrm{D}t} + f\,u = -\frac{1}{\rho}\frac{\partial p}{\partial y} \tag{24.21}$$

$$\frac{\mathrm{D}w}{\mathrm{D}t} = -\frac{1}{\rho}\frac{\partial p}{\partial z} - g, \tag{24.22}$$

where we introduced the constant Coriolis parameter

$$f = 2\,\Omega \sin\phi_0 \equiv f_0. \tag{24.23}$$

The f-plane approximation is the simplest model for a rotating fluid. It provides a useful member in a hierarchy of theoretical models of geophysical flows.

24.2.3 β-plane approximation

Rossby waves are planetary scale waves that sample the earth's spherical nature. The essential ingredient in 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 while retaining the flat plane geometry. We thus write

$$f = f_0 + R^{-1}\,(2\,\Omega\,\cos\phi_0)\,(y - y_0), \tag{24.24}$$

or more simply

$$f = f_0 + \beta\,y \tag{24.25}$$

$$\beta = \frac{\partial f}{\partial y} = \frac{2\,\Omega\,\cos\phi_0}{R}, \tag{24.26}$$

which constitutes the β-plane approximation.

24.3 Hydrostatic approximation

Making the hydrostatic approximation in the vertical momentum equation leads to

$$\frac{\partial p}{\partial z} = -\rho\,g. \tag{24.27}$$

Vertically integrating upward from a point in the ocean to the ocean surface leads to the hydrostatic pressure

$$p(x,y,z,t) = p[\eta(x,y,t)] + g\int_z^\eta \rho(x,y,z',t)\,\mathrm{d}z', \tag{24.28}$$

where $p(\eta)$ is the pressure at the top surface of the ocean, $z = \eta(x, y, t)$, arising from the weight of the overlying atmosphere. A similar integration applies to the atmosphere

$$p(x, y, z, t) = g \int_z^{z_{\text{top}}} \rho(x, y, z', t) \, \mathrm{d}z', \tag{24.29}$$

where we assume g is a constant over the vertical extent of the atmosphere. In either the ocean or atmosphere, the hydrostatic pressure at a vertical position, z, equals to the weight per horizontal area of material above that depth.

In addition to static flows, the hydrostatic balance is accurately maintained for a wide range of scales in a moving atmosphere and ocean. We here present a scale analysis for the hydrostatic balance in both unstratified and stratified fluids. 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

$$\alpha_{\text{aspect}} \equiv \frac{H}{L} \ll 1, \tag{24.30}$$

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 balance is concered with balances over a fluid column, it is sufficient to ignore rotation when performing a scale analysis.

24.3.1 Preliminaries

Consider the vertical momentum equation (24.18) from the tangent plane and Traditional approximations (Section 24.2), along with the associated scales for the various terms

$$\frac{\mathrm{D}w}{\mathrm{D}t} = -\frac{1}{\rho}\frac{\partial p}{\partial z} - g \tag{24.31}$$

$$\frac{W}{T} + \frac{U\,W}{L} + \frac{W\,W}{H} = -\frac{1}{\rho}\frac{\partial p}{\partial z} - g. \tag{24.32}$$

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 alternatively consider time to scale according to a wave speed and wave length, $T \sim \lambda/c$.

To get a sense for the numbers, consider the atmospheric case of $W = 10^{-2}$ m s^{-1}, $L = 10^5$ m, $H = 10^3$ m, $U = 10$ m s^{-1}. These numbers lead to $T = L/U = 10^4$ s and to the values for the vertical momentum equation

$$10^{-6} \text{ m s}^{-2} \sim -\frac{1}{\rho}\frac{\partial p}{\partial z} - g. \tag{24.33}$$

With $g \sim 10$ m s^{-2}, the only term that can balance the gravitational acceleration is the vertical pressure gradient. We thus conclude that large scale motion maintains a hydrostatic balance whereby $\partial p/\partial z = -\rho\, g$.

24.3.2 Scaling relations

We now proceed more formally by non-dimensionalizing the non-rotating inviscid oceanic Boussinesq momentum equations

$$\frac{\mathrm{D}\boldsymbol{u}}{\mathrm{D}t} = -\nabla_z \psi \tag{24.34}$$

$$\frac{\mathrm{D}w}{\mathrm{D}t} = -\frac{\partial \psi}{\partial z} + b \tag{24.35}$$

$$\nabla \cdot \boldsymbol{v} = 0, \tag{24.36}$$

where

$$\nabla_z = \hat{\boldsymbol{x}}\frac{\partial}{\partial x} + \hat{\boldsymbol{y}}\frac{\partial}{\partial y} \tag{24.37}$$

is the horizontal gradient operator. For the present analysis, ignore the potential impacts from vertical stratification (we include stratification in Section 25.7). To non-dimensionalize, introduce the dimensional scales (capital letters) and corresponding non-dimensional variables (hat variables) according to

$$(x, y) = L\,(\hat{x}, \hat{y}) \qquad z = H\,\hat{z} \qquad \boldsymbol{u} = U\,\hat{\boldsymbol{u}} \qquad w = W\,\hat{w} \tag{24.38}$$

$$t = T\,\hat{t} \qquad \psi = \Psi\,\hat{\psi} \qquad b = B\,\hat{b}. \tag{24.39}$$

Importantly, the dimensional scales are constrained through the equations of motion. To expose the constraints requires us to impose our subjective input based on the regimes of interest. To start, assume, as before, that the time scale is determined by the horizontal velocity and the length scale

$$T = \frac{L}{U}. \tag{24.40}$$

Secondly, scale the vertical velocity according to the continuity equation

$$\nabla_z \cdot \boldsymbol{u} + \frac{\partial w}{\partial z} = 0 \Rightarrow W \sim \frac{UH}{L} = U\alpha_{\text{aspect}}. \tag{24.41}$$

Third, scale the pressure according to the non-rotating balance of the horizontal advection and the horizontal pressure gradient

$$\frac{UU}{L} \sim \frac{\Psi}{L} \Rightarrow \Psi \sim U^2. \tag{24.42}$$

Fourth, scale buoyancy according to the hydrostatic balance

$$B \sim \frac{\Psi}{H} = \frac{U^2}{H}. \tag{24.43}$$

Bringing these results together leads to

$$\frac{U}{T}\frac{\partial \hat{u}}{\partial \hat{t}} + \frac{U^2}{L}\hat{u}\frac{\partial \hat{u}}{\partial \hat{x}} + \frac{U^2}{L}\hat{v}\frac{\partial \hat{u}}{\partial \hat{y}} + \frac{UW}{H}\hat{w}\frac{\partial \hat{u}}{\partial \hat{z}} = -\frac{\Psi}{L}\frac{\partial \hat{\psi}}{\partial \hat{x}} \tag{24.44}$$

$$\frac{U}{T}\frac{\partial \hat{v}}{\partial \hat{t}} + \frac{U^2}{L}\hat{u}\frac{\partial \hat{v}}{\partial \hat{x}} + \frac{U^2}{L}\hat{v}\frac{\partial \hat{v}}{\partial \hat{y}} + \frac{UW}{H}\hat{w}\frac{\partial \hat{v}}{\partial \hat{z}} = -\frac{\Psi}{L}\frac{\partial \hat{\psi}}{\partial \hat{y}} \tag{24.45}$$

$$\frac{W}{T}\frac{\partial \hat{w}}{\partial \hat{t}} + \frac{UW}{L}\hat{u}\frac{\partial \hat{w}}{\partial \hat{x}} + \frac{UW}{L}\hat{v}\frac{\partial \hat{w}}{\partial \hat{y}} + \frac{WW}{H}\hat{w}\frac{\partial \hat{w}}{\partial \hat{z}} = -\frac{\Psi}{H}\frac{\partial \hat{\psi}}{\partial \hat{z}} + B\,\hat{b}. \tag{24.46}$$

Tidying up these relations leads to the non-dimensional oceanic Boussinesq equations

$$\frac{\mathrm{D}\hat{u}}{\mathrm{D}\hat{t}} = -\hat{\nabla}\hat{\psi} \tag{24.47}$$

$$\alpha^2_{\text{aspect}}\frac{\mathrm{D}\hat{w}}{\mathrm{D}\hat{t}} = -\frac{\partial \hat{\psi}}{\partial \hat{z}} + \hat{b}, \tag{24.48}$$

where we introduced the non-dimensional material time derivative

$$\frac{\mathrm{D}}{\mathrm{D}\hat{t}} = \frac{\partial}{\partial \hat{t}} + \hat{\boldsymbol{v}}\cdot\hat{\nabla} \tag{24.49}$$

and the non-dimensional gradient operator

$$\hat{\nabla} = \frac{\partial}{\partial \hat{x}} + \frac{\partial}{\partial \hat{y}} + \frac{\partial}{\partial \hat{z}}. \tag{24.50}$$

The non-dimensional vertical momentum equation (24.48) reveals that for small aspect ratio flow, the vertical momentum equation reduces to the hydrostatic balance

$$\frac{\partial \hat{\psi}}{\partial \hat{z}} = \hat{b} \qquad \text{for } \alpha^2_{\text{aspect}} \ll 1. \tag{24.51}$$

24.3.3 Computing the horizontal hydrostatic pressure gradient

We offer an example of how horizontal differences in density lead to horizontal pressure gradients through the hydrostatic relation. This example is emblematic of how one can determine the sign for horizontal gradients in a hydrostatic fluid, and thus to determine the direction for the pressure force.

Two columns with equal mass yet different densities

For this purpose consider two adjacent columns of seawater with equal mass but with distinct density, and assume the density in each column is constant throughout the respective columns. Figure 24.1 offers a schematic, where we make the additional assumption that the two columns sit on a flat surface. 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 volumn so that its free surface sits higher

$$\eta_{\mathrm{B}} > \eta_{\mathrm{A}}. \tag{24.52}$$

So, 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 pressures are identical since the two columns have equal mass.

Computing $p(z)$ 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

$$p_{\text{bot}} = g\,\rho_{\text{A}}\,(\eta_{\text{A}} - \eta_{\text{b}}) = g\,\rho_{\text{B}}\,(\eta_{\text{B}} - \eta_{\text{b}}), \tag{24.53}$$

where $z = \eta_{\text{b}}(x, y)$ is the vertical position at the bottom and $z = \eta(x, y, t)$ is the surface. Hence, 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

$$p_{\text{A}}(z) = g\,\rho_{\text{A}}\,(\eta_{\text{A}} - z) = p_{\text{bot}} - g\,\rho_{\text{A}}\,(-\eta_{\text{b}} + z). \tag{24.54}$$

The second equality arose from substituting the bottom pressure from equation (24.53) to eliminate the surface height η_{A}. Likewise, the pressure in column B is given by

$$p_{\text{B}}(z) = g\,\rho_{\text{B}}\,(\eta_{\text{B}} - z) = p_{\text{bot}} - g\,\rho_{\text{B}}\,(-\eta_{\text{b}} + z). \tag{24.55}$$

We can now take the difference between the two hydrostatic pressures to find

$$p_{\text{B}}(z) - p_{\text{A}}(z) = g\,(-\eta_{\text{b}} + z)\,(\rho_{\text{A}} - \rho_{\text{B}}) > 0. \tag{24.56}$$

Hence, at any point above the bottom, the hydrostatic pressure in column B is greater than that in column A. 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

$$F_{\text{pressure B to A}} = \int_{-H}^{\eta_{\text{A}}} [p_{\text{B}}(z) - p_{\text{A}}(z)]\,\mathrm{d}z = (g/2)\,(\rho_{\text{A}} - \rho_{\text{B}})\,(\eta_{\text{A}} - \eta_{\text{b}})^2. \tag{24.57}$$

Inferring pressure gradients starting from the top

Another way to understand why the pressure force points from column B to 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 $z = \eta_{\text{B}}$ the pressure increases in column B immediately, whereas the pressure in column A remains zero until entering the water at $z = \eta_{\text{A}} < \eta_{\text{B}}$. 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 24.1.

24.3.4 Concerning vertical motion in hydrostatic fluids

For a static fluid with identically zero acceleration, the vertical pressure gradient precisely balances the weight of fluid thus realizing exact hydrostatic balance. We discussed this static solution to the equations of motion in Section 18.3. 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 acclerations acting on a fluid element. In this case, the vertical momentum equation, even for the moving fluid, remains approximately in hydrostatic balance.

Unbalanced vertical accelerations still exist in a moving hydrostatic fluid, thus allowing for nonzero vertical motion. Yet these vertical motions are not prognosed by the vertical momentum

Figure 24.1: *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: $p_A(z = \eta_b) = p_B(z = \eta_b) = p_{bot}$. 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 (24.56) 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 29.5. Compare this schematic to Figure 26.3, which discusses the depth dependence of the horizontal gradient in hydrostatic pressure as per $\partial(\nabla_z p)/\partial z = -g\,\nabla_z \rho$.*

equation in a hydrostatic fluid. Instead, they are diagnosed through constraints on the fluid motion. Continuity provides the canonical kinematic constraint on fluid motion. Another constraint arises from vorticity and potential vorticity (Part VI). For example, as seen in Section 25.2.3, horizontal velocity divergence in an incompressible fluid is balanced by vertical velocity convergence as per equation (25.14). The vertical pressure forces required to produce the converging vertical motion are those precisely needed to maintain continuity. In a hydrostatic fluid, we do not directly compute these forces and use them to drive vertical motion. Instead, the motion is induced through the kinematic constraint. The associated forces, if desired, can be diagnosed given the velocity and the accelerations.

24.3.5 Gopotential thickness in an ideal gas atmosphere

From the hydrostatic equation (18.33) we know that the pressure at a geopotential Φ_1 equals to

$$p(\Phi_1) = \int_{\Phi_1}^{\infty} \rho(\Phi)\,\mathrm{d}\Phi \tag{24.58}$$

where we assumed that $p(\Phi_2 = \infty) = 0$. Equation (18.32) allows us to write the integrand as

$$\mathrm{d}\Phi = -\rho^{-1}\,\mathrm{d}p = -\frac{T\,R^{\mathrm{atm}}\,\mathrm{d}p}{p}, \tag{24.59}$$

where the second equality assumed the atmosphere to be an ideal gas (Section 22.8) so that the specific volume is given by

$$\rho^{-1} = \frac{T\,R^{\mathrm{atm}}}{p}, \tag{24.60}$$

with R^{atm} (equation (22.92)) the *specific gas constant* as defined by the universal gas constant normalized by the molar mass for the atmosphere. Vertical integration of equation (24.59) leads to the *hypsometric equation*, which provides the geopotential thickness between two pressure isosurfaces

$$\Phi(z_2) - \Phi(z_1) = -R^{\text{atm}} \int_{p_1}^{p_2} T \, \mathrm{d}(\ln p). \tag{24.61}$$

Recall that $\mathrm{d}p < 0$ if $\mathrm{d}z > 0$ since the hydrostatic pressure decreases when moving up in the atmosphere. We define the *geopotential height* according to

$$Z = \Phi / g_0, \tag{24.62}$$

where g_0 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 (24.61) says that the geopotential thickness between two isobars is

$$Z_2 - Z_1 = \frac{R^{\text{atm}}}{g_0} \int_{p_2}^{p_1} T \, \mathrm{d}(\ln p). \tag{24.63}$$

Defining the layer mean temperature

$$\langle T \rangle = \frac{\int_{p_2}^{p_1} T \, \mathrm{d}(\ln p)}{\int_{p_2}^{p_1} \mathrm{d}(\ln p)} \tag{24.64}$$

and the layer mean scale height

$$H = \frac{R^{\text{atm}} \langle T \rangle}{g_0} \tag{24.65}$$

leads to the geopotential thickness

$$\Delta Z = Z_2 - Z_1 = H \, \ln(p_1/p_2). \tag{24.66}$$

The geopotential thickness is thus proportional to the mean temperature within the pressure layer. Furthermore, if the atmosphere has constant temperature T (unrealistic but useful special limit), then the geopotential height of a given pressure surface equals to

$$Z = H \ln(p_{\text{s}}/p), \tag{24.67}$$

where p_{s} is the surface pressure. Correspondingly, the pressure in an isothermal atmosphere decreases exponentially with geopotential height

$$p(Z) = p_{\text{s}} \exp(-Z/H). \tag{24.68}$$

24.3.6 Ocean dynamic topography

There are occasions where it is useful to study the thickness of a layer separated by isobars. For example, it is often assumed in dynamical oceanography that there is a pressure at which baroclinic currents vanish, with the reduced gravity model of Section 29.4 studying the associated dynamical implications. 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

$$\mathcal{D}(\mathcal{P}) = \eta - z(\mathcal{P}). \tag{24.69}$$

Assuming a hydrostatic balance allows us to relate this expression to the vertical integral between two pressure surfaces of the specific volume ρ^{-1}

$$\mathcal{D}(\mathcal{P}) = \int_{z(\mathcal{P})}^{\eta} \mathrm{d}z = g^{-1} \int_{p_\mathrm{a}}^{\mathcal{P}} \frac{\mathrm{d}p}{\rho}, \tag{24.70}$$

where the second step used the hydrostatic balance. We refer to the thickness $\mathcal{D}(\mathcal{P})$ as the *dynamic topography* with respect to a reference pressure $\mathcal{P}$. Note that it is sometimes also called the *steric sea level*. Evolution of the dynamic topography $\mathcal{D}$ arises from changes in the pressure applied to the free surface as well as changes in the specific volume

$$g \frac{\partial \mathcal{D}(\mathcal{P})}{\partial t} = -\frac{1}{\rho(\eta)} \frac{\partial p_\mathrm{a}}{\partial t} + \int_{p_\mathrm{a}}^{\mathcal{P}} \frac{\partial \rho^{-1}}{\partial t} \, \mathrm{d}p, \tag{24.71}$$

where the time derivative acting on the specific volume is taken on surfaces of constant pressure. If the depth $z(\mathcal{P})$ of the constant pressure surface is static, then the evolution of layer thickness $\mathcal{D}(\mathcal{P})$ is identical to the sea level η. In general, there is no such static pressure level, thus making the time tendencies differ, though certain situations may warrant this approximation.

24.3.7 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 field in Section 24.3.3 is motivated by similar considerations presented in Chapter 2 of the descriptive oceanography text from *Tomczak and Godfrey* (1994). Section 1.6 of *Holton* (1992) provides more discussion of properties of a hydrostatic atmosphere. Further discussion of dynamic topography is given in Appendix B.4 of *Griffies et al.* (2014) as well as in *Tomczak and Godfrey* (1994).

24.4 Exercises

EXERCISE 24.1: PRIMITIVE EQUATIONS AND AXIAL ANGULAR MOMENTUM

The axial angular momentum of a fluid element satisfying the primitive equations is given by

$$L^z = (\rho \, \delta V) \, R_\perp \, (u + R_\perp \, \Omega) \equiv (\rho \, \delta V) \, l^z \tag{24.72}$$

where

$$R_\perp = R_\mathrm{e} \cos \phi \tag{24.73}$$

is the distance from the polar rotation axis to a point on the sphere with radius R_e, and

$$l^z = R_\perp \, (u + R_\perp \, \Omega) \tag{24.74}$$

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 (24.7) implies that the material evolution of axial angular momentum per mass is given by

$$\frac{\mathrm{D}l^z}{\mathrm{D}t} = -\frac{1}{\rho} \frac{\partial p}{\partial \lambda}. \tag{24.75}$$

(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 10.1.1.

(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 your symmetry argument given in the previous part of this exercise?

EXERCISE 24.2: RELATIONS FOR AN ATMOSPHERE

In this exercise, we establish some relations for an atmosphere. Assume for this problem 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. Answering the parts to this question will require reviewing elements of thermodynamics detailed in Chapter 22.

(a) For an ideal gas atmosphere in 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

$$\int_0^{z_{\text{top}}} (\Phi + \mathcal{I})\, \rho \, \mathrm{d}z = \int_0^{z_{\text{top}}} \mathcal{H}\, \rho \, \mathrm{d}z, \tag{24.76}$$

where

$$\mathcal{H} = p\, \alpha + \mathcal{I} \tag{24.77}$$

is the enthalpy per mass,

$$\Phi = g\, z \tag{24.78}$$

is the geopotentia, which is also the gravitational potential energy per mass (Section 9.12.2), 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 $z = z_{\text{top}}$.

(b) For an ideal gas atmosphere in hydrostatic balance, show that

$$\frac{\mathrm{d}\sigma}{\mathrm{d}z} = \Pi \frac{\mathrm{d}\theta}{\mathrm{d}z}, \tag{24.79}$$

where

$$\sigma = \mathcal{H} + \Phi \tag{24.80}$$

is the dry static energy and

$$\Pi = \frac{c_{\text{p}}\, T}{\theta} \tag{24.81}$$

is the *Exner function*.

(c) For an ideal gas atmosphere (either hydrostatic or non-hydrostatic), derive the following expression for the pressure gradient force

$$-\frac{1}{\rho}\,\nabla p = -\theta\,\nabla\Pi. \tag{24.82}$$

(d) For an ideal gas atmosphere (either hydrostatic or non-hydrostatic), derive the following expression for the pressure gradient force

$$-\frac{1}{\rho}\,\nabla p = -\frac{c_s^2}{\rho\,\theta}\,\nabla(\rho\,\theta), \tag{24.83}$$

where c_s is the sound speed.

(e) Show that for a hydrostatic atmosphere with an arbitrary equation of state

$$\int_0^{p_{\mathrm{s}}} (\Phi + \mathcal{I})\,\mathrm{d}p = \int_0^{p_{\mathrm{s}}} \mathcal{H}\,\mathrm{d}p. \tag{24.84}$$

That is, show that the relation in the first part of this problem holds even without making the ideal gas assumption.

EXERCISE 24.3: MASS BALANCE FOR A COLUMN OF HYDROSTATIC FLUID

Return to Exercise 14.2. Show that for a hydrostatic fluid the mass balance for a fluid column (equation (14.88)) takes the form

$$\frac{1}{g}\frac{\partial\,(p_{\mathrm{b}} - p_{\mathrm{a}})}{\partial t} = -\nabla\cdot\left[\int_{\eta_{\mathrm{b}}}^{\eta} \mathbf{u}\,\rho\,\mathrm{d}z\right] + Q_{\mathrm{m}}, \tag{24.85}$$

where

$$p_{\mathrm{b}} = p_{\mathrm{a}} + g\int_{\eta_{\mathrm{b}}}^{\eta} \rho\,\mathrm{d}z \tag{24.86}$$

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 (24.86) into (14.88).

EXERCISE 24.4: ROTATING TANK FLUID DYNAMICS

Consider a rotating circular flat bottom laboratory-scale tank filled with uniform density water. Let the rotational axis be through the center of the circle with angular velocity $\mathbf{\Omega} = \Omega\,\hat{\boldsymbol{z}}$. Orient the vertical coordinate so that the bottom is at $z = 0$ and water surface at $z = \eta(x, y, t)$. Let the gravitational acceleration be uniform $-g_{\mathrm{e}}\,\hat{\boldsymbol{z}}$ (Section 9.12). Ignore friction throughout.

(a) A laboratory frame observer is not inertial, since he/she is on the rotating planet. But for the purpose of describing fluid motion in a laboratory scale rotating tank, it seems intuitive that one can ignore the Coriolis force associated with the rotating earth reference frame. Discuss why this intuition is sensible, given a typical rotational speed for a rotating tank (e.g., the speed of a 45 rpm record player) versus that of the planet. We offer scaling analysis in Section 26.2 to further support this intuition.

(b) Derive the equations of motion for the fluid in the reference frame of an observer in the laboratory, assumed to be in an inertial reference frame.

(c) Derive the equations of motion in the frame rotating with the tank.

(d) What is the geometric shape of the free surface when the fluid undergoes solid-body rotation. Neglect any variations in the atmospheric pressure applied to the upper surface.

25

Oceanic Boussinesq fluid

Ocean density deviates no more than a few percent relative to the volume mean density. Although small, the density deviations are crucial for driving large-scale ocean circulations derived from variations in temperature and salinity (*thermohaline circulation*). The reason is that the density variations are multiplied by the relatively large gravitational acceleration when computing pressure forces. The oceanic Boussinesq approximation offers a systematic means to exploit the small density deviations where they can be exploited, while retaining the variations where they are critical such as when multiplied by gravity. In brief, the oceanic Boussinesq approximation maintains a compressibile thermodynamics and an incompressible kinematics. Doing so 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.

READER'S GUIDE TO THIS CHAPTER

This chapter focuses on deriving the oceanic Boussinesq equations and exploring their physical properties. These considerations have a much broader application than just to the ocean, given that many of the physical and mathematical features also hold for atmospheric flow satisfying the anelastic approximation. In studying the oceanic Boussinesq fluid, we encounter many elements of fluid mechanics found in other chapters from this part of the book, thus offering an opportunity to bolster those earlier lessons.

25.1 Loose threads **432**
25.2 The oceanic Boussinesq approximation **432**
25.2.1 Isolating the dynamically active pressure field 432
25.2.2 Boussinesq momentum equation 433
25.2.3 Mass continuity 434
25.2.4 A non-divergent velocity with density evolution 434
25.2.5 Thermodynamic equation 435
25.2.6 Version I of the Boussinesq equations 435
25.2.7 Version II of the oceanic Boussinesq equations 436
25.2.8 Comments and further study 437
25.3 Subgrid scale processes and boundary conditions **437**
25.3.1 Material evolution of buoyancy 437
25.3.2 Mechanical forcing and dissipation 438
25.4 Mechanical energy analysis part I **439**

25.4.1 Kinetic energy . . . 439
25.4.2 Gravitational and mechanical energies . . . 440
25.5 Mechanical energy analysis part II . . . **441**
25.5.1 Boussinesq dynamic enthalpy . . . 441
25.5.2 Comments . . . 443
25.6 Available potential energy . . . **443**
25.6.1 Analytic continuation of buoyancy surfaces . . . 444
25.6.2 The dual relation between height and buoyancy . . . 447
25.6.3 Exact expression for APE . . . 448
25.6.4 Approximate expression for APE . . . 449
25.6.5 Comments . . . 451
25.7 Stratification effects on hydrostatic scaling . . . **451**
25.8 Exercises . . . **454**

25.1 Loose threads

- Pressure in an incompressible fluid as per appendix to Jevangee and Romps (2015). In an incompressible fluid, pressure must conform to the needs of continuity.

25.2 The oceanic Boussinesq approximation

In this section we derive the oceanic Boussinesq approximation by isolating portions of the buoyancy and pressure contributions and then making use of basic scaling analysis.

25.2.1 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. This decomposition holds even when the fluid is non-hydrostatic. Most conveniently we consider a 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 convenenient 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 we start by decomposing density according to

$$\rho(\boldsymbol{x}, t) = \rho_0(z) + \delta\rho(\boldsymbol{x}, t) \tag{25.1}$$

where the deviation density is much smaller than the reference density

$$\delta\rho \ll \rho_0. \tag{25.2}$$

The following formulation for the momentum equation holds for the general case of $\rho_0(z)$. However, we note in Section 25.2.2 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.[1]

[1] Exercise 25.1 offers a modest means to generalize this assumption.

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, $\delta p(\boldsymbol{x}, t)$,

$$p(\boldsymbol{x}, t) = p_0(z) + \delta p(\boldsymbol{x}, t). \tag{25.3}$$

The background pressure is assumed to be in hydrostatic balance with the reference density

$$\frac{\mathrm{d}p_0}{\mathrm{d}z} = -\rho_0\, g. \tag{25.4}$$

By definition, both p_0 and ρ_0 are static.

With the above density and pressure decompositions, the momentum equation

$$\rho \left[\frac{\mathrm{D}\boldsymbol{v}}{\mathrm{D}t} + 2\,\boldsymbol{\Omega} \wedge \boldsymbol{v} \right] = -\nabla p - \hat{\boldsymbol{z}}\, g\, \rho \tag{25.5}$$

takes the equivalent form

$$(\rho_0 + \delta\rho) \left[\frac{\mathrm{D}\boldsymbol{v}}{\mathrm{D}t} + 2\,\boldsymbol{\Omega} \wedge \boldsymbol{v} \right] = -\nabla \delta p - g\, \delta\rho\, \hat{\boldsymbol{z}} - \left[\frac{\mathrm{d}p_0}{\mathrm{d}z} + \rho_0\, g \right] \hat{\boldsymbol{z}} \tag{25.6a}$$

$$= -\nabla \delta p - g\, \delta\rho\, \hat{\boldsymbol{z}}, \tag{25.6b}$$

where we used the hydrostatic balance (25.4) for the second equality. As advertised above, the background hydrostatic pressure, p_0, leads to no motion since it drops out from the momentum equation. Rather, only the gradients of δp lead to motion. We have thus succeeded in isolating the dynamically relevant portion of the pressure field, with this decomposition useful for general flows even when not making the Boussinesq or hydrostatic approximations.

25.2.2 Boussinesq momentum equation

Dividing the momentum equation (25.6b) by the reference density, and dropping the small term $\delta\rho/\rho_0$ multiplying the material acceleration on the left hand side, yields the Boussinesq momentum equation

$$\frac{\mathrm{D}\boldsymbol{v}}{\mathrm{D}t} + 2\,\boldsymbol{\Omega} \wedge \boldsymbol{v} = -\frac{1}{\rho_0}\, \nabla \delta p + b\, \hat{\boldsymbol{z}}, \tag{25.7}$$

where we introduced the buoyancy field defined relative to the background density

$$b = -\frac{g\, \delta\rho}{\rho_0} = -\frac{g\, (\rho - \rho_0)}{\rho_0} \tag{25.8}$$

As defined, buoyancy is positive when the *in situ* density is less than the reference density so that $\delta\rho = \rho - \rho_0 < 0$. That is, $b > 0$ when the fluid element is lighter (more buoyant) than the reference density. Buoyancy is the product of the gravitational acceleration, which is a relatively large term, and the small number $\delta\rho/\rho_0$. Their product is not small, in which case the product cannot be neglected from the momentum equation. In effect, the Boussinesq momentum equation ignores all density variations *except* when multiplied by gravity.

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

$$\varphi = \frac{\delta p}{\rho_0} = \frac{p - p_0(z)}{\rho_0}. \tag{25.9}$$

In this case the Boussinesq momentum equation takes the form

$$\frac{\mathrm{D}\boldsymbol{v}}{\mathrm{D}t} + 2\,\boldsymbol{\Omega} \wedge \boldsymbol{v} = -\nabla\varphi + b\,\hat{\boldsymbol{z}}. \tag{25.10}$$

Although we have yet to discuss vorticity and potential vorticity, we note that this form of the pressure gradient acceleration is particularly useful given that the curl of the right hand side eliminates pressure from the vorticity equation (Section 34.7.1). In contrast, for the general form with $\rho_0(z)$, the Boussinesq baroclinicity vector (equation (34.97)) has a contribution from both pressure and buoyancy

$$\boldsymbol{B} = \left[\nabla b - \frac{1}{\rho_0^2}\frac{\partial\rho_0}{\partial z}\,\nabla\delta p\right] \wedge \hat{\boldsymbol{z}}. \tag{25.11}$$

The additional pressure contribution complicates the development of potential vorticity (Section 36.5.1), thus serving as motivation to use a space-time constant reference density so that $\partial_z\rho_0 = 0$. Following this motivation, we generally assume ρ_0 is a constant in this book. Even so, in the next subsection we show how to split the difference by defining a slightly more general buoyancy field while retaining a constant ρ_0.

25.2.3 Mass continuity

The mass continuity equation

$$\frac{\mathrm{D}\rho}{\mathrm{D}t} = -\rho\,\nabla\cdot\boldsymbol{v}, \tag{25.12}$$

takes the form

$$\frac{\mathrm{D}\delta\rho}{\mathrm{D}t} = -(\rho_0 + \delta\rho)\,(\nabla_z\cdot\boldsymbol{u} + \partial_z w). \tag{25.13}$$

The material time derivative on the left hand side is much smaller than either of the two terms appearing on the right hand side. Consequently, to leading order, the divergence of the horizontal velocity balances the vertical convergence of the vertical velocity

$$\nabla_z\cdot\boldsymbol{u} = -\partial_z w. \tag{25.14}$$

That is, the velocity field for a Boussinesq fluid is non-divergent

$$\nabla\cdot\boldsymbol{v} = 0, \tag{25.15}$$

so that the flow is incompressible.

25.2.4 A non-divergent velocity with density evolution

We here offer some comments regarding the use of a non-divergent velocity for the Boussinesq system, while still having a non-zero material evolution of density.

Comments about density evolution and the thermohaline circulation

Use of a non-divergent velocity for the Boussinesq equations does not mean that the material time evolution of ρ vanishes identically. Instead, the scaling in Section 25.2.3 focuses just on the mass continuity equation. We must additionally acknowledge that as temperature and salinity evolve, so too does density as determined through the equation of state. Such changes in density translate into changes in pressure, which in turn drive the large-scale *thermohaline circulation*.

Divergent and non-divergent velocity components

As noted above, 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, $\boldsymbol{v}^{\mathrm{d}}$, that balances the material evolution of density

$$\frac{1}{\delta\rho}\frac{\mathrm{D}\delta\rho}{\mathrm{D}t} = -\nabla\cdot\boldsymbol{v}^{\mathrm{d}} \neq 0. \tag{25.16}$$

The divergent velocity $\boldsymbol{v}^{\mathrm{d}}$ is not used for any of the Boussinesq dynamical equations. Nonetheless, $\boldsymbol{v}^{\mathrm{d}} \neq 0$, as its divergence is required to balance the material evolution of density according to equation (25.16).

25.2.5 Thermodynamic equation

The thermodynamic equation provides a prognostic relation for the buoyancy. There are various forms for this relation, depending on assumptions made in determining the density. We discussed the flavors for density in Section 21.3. For purposes of realistic ocean modeling, the most accurate expression for density is critical. For idealized modeling, it is common to assume buoyancy equals to a constant times the Conservative Temperature. These considerations prompt us to in Sections 25.2.6 and 25.2.7 to summarize two forms for the oceanic Boussinesq fluid commonly encountered in this book.

25.2.6 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

$$\frac{\mathrm{D}\boldsymbol{v}}{\mathrm{D}t} + 2\,\boldsymbol{\Omega}\wedge\boldsymbol{v} = -\nabla\varphi + b\,\hat{\boldsymbol{z}} + \boldsymbol{F} \qquad \text{velocity equation} \tag{25.17a}$$

$$\nabla\cdot\boldsymbol{v} = 0 \qquad \text{continuity equation} \tag{25.17b}$$

$$\frac{\mathrm{D}b}{\mathrm{D}t} = \dot{b} \qquad \text{buoyancy equation} \tag{25.17c}$$

$$b = -\frac{g\,\delta\rho}{\rho_0} = -\frac{g\,(\rho-\rho_0)}{\rho_0} \qquad \text{buoyancy defined} \tag{25.17d}$$

$$\varphi = \frac{\delta p}{\rho_0} = \frac{p - p_0(z)}{\rho_0} \qquad \text{dynamic pressure defined} \tag{25.17e}$$

$$\rho = \rho_0\,(1 - \alpha\,\Theta + \beta\,S) \qquad \text{linear equation of state} \tag{25.17f}$$

$$\frac{\mathrm{d}p_0}{\mathrm{d}z} = -\rho_0\,g \qquad \text{background hydrostatic pressure.} \tag{25.17g}$$

We offer the following comments on these equations

- EQUATION OF STATE: The equation of state, (25.17f), 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

nonlinear processes such as cabbeling and thermobaricity (see Section 49.4). These processes are important for certain features of the ocean, thus prompting the more general equation set in Section 25.2.7.

- HYDROSTATIC APPROXIMATION: Most numerical models of the large-scale ocean circulation are hydrostatic. Making the hydrostatic approximation in the velocity equation (25.17a) leads to the split into a horizontal velocity equation plus a vertical hydrostatic balance

$$\frac{\mathrm{D}\boldsymbol{u}}{\mathrm{D}t} + 2\,\boldsymbol{\Omega} \wedge \boldsymbol{v} = -\nabla_z \varphi + \boldsymbol{F}^h \qquad \text{horizontal velocity equation} \tag{25.18a}$$

$$\frac{\partial \varphi}{\partial z} = b \qquad \text{hydrostatic balance,} \tag{25.18b}$$

where ∇_z is the horizontal gradient operator and $\boldsymbol{F}^h$ is the horizontal frictional acceleration.

25.2.7 Version II of the oceanic Boussinesq equations

The non-hydrostatic Boussinesq equations (25.17a)-(25.17g) 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 movitated to consider the equations in the form

$$\frac{\mathrm{D}\boldsymbol{v}}{\mathrm{D}t} + 2\,\boldsymbol{\Omega} \wedge \boldsymbol{v} = -(1/\rho_0)\,(\nabla p + \rho\,\nabla\Phi) + \boldsymbol{F} \qquad \text{velocity equation} \tag{25.19a}$$

$$\nabla \cdot \boldsymbol{v} = 0 \qquad \text{continuity equation} \tag{25.19b}$$

$$\frac{\mathrm{D}S}{\mathrm{D}t} = \dot{S} \qquad \text{salinity equation} \tag{25.19c}$$

$$\frac{\mathrm{D}\Theta}{\mathrm{D}t} = \dot{\Theta} \qquad \text{Conservative Temperature equation} \tag{25.19d}$$

$$\rho = \rho(S, \Theta, \Phi) \qquad \text{equation of state.} \tag{25.19e}$$

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 43)

$$\Phi = \Phi(\boldsymbol{x}, t). \tag{25.20}$$

- EQUATION OF STATE: The equation of state is a function of salinity, Conservative Temperature, and geopotential, thus allowing for nonlinear density transformation processes such as cabelling and thermobaricity (see Section 49.4). Even so, the pressure dependence in the equation of state is computed as per a homogeneous and resting hydrostatic fluid

$$\rho(S, \Theta, \Phi) = \rho(S, \Theta, p = -\rho_0\,\Phi). \tag{25.21}$$

In Section 25.5 we see why it is appropriate to take this functional form rather than the more general form discussed in Section 21.3, in which density is a function of the full *in situ* pressure rather than the geopotential.

- 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 (25.19a) splits into a horizontal velocity equation and hydrostatic balance

$$\frac{\mathrm{D}\boldsymbol{u}}{\mathrm{D}t} + 2\,\boldsymbol{\Omega} \wedge \boldsymbol{v} = -(1/\rho_0)\,(\nabla_z p + \rho\,\nabla_z \Phi) + \boldsymbol{F}^h \qquad \text{horizontal velocity equation} \tag{25.22a}$$

$$\frac{\partial p}{\partial z} = -g\,\rho \qquad \text{hydrostatic balance.} \tag{25.22b}$$

25.2.8 Comments and further study

We offer the following points concerning the oceanic Boussinesq approximation.

- 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 assumes a linear equation of state. However, as we saw in Section 25.2.7, the oceanic Boussinesq approximation does not generally assume a linear equation of state, thus making it suitable for realistic ocean circulation studies.

- 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.

- FURTHER READING: 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).

25.3 Subgrid scale processes and boundary conditions

The Boussinesq equations from Section 25.2 must be supplemented with boundary conditions as well as subgrid scale parameterizations. We examine boundary stresses in Chapter 19.8; study tracer subgrid scale parameterizations in Chapter 49; and detail the variety of boundary buoyancy fluxes in Section 50.3. Even so, there is far more to each of these stories than provided in this book. We here merely summarize certain facets sufficient to give a taste of the issues and to help work through exercises at the end of this chapter.

25.3.1 Material evolution of buoyancy

The term $\dot{b}$ appearing in the buoyancy equation (25.17c) 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 statified fluid. With a linear equation of state, we can write the material evolution as the convergence of a subgrid scale buoyancy flux

$$\frac{\mathrm{D}b}{\mathrm{D}t} = -\nabla \cdot \boldsymbol{F}^{\mathrm{b}}. \tag{25.23}$$

A particularly common form for this flux is given by

$$\boldsymbol{F}^{\mathrm{b}} = -\kappa \frac{\partial b}{\partial z} \hat{\boldsymbol{z}} + \boldsymbol{v}^* b. \tag{25.24}$$

The first term is a downgradient vertical diffusive flux with the vertical *eddy diffusivity*, $\kappa > 0$, a function of the flow state so that

$$\kappa = \kappa(x, y, z, t). \tag{25.25}$$

The second term is an advective flux, where the advective velocity, $\boldsymbol{v}^* = (\boldsymbol{u}^*, w^*)$, is assumed to be non-divergent

$$\nabla \cdot \boldsymbol{v}^* = \nabla_z \cdot \boldsymbol{u}^* + \frac{\partial w^*}{\partial z} = 0. \tag{25.26}$$

The velocity $\boldsymbol{v}^*$ is commonly termed the *eddy-induced* velocity, with a particular choice for its parameterization discussed in Exercise 25.9 and further examined in Section 49.1.

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

$$\hat{\boldsymbol{n}} \cdot \boldsymbol{F}^{\mathrm{b}} = Q_{\mathrm{surf}} \qquad \text{at } z = \eta(x, y, t) \tag{25.27a}$$

$$(-\hat{\boldsymbol{n}}) \cdot \boldsymbol{F}^{\mathrm{b}} = Q_{\mathrm{bot}} \qquad \text{at } z = \eta_{\mathrm{b}}(x, y) \tag{25.27b}$$

$$\boldsymbol{v}^* \cdot \hat{\boldsymbol{n}} = 0 \qquad \text{all boundaries,} \tag{25.27c}$$

where Q_{surf} is the surface buoyancy flux, Q_{bott} is the bottom buoyancy flux (e.g., geothermal heating), and $\hat{\boldsymbol{n}}$ is the outward normal at the boundaries. Both Q_{surf} and Q_{bot} are positive upward. In Section 50.3 we detail the plethora of processes leading to boundary fluxes of buoyancy.

25.3.2 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*, $\nu^{\mathrm{eddy}} > 0$, in which case the horizontal frictional acceleration in the ocean interior is written as the vertical divergence of a turbulent stress

$$\boldsymbol{F}^h = \frac{1}{\rho_0} \frac{\partial \boldsymbol{\tau}}{\partial z} = \frac{\partial}{\partial z} \left[\nu^{\mathrm{eddy}} \frac{\partial \boldsymbol{u}}{\partial z} \right], \tag{25.28}$$

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 19.6), the eddy viscosity is a function of the flow state, which in turn means that

$$\nu^{\mathrm{eddy}} = \nu^{\mathrm{eddy}}(\boldsymbol{x}, t). \tag{25.29}$$

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

$$\boldsymbol{\tau} = \boldsymbol{\tau}^{\mathrm{wind}} \qquad \text{at } z = \eta \tag{25.30}$$

$$\boldsymbol{\tau} = \boldsymbol{\tau}^{\mathrm{bott}} \qquad \text{at } z = \eta_{\mathrm{b}}(x, y) \tag{25.31}$$

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 19.8), the stress imparted to the ocean is equal and opposite the stress imparted to the atmosphere at its lower boundary. At the ocean bottom, $z = \eta_{\text{b}}$, we often parameterize subgrid scale interactions with bottom topography via a quadratic bottom drag

$$\boldsymbol{\tau}^{\text{bott}} = -C_{\text{D}}\, \boldsymbol{u}\, |\boldsymbol{u}|, \tag{25.32}$$

where $C_{\text{D}} > 0$ 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.

25.4 Mechanical energy analysis part I

The volume of an incompressible fluid element remains materially invariant even as pressure acts on the element, which means that pressure can perform no mechanical work on an incompressible fluid. In turn, an incompressible fluid has no direct 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 and Section 25.5.

In this section we develop the mechanical energy budget for an unforced non-hydrostatic Boussinesq fluid assuming that density is a function only of salinity and Conservative Temperature,

$$\rho = \rho(S, \Theta). \tag{25.33}$$

The more general pressure-dependent equation of state is considered in Section 25.5. Furthermore, we choose to work with the velocity, density, and continuity equations in the form

$$\rho_0 \left[\frac{\mathrm{D}\boldsymbol{v}}{\mathrm{D}t} + 2\,\boldsymbol{\Omega} \wedge \boldsymbol{v} \right] = -(\nabla p + \rho\, \nabla \Phi) \tag{25.34a}$$

$$\frac{\mathrm{D}\rho}{\mathrm{D}t} = \frac{\partial \rho}{\partial S}\, \frac{\mathrm{D}S}{\mathrm{D}t} + \frac{\partial \rho}{\partial \Theta}\, \frac{\mathrm{D}\Theta}{\mathrm{D}t} = \dot{\rho} \tag{25.34b}$$

$$\nabla \cdot \boldsymbol{v} = 0. \tag{25.34c}$$

Note that the geopotential, Φ, is generally a function of space and time, $\Phi = \Phi(\boldsymbol{x}, t)$, which is appropriate when studying astronomical tides or mass inhomogeneities (see Chapter 43).

25.4.1 Kinetic energy

To obtain a kinetic energy equation, start by taking the dot product of the velocity, $\boldsymbol{v}$, and the momentum equation (25.128a) and note that the Coriolis acceleration drops out since it is orthogonal to the velocity

$$\boldsymbol{v} \cdot (2\,\boldsymbol{\Omega} \wedge \boldsymbol{v}) = 0. \tag{25.35}$$

The material time derivative takes the form

$$v_i \left[\frac{\partial v_i}{\partial t} + v_j \partial_j v_i \right] = \frac{\partial \mathcal{K}}{\partial t} + \boldsymbol{v} \cdot \nabla \mathcal{K} = \frac{\mathrm{D}\mathcal{K}}{\mathrm{D}t}, \tag{25.36}$$

where we introduced the kinetic energy per mass

$$\mathcal{K} = \boldsymbol{v} \cdot \boldsymbol{v}/2. \tag{25.37}$$

The equation for the Boussinesq kinetic energy per volume thus takes the form

$$\rho_0 \frac{\mathrm{D}\mathcal{K}}{\mathrm{D}t} = -\boldsymbol{v} \cdot \nabla p - \rho\, \boldsymbol{v} \cdot \nabla \Phi \tag{25.38}$$

Alternatively, we can write this equation in the Eulerian flux-form

$$\frac{\partial(\rho_0\, \mathcal{K})}{\partial t} + \nabla \cdot [\boldsymbol{v}\, (\rho_0\, \mathcal{K} + p)] = -\rho\, \boldsymbol{v} \cdot \nabla \Phi, \tag{25.39}$$

where we used $\nabla \cdot \boldsymbol{v} = 0$ to write $\boldsymbol{v} \cdot \nabla p = \nabla \cdot (\boldsymbol{v}\, p)$. Note that in Exercise 25.3 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 $-\rho\, \boldsymbol{v} \cdot \nabla \Phi$ in the kinetic energy equation (25.39) is a source/sink that arises from fluid motion crossing surfaces of constant geopotential. Moving a fluid element down the geopotential gradient ($\boldsymbol{v} \cdot \nabla \Phi < 0$) 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 $\boldsymbol{v} \cdot \nabla \Phi = \rho\, g\, w$. We have more to say regarding buoyancy work when discussing the mechanical energy budget in Section 25.4.2.

25.4.2 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

$$\frac{\mathrm{D}\Phi}{\mathrm{D}t} = \partial_t \Phi + \boldsymbol{v} \cdot \nabla \Phi. \tag{25.40}$$

The time-dependent geopotential provides an external source of potential energy to the system. Additionally, motion moving up the gradient of the geopotential ($\boldsymbol{v} \cdot \nabla \Phi > 0$) increases the potential energy per mass, and conversely for motion down the geopotential gradient. Adding to the kinetic energy equation (25.38) renders the material evolution

$$\rho_0 \frac{\mathrm{D}\mathcal{K}}{\mathrm{D}t} + \rho \frac{\mathrm{D}\Phi}{\mathrm{D}t} = -\nabla \cdot (\boldsymbol{v}\, p) + \rho\, \partial_t \Phi. \tag{25.41}$$

Note how the buoyancy work source, $\rho\, \boldsymbol{v} \cdot \nabla \Phi$, dropped out from this budget. Consequently, this term provides a reversible transfer of energy between gravitational potential energy per volume and the kinetic energy per volume. We saw the same transfer in Section 23.3.4 when studying the mechanical energy budget for a compressible fluid.

Equation (25.41) has nearly the same form as that for the compressible fluid given by equation (23.43). However, for the Boussinesq fluid it does not lead to a flux-form Eulerian conservation law, even in the perfect fluid limit. To develop a closed Boussinesq mechanical energy budget, add $\Phi\, \dot{\rho}$ to both sides of equation (25.41) to render the material evolution for the mechanical energy *per volume*

$$\frac{\mathrm{D}}{\mathrm{D}t} [\rho_0\, \mathcal{K} + \rho\, \Phi] = -\nabla \cdot (\boldsymbol{v}\, p) + \rho\, \partial_t \Phi + \dot{\rho}\, \Phi, \tag{25.42}$$

which has the Eulerian flux-form expression

$$\partial_t\, (\rho_0\, \mathcal{K} + \rho\, \Phi) + \nabla \cdot [\boldsymbol{v} \cdot (\rho_0\, \mathcal{K} + \rho\, \Phi + p)] = \rho\, \partial_t \Phi + \Phi\, \dot{\rho}. \tag{25.43}$$

In the absence of mixing and boundary processes, the global integrated Boussinesq mechanical energy per volume

$$\rho_0 \, \mathcal{M} = \rho_0 \, \mathcal{K} + \rho \, \Phi \tag{25.44}$$

evolves according to

$$\rho_0 \, \frac{\mathrm{d}}{\mathrm{d}t} \left[\int \mathcal{M} \, \mathrm{d}V \right] = \int \rho \, \partial_t \Phi \, \mathrm{d}V. \tag{25.45}$$

Time changes to the geopotential, due to astronomical forces or changes to the land mass distribution, lead to changes in the mechanical energy. In their absence, and without irreversible and boundary effects, the volume integrated Boussinesq mechanical energy per volume is constant.

25.5 Mechanical energy analysis part II

In this section we build on the analysis from Section 25.4, here allowing density to be a function of salinity, Conservative Temperature, and pressure so that

$$\frac{\mathrm{D}\rho}{\mathrm{D}t} = \frac{\partial \rho}{\partial S} \frac{\mathrm{D}S}{\mathrm{D}t} + \frac{\partial \rho}{\partial \Theta} \frac{\mathrm{D}\Theta}{\mathrm{D}t} + \frac{\partial \rho}{\partial p} \frac{\mathrm{D}p}{\mathrm{D}t}. \tag{25.46}$$

Density thus materially evolves even in the absence of mixing or diabatic processes since $\mathrm{D}p/\mathrm{D}t$ is nonzero whenever flow crosses isobars. Hence, the Eulerian flux-form mechanical energy equation (25.43) now takes on the form

$$\partial_t \left(\rho_0 \, \mathcal{K} + \rho \, \Phi \right) + \nabla \cdot \left[\boldsymbol{v} \cdot \left(\rho_0 \, \mathcal{K} + \rho \, \Phi + p \right) \right] = \rho \, \partial_t \Phi + \Phi \left[\frac{\partial \rho}{\partial S} \, \dot{S} + \frac{\partial \rho}{\partial \Theta} \, \dot{\Theta} + \frac{\partial \rho}{\partial p} \, \dot{p} \right]. \tag{25.47}$$

Pressure dependence to the equation of state leads to a mechanical energy source even in the absence of mixing and diabatic processes. However, with some finesse from *Young* (2010), we can recover a closed Boussinesq mechanical energy budget by modifying the potential energy contribution.

25.5.1 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 ourself with the necessary thermodynamic formalism.

Material time changes to an integral of density over pressure

Consider a thermodynamic function, $\widetilde{\Pi}(S, \Theta, p)$, defined according to the pressure integral of the *in situ* density

$$\widetilde{\Pi}(S, \Theta, p) \equiv - \int_p^{p_\mathrm{r}} \rho(S, \Theta, p') \, \mathrm{d}p' \Longrightarrow \left[\frac{\partial \widetilde{\Pi}}{\partial p} \right]_{S,\Theta} = \rho(S, \Theta, p), \tag{25.48}$$

where p_r is an arbitrary constant pressure. Notably, the integration is taken over pressure in a thermodynamic state space rather than an integral over a region in space.[2] Accordingly, the

[2] This is an example where the discussion in Section 23.2.2 is key, whereby we must distinguish between functions in a thermodynamic state-space versus functions as space-time fields.

infinitesimal increment of $\widetilde{\Pi}$ is given by

$$\delta\widetilde{\Pi} = \delta S \left[\frac{\partial\widetilde{\Pi}}{\partial S}\right]_{\Theta,p} + \delta\Theta \left[\frac{\partial\widetilde{\Pi}}{\partial \Theta}\right]_{S,p} + \rho(S,\Theta,p)\,\delta p, \tag{25.49}$$

and if the increment is computed by following a moving fluid element then we are led to the material time derivative

$$\frac{\mathrm{D}\widetilde{\Pi}}{\mathrm{D}t} = \frac{\mathrm{D}S}{\mathrm{D}t} \left[\frac{\partial\widetilde{\Pi}}{\partial S}\right]_{\Theta,p} + \frac{\mathrm{D}\Theta}{\mathrm{D}t} \left[\frac{\partial\widetilde{\Pi}}{\partial \Theta}\right]_{S,p} + \rho(S,\Theta,p)\,\frac{\mathrm{D}p}{\mathrm{D}t}. \tag{25.50}$$

Material time changes to an integral of density over geopotential

Using the same formalism as above, now consider a thermodynamic function of salinity, Conservative Temperature, and geopotential

$$\Pi(S,\Theta,\Phi) \equiv -\int_{\Phi}^{\Phi_{\mathrm{r}}} \rho(S,\Theta,\Phi')\,\mathrm{d}\Phi' \Longrightarrow \left[\frac{\partial\Pi}{\partial\Phi}\right]_{S,\Theta} = \rho(S,\Theta,\Phi), \tag{25.51}$$

where Φ_{r} is an arbitrary constant geopotential. We offer the following three comments concerning Π.

- For density that is independent of the geopotential, then $\Pi = \rho\,\Phi$ if we take $\Phi_{\mathrm{r}} = 0$. Hence, we can consider $\Pi(S,\Theta,\Phi)$ to be a generalization of the gravitational potential energy per volume.
- Following *Young* (2010), we refer to Π as the *Boussinesq dynamic enthalpy*, though it is beyond our scope to motivate this name, with the reader encouraged to study *Young* (2010) for details.
- One might consider Π to be the difference in hydrostatic pressure between Φ and Φ_{r} (e.g., see equation (18.33)). However, that interpretation only holds if one integrates vertically in space between two geopotentials, with that integration generally crossing surfaces of constant S and Θ. In contrast, the integration in equation (25.48) is taken over the geopotential in a thermodynamic state space. That is, we consider the geopotential to be a replacement for pressure within a Boussinesq thermodynamic state space, with S and Θ held fixed during the integration.

Using the same formalism to derive $\mathrm{D}\widetilde{\Pi}/\mathrm{D}t$ in equation (25.50), we have the material time derivative of the Boussinesq dynamic enthalpy

$$\frac{\mathrm{D}\Pi}{\mathrm{D}t} = \frac{\mathrm{D}S}{\mathrm{D}t} \left[\frac{\partial\Pi}{\partial S}\right]_{\Theta,\Phi} + \frac{\mathrm{D}\Theta}{\mathrm{D}t} \left[\frac{\partial\Pi}{\partial \Theta}\right]_{S,\Phi} + \rho(S,\Theta,\Phi)\,\frac{\mathrm{D}\Phi}{\mathrm{D}t}. \tag{25.52}$$

The final term, $\rho\,\dot{\Phi}$, also appears in the mechanical energy budget, thus motivating us to consider the sum

$$\frac{\mathrm{D}}{\mathrm{D}t}\left[\rho_0\,\mathscr{K} + \Pi\right] = -[\boldsymbol{v}\cdot\nabla p + \rho\,\boldsymbol{v}\cdot\nabla\Phi] + \dot{S}\left[\frac{\partial\Pi}{\partial S}\right]_{\Theta,\Phi} + \dot{\Theta}\left[\frac{\partial\Pi}{\partial \Theta}\right]_{S,\Phi} + \rho\left(\partial_t\Phi + \boldsymbol{v}\cdot\nabla\Phi\right) \tag{25.53a}$$

$$= -\boldsymbol{v}\cdot\nabla p + \rho\,\partial_t\Phi + \dot{S}\left[\frac{\partial\Pi}{\partial S}\right]_{\Theta,\Phi} + \dot{\Theta}\left[\frac{\partial\Pi}{\partial \Theta}\right]_{S,\Phi}, \tag{25.53b}$$

whose Eulerian flux-form expression is given by

$$\partial_t \left(\rho_0 \mathcal{K} + \Pi\right) + \nabla \cdot \left[\boldsymbol{v} \cdot \left(\rho_0 \mathcal{K} + \Pi + p\right)\right] = \rho \, \partial_t \Phi + \dot{S} \left[\frac{\partial \Pi}{\partial S}\right]_{\Theta,\Phi} + \dot{\Theta} \left[\frac{\partial \Pi}{\partial \Theta}\right]_{S,\Phi} . \tag{25.54}$$

The source terms on the right hand side vanish in the absence of mixing and diabatic processes, and with a time-independent geopotential. We have thus succeeded in deriving a closed energy budget for a Boussinesq fluid, with

$$\rho_0 \, \mathcal{M} = \rho_0 \, \mathcal{K} + \Pi \xrightarrow{\rho = \rho(S,\Theta)} \rho_0 \, \mathcal{K} + \rho \, \Phi \tag{25.55}$$

the appropriate Boussinesq form of the mechanical energy per volume.

25.5.2 Comments

Shielding from internal energy

There are further niceties required to unravel the energetics of the oceanic Boussinesq fluid, with full 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 23.3 and 23.4 are relatively straightforward.

The difficulty with compressible energetics arises from the internal energy. Namely, since geophysical flows have speeds that are tiny compared to molecular speeds (see Section 11.2), the mechanical energy associated with geophysical flow is tiny relative to internal energy. 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 us to focus on the Boussinesq mechanical energy arising just from the fluid flow while being shielded from the largely irrelevant internal energy. That shielding offers a strong motivation for making use of the oceanic Boussinesq approximation.

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

$$\rho = \rho[S(\boldsymbol{x}, t), \Theta(\boldsymbol{x}, t), p = -\rho_0 \, g \, z]. \tag{25.56}$$

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

$$\rho = \rho[S(\boldsymbol{x}, t), \Theta(\boldsymbol{x}, t), p = -\rho_0 \, \Phi(\boldsymbol{x}, t)]. \tag{25.57}$$

25.6 Available potential energy

The gravitational potential energy per mass of a fluid element as measured relative to the geopotential, $z = 0$, 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 25.4 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 zero horizontal gradients. 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 34.7.1, 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 25.2.2 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 25.1 illustrates these notions. In the remainder of this section we provide further details to support this conceptual description of APE. Prior to doing so, we take an important diversion in Sections 25.6.1 and 25.6.2 to expose certain kinematic treatments of buoyancy surfaces with particular attention given to boundaries.

Figure 25.1: *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 25.6.3, 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 25.6.2).*

25.6.1 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 25.2 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 horizontal regions where they do not exist; i.e., where they are outcropped or incropped.

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 25.6.3. 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 51). 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 25.2: *Geometry of buoyancy surfaces in a flat bottom box of perfect Boussinesq fluid with height H and horizontal area $\int \mathrm{d}A = A$. 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 25.6.2) is $\overline{\eta(\mathcal{B})} = A^{-1}\int \eta(x, y, \mathcal{B}, t)\,\mathrm{d}A$; and its corresponding anomalous height is $\eta'(x, y, \mathcal{B}, t) = \eta(x, y, \mathcal{B}, t) - \overline{\eta(\mathcal{B})}$. As a complement, a particular point on the $z = \overline{\eta(\mathcal{B})}$ height surface has buoyancy $b = B(x, y, \overline{\eta(\mathcal{B})}, t)$, which then leads to an anomalous buoyancy $b'(x, y, \overline{\eta(\mathcal{B})}, t) = B(x, y, \overline{\eta(\mathcal{B})}, t) - \mathcal{B}$, where $\overline{B(x, y, \overline{\eta(\mathcal{B})}, t)} = \mathcal{B}$ (see equation (25.65)). 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 $\overline{\eta(\mathcal{B}_1)} > \overline{\eta(\mathcal{B}_2)}$ if $\mathcal{B}_1 > \mathcal{B}_2$. When focused on a single buoyancy surface, we can reduce notational clutter by writing, for example, $\overline{\eta}$ rather than $\overline{\eta(\mathcal{B})}$, as well as $b'(\overline{\eta}) = B(\overline{\eta}) - \mathcal{B}$, and $\eta' = \eta - \overline{\eta}$.*

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

$$\overline{\eta(\mathcal{B}, t)}^{\text{buoyancy}} = \frac{\int_{\mathcal{B}} \eta(x, y, \mathcal{B}, t)\,\mathrm{d}\mathcal{S}}{\int_{\mathcal{B}} \mathrm{d}\mathcal{S}}, \tag{25.58}$$

where $\int_{\mathcal{B}} \mathrm{d}\mathcal{S}$ 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

$$\overline{\eta(\mathcal{B})} = \frac{\int \eta(x, y, \mathcal{B}, t)\, \mathrm{d}A}{\int \mathrm{d}A}, \tag{25.59}$$

where

$$A = \int \mathrm{d}A = \int_{\text{fluid domain}} \mathrm{d}x\, \mathrm{d}y \tag{25.60}$$

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 25.6.2 that the area mean height is time-independent for all buoyancy surfaces. Furthermore, this area mean height satisfies the monotonicity property

$$\overline{\eta(\mathcal{B}_1)} > \overline{\eta(\mathcal{B}_2)} \qquad \text{if} \qquad \mathcal{B}_1 > \mathcal{B}_2. \tag{25.61}$$

We make use of both the time-independence of the mean height and the monotonicity property when formulating the available potential energy in Section 25.6.3.

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 $b = \mathcal{B}_{\text{oucrop}}$ in Figure 25.2 an example. Furthermore, let it cover less horizontal area than the full domain area. If we horizontally integrate juse over regions where the surface does not outcrop, but we still normalize by the total horizontal area of the domain, then the area mean height will be less than certain of other buoyancy surfaces whose buoyancy is less and yet whose horizontal area is more, thus breaking the monotonicity property 25.61. 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 contination of the surface within the bottom boundary so that

$$\eta(x, y, \mathcal{B}, t) = \left[\begin{array}{ll} H & \text{if } (x, y) \in \text{outcrop region} \\ 0 & \text{if } (x, y) \in \text{incrop region} \\ \eta(x, y, \mathcal{B}, t) & \text{otherwise.} \end{array} \right. \tag{25.62}$$

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 sequency extends from $0 \leq \overline{\eta(\mathcal{B})} \leq H$.

Analytic continuation of buoyancy stratification at outcrops and incrops

What does the analytic continuation (25.62) 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.

25.6.2 The dual relation between height and buoyancy

In deriving an expression for the APE in Section 25.6.3, we will find it useful to have relations between the unsorted and sorted buoyancy fields and to exploit 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 25.2.

Volume beneath a buoyancy surface using height coordinates

Making use of notation from Figure 25.2, the volume of fluid contained beneath an arbitrary buoyancy surface is

$$V(\mathcal{B}) = \int \mathrm{d}A \int_0^{\eta(x,y,\mathcal{B},t)} \mathrm{d}z = \int \eta(x,y,\mathcal{B},t)\,\mathrm{d}A = A\,\overline{\eta(\mathcal{B})}, \tag{25.63}$$

The following properties result from volume conservation in a perfect incompressible fluid (see Chapter 16).

- 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 t argument from $V(\mathcal{B})$ and $\overline{\eta(\mathcal{B})}$ in equation (25.63).
- 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 (25.63) 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

$$\overline{B(z)} = A^{-1} \int B(x,y,z,t)\,\mathrm{d}A = \text{time independent}. \tag{25.64}$$

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 $\mathcal{B}$

$$\overline{B[\overline{\eta(\mathcal{B})}]} = A^{-1} \int B(x, y, z = \overline{\eta(\mathcal{B})}, t)\, \mathrm{d}A = \mathcal{B}. \tag{25.65}$$

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

$$V(\mathcal{B}) = \int \mathrm{d}A \int_0^{\eta(x,y,\mathcal{B},t)} \mathrm{d}z = \int \mathrm{d}A \int_{b(x,y,0,t)}^{\mathcal{B}} \frac{\partial z}{\partial b}\, \mathrm{d}b = \int \mathrm{d}A \int_{b(x,y,0,t)}^{\mathcal{B}} \frac{\mathrm{d}b}{N^2} \tag{25.66}$$

where $b(x, y, 0, t)$ is the buoyancy at the bottom of the domain and

$$N^2 = \frac{\partial b}{\partial z} \tag{25.67}$$

is the squared buoyancy frequency. As noted in Section 25.6.1 and illustrated in Figure 25.2, we analytically continue the buoyancy surfaces into the surface and bottom boundaries so that

$$N^{-2}(x, y, \mathcal{B}) = \left[\begin{array}{ll} = 0 & \text{if } \mathcal{B} > b(x, y, H) \text{ (surface outcrop region)} \\ = 0 & \text{if } \mathcal{B} < b(x, y, 0) \text{ (bottom incrop region)} \\ = N^{-2}(x, y, \mathcal{B}) & \text{if } b(x, y, 0) \le \mathcal{B} \le b(x, y, H). \end{array} \right. \tag{25.68}$$

In this manner we can replace the lower limit in equation (25.66) with a buoyancy below any found in the domain, which we write as b_{min} so that

$$V(\mathcal{B}) = \int \mathrm{d}A \int_{b_{\mathrm{min}}}^{\mathcal{B}} \frac{\mathrm{d}b}{N^2} = \int_{b_{\mathrm{min}}}^{\mathcal{B}} \mathrm{d}b \int \frac{\mathrm{d}A}{N^2}. \tag{25.69}$$

Being able to commute the area and buoyancy integrals proves useful in the following.

25.6.3 Exact expression for APE

In this section we develop an expression for the APE of the perfect stably stratified Boussinesq fluid in a box-domain. To start, we consider the gravitational potential energy of the fluid in Figure 25.2, relative to a constant density background state with $\rho = \rho_0$

$$\mathcal{P} = -\int \mathrm{d}A \int_0^H b\, z\, \mathrm{d}z = -\frac{1}{2} \int \mathrm{d}A \int_0^H b\, \mathrm{d}(z^2). \tag{25.70}$$

Integration by parts leads to the equivalent expression

$$\mathcal{P} = -\frac{1}{2} \int \mathrm{d}A \int_0^H \mathrm{d}(b\, z^2) + \frac{1}{2} \int \mathrm{d}A \int_{b(x,y,0)}^{b(x,y,H)} \eta^2(x, y, b)\, \mathrm{d}b \tag{25.71}$$

$$= -\frac{A\, H^2}{2}\, \overline{b(H)} + \frac{1}{2} \int \mathrm{d}A \int_{b(x,y,0)}^{b(x,y,H)} \eta^2(x, y, b)\, \mathrm{d}b, \tag{25.72}$$

where $\overline{b(H)}$ is the area averaged buoyancy at the top of the fluid domain, $z = H$. As discussed in Section 25.6.1, 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 (25.72) 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

$$\mathscr{P} = \frac{A}{2}\left[-H^2\, b_{\max} + \int_{b_{\min}}^{b_{\max}} \overline{\eta^2(b)}\, \mathrm{d}b\right], \tag{25.73}$$

where $\overline{\eta^2(b)}$ is the area mean of the squared height of a buoyancy surface. The same calculation for the reference buoyancy, $b_{\mathrm{ref}}(z)$, leads to

$$\mathscr{P}_{\mathrm{ref}} = \frac{A}{2}\left[-H^2\, b_{\max} + \int_{b_{\min}}^{b_{\max}} \overline{\eta(b)}^2\, \mathrm{d}b\right], \tag{25.74}$$

where we noted that the height of a reference buoyancy surface equals to the area mean of the corresponding buoyancy surface

$$\eta(b_{\mathrm{ref}}) = \overline{\eta}(b = b_{\mathrm{ref}}), \tag{25.75}$$

and the reference buoyancy at the surface boundary equals to the maximum buoyancy, $b_{\mathrm{ref}}(H) = b_{\max}$.

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

$$\mathscr{P}_{\mathrm{APE}} = \mathscr{P} - \mathscr{P}_{\mathrm{ref}} = \frac{A}{2}\left[\int_{b_{\min}}^{b_{\max}} [\overline{\eta^2(b)} - \overline{\eta(b)}^2]\, \mathrm{d}b\right] = \int_{b_{\min}}^{b_{\max}} \overline{(\eta')^2}\, \mathrm{d}b \geq 0, \tag{25.76}$$

where (see Figure 25.2)

$$\eta(x, y, b, t) = \overline{\eta(b)} + \eta'(x, y, b, t). \tag{25.77}$$

The positive definite nature of the APE arises since either a positive or negative buoyancy surface undulation gives rise to motion.

Equation (25.76) 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 30.7. It is the natural expression when working in buoyancy coordinates, whereby the APE is determined by height variations of constant buoyancy surfaces.

25.6.4 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 25.2 if confused by signs)

$$\eta(\mathcal{B}) \approx \overline{\eta(\mathcal{B})} + \left[\frac{\partial z}{\partial b}\right]_{z=\overline{\eta}} [\mathcal{B} - B(\overline{\eta})] = \overline{\eta} - \frac{b'}{N^2} \approx \overline{\eta} - \frac{b'}{N_{\mathrm{ref}}^2}, \tag{25.78}$$

where the final step set $N^2(x, y, \overline{\eta}) \approx N_{\mathrm{ref}}^2(\overline{\eta})$, which is valid to the same order as the approximation. We are thus lead to the approximate expression

$$\eta' = \eta - \overline{\eta} \approx -\frac{b'}{N_{\mathrm{ref}}^2} \tag{25.79}$$

so that the APE is given approximately by

$$\mathscr{P}_{\rm APE} \approx A \left[\int_0^H \frac{\overline{(b')^2}}{2\, N_{\rm ref}^2} \, {\rm d}z \right] , \tag{25.80}$$

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 25.3 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 (25.80).

Figure 25.3: *Illustrating how to sort buoyancy to determine the background 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 $b_{\rm ref}(z)$. 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

$$b(x, y, z, t) = b_{\rm ref}(z) + b'(x, y, z, t), \tag{25.81}$$

so that the perfect fluid buoyancy equation takes on the form

$$\frac{\mathrm{D}b}{\mathrm{D}t} = 0 \Longrightarrow \frac{\mathrm{D}b'}{\mathrm{D}t} = -w\, N_{\mathrm{ref}}^2. \tag{25.82}$$

Multiplying by b' leads to

$$\frac{\mathrm{D}[(b')^2/2]}{\mathrm{D}t} = -w\, b'\, N_{\mathrm{ref}}^2, \tag{25.83}$$

and then dividing by N_{ref}^2 renders

$$\frac{\mathrm{D}P_{\mathrm{APE}}}{\partial t} + \boldsymbol{v}\cdot\nabla P_{\mathrm{APE}} = -w\, b'\left[1 - \frac{b'}{2}\,\frac{\partial(1/N_{\mathrm{ref}}^2)}{\partial z}\right], \tag{25.84}$$

where we defined the approximate APE per unit volume

$$P_{\mathrm{APE}} = [b'/(2\, N_{\mathrm{ref}})]^2. \tag{25.85}$$

In the case of 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 (25.84) to the kinetic energy equation (25.39) leads to

$$\frac{\partial(\mathcal{K} + P_{\mathrm{APE}})}{\partial t} + \boldsymbol{v}\cdot\nabla(\mathcal{K} + P_{\mathrm{APE}} + \varphi) = w\left[b_{\mathrm{ref}} + \frac{(b')^2}{2}\,\frac{\partial(1/N_{\mathrm{ref}}^2)}{\partial z}\right]. \tag{25.86}$$

It is notable that an area average eliminates the $w\, b_{\mathrm{ref}}$ term since

$$\overline{w} = A^{-1}\int w\,\mathrm{d}A = 0, \tag{25.87}$$

which follows from $\nabla\cdot\boldsymbol{v} = 0$ (see Exercise 16.5). Further simplifications arise with depth-independent N_{ref}, with the corresponding space-time spectra studied by *Bühler et al.* (2014).

25.6.5 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.

25.7 Stratification effects on hydrostatic scaling

Stratification impacts how the hydrostatic equation scales. To include stratification requires us to include the thermodynamics equation, which for the Boussinesq system is the evolution of buoyancy. It is also sufficient to ignore irreversible effects and rotation, in which case the Boussinesq equations take on the form

$$\frac{\mathrm{D}\boldsymbol{u}}{\mathrm{D}t} = -\nabla_z\varphi \tag{25.88}$$

$$\frac{\mathrm{D}w}{\mathrm{D}t} = -\frac{\partial\varphi}{\partial z} + b \tag{25.89}$$

$$\nabla\cdot\boldsymbol{v} = 0 \tag{25.90}$$

$$\frac{\mathrm{D}b}{\mathrm{D}t} = 0. \tag{25.91}$$

To help isolate the dynamically important portion of stratification, we proceed much like in Section 25.2.1 whereby buoyancy is written

$$b = b'(x, y, z, t) + \tilde{b}(z). \tag{25.92}$$

The static buoyancy $\tilde{b}(z)$ encompasses a background stratification that is in hydrostatic balance with its corresponding portion of the pressure field

$$\frac{\partial \tilde{\varphi}}{\partial z} = \tilde{b}(z). \tag{25.93}$$

The Boussinesq equations thus take the form

$$\frac{\mathrm{D}\boldsymbol{u}}{\mathrm{D}t} = -\nabla_z \varphi' \tag{25.94}$$

$$\frac{\mathrm{D}w}{\mathrm{D}t} = -\frac{\partial \varphi'}{\partial z} + b' \tag{25.95}$$

$$\nabla \cdot \boldsymbol{v} = 0 \tag{25.96}$$

$$\frac{\mathrm{D}b'}{\mathrm{D}t} = -w\, N^2, \tag{25.97}$$

where

$$N^2 = \frac{\mathrm{d}\tilde{b}}{\mathrm{d}z} \tag{25.98}$$

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.

Now introduce the dimensional scales and corresponding non-dimensional quantities

$$(x, y) = L\,(\hat{x}, \hat{y}) \qquad z = H\,\hat{z} \qquad \boldsymbol{u} = U\,\hat{\boldsymbol{u}} \qquad w = W\,\hat{w} \tag{25.99}$$

$$t = T\,\hat{t} \qquad \varphi' = \Phi\,\hat{\varphi}' \qquad b' = B\,\hat{b}' \qquad N^2 = \overline{N}^2\,\hat{N}^2. \tag{25.100}$$

As for the previous case in Section 24.3.2, we impose our subjective regime choices to help constrain the dimensional variables. In addition to those choices considered in the absence of stratification, we acknowledge that the vertical velocity will likely be reduced in the presence of stratification, given that vertical stratification acts to suppress vertical motion. We thus introduce a non-dimensional number ϵ so that

$$w = W\,\hat{w} = \epsilon \left[\frac{HU}{L}\right] \hat{w}. \tag{25.101}$$

Inserting into the stratified Boussinesq equations then leads to the non-dimensional system

$$\frac{\mathrm{D}\hat{u}}{\mathrm{D}\hat{t}} = -\hat{\nabla}\hat{\varphi}' \tag{25.102}$$

$$\epsilon\, \alpha^2_{\text{aspect}} \frac{\mathrm{D}\hat{w}}{\mathrm{D}\hat{t}} = -\frac{\partial \hat{\varphi}'}{\partial \hat{z}} + \hat{b}' \tag{25.103}$$

$$\left[\frac{U^2}{\overline{N}^2\, H^2}\right] \frac{\hat{\mathrm{D}}\hat{b}'}{\mathrm{D}\hat{t}} + \epsilon\, \hat{N}^2\, \hat{w} = 0 \tag{25.104}$$

$$\hat{\nabla} \cdot \hat{\boldsymbol{u}} + \epsilon \frac{\partial \hat{w}}{\partial \hat{z}} = 0 \tag{25.105}$$

where we introduced the non-dimensional material time derivative

$$\frac{\mathrm{D}}{\mathrm{D}\hat{t}} = \frac{\partial}{\partial \hat{t}} + \hat{\boldsymbol{u}} \cdot \hat{\nabla}_z + \epsilon\, \hat{w}\, \frac{\partial}{\partial \hat{z}}. \tag{25.106}$$

At this point we make a choice for the parameter ϵ. There are many choices that one could consider. For our interests it is suitable to set ϵ equal to the squared Froude number

$$\epsilon = \mathrm{Fr}^2 = \frac{U^2}{\overline{N}^2\, H^2}. \tag{25.107}$$

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 $Fr > 1$ a common indicator of hydraulic instability (see Exercise 29.3 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

$$\mathrm{Ri} = \mathrm{Fr}^{-2} = \frac{\overline{N}^2\, H^2}{U^2}. \tag{25.108}$$

It is a matter of taste whether one works with Fr or Ri.

With this choice the vertical velocity scale is given by

$$W = \mathrm{Fr}^2 \left[\frac{HU}{L}\right]. \tag{25.109}$$

We see that for $\mathrm{Fr} < 1$ that the stratification reduces the scale for the vertical velocity. With this choice for ϵ, the non-dimensional Boussinesq equations take the form

$$\frac{\mathrm{D}\hat{u}}{\mathrm{D}\hat{t}} = -\hat{\nabla}_z \hat{\varphi}' \tag{25.110}$$

$$\mathrm{Fr}^2\, \alpha_{\text{aspect}}^2\, \frac{\mathrm{D}\hat{w}}{\mathrm{D}\hat{t}} = -\frac{\partial \hat{\varphi}'}{\partial \hat{z}} + \hat{b}' \tag{25.111}$$

$$\frac{\hat{\mathrm{D}}\hat{b}'}{\mathrm{D}\hat{t}} + \hat{N}^2\, \hat{w} = 0 \tag{25.112}$$

$$\hat{\nabla} \cdot \hat{\boldsymbol{u}} + \mathrm{Fr}^2\, \frac{\partial \hat{w}}{\partial \hat{z}} = 0. \tag{25.113}$$

The condition for hydrostatic balance in a stratified fluid thus takes the form

$$\mathrm{Fr}^2\, \alpha_{\text{aspect}}^2 \ll 1. \tag{25.114}$$

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

$$\left|\hat{\nabla} \cdot \hat{\boldsymbol{u}}\right| = \left|\mathrm{Fr}^2\, \frac{\partial \hat{w}}{\partial \hat{z}}\right| \ll \left|\frac{\partial \hat{w}}{\partial \hat{z}}\right|. \tag{25.115}$$

25.8 Exercises

EXERCISE 25.1: 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, $\overline{\rho}(z)$, 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

$$\rho(\boldsymbol{x}, t) = \rho_0 + \delta\rho(\boldsymbol{x}, t), \tag{25.116}$$

thus leading to the Boussinesq momentum equation

$$\frac{\mathrm{D}\boldsymbol{v}}{\mathrm{D}t} + 2\,\boldsymbol{\Omega} \wedge \boldsymbol{v} = -\rho_0^{-1}\,(\nabla p + \hat{\boldsymbol{z}}\,g\,\rho). \tag{25.117}$$

But rather than take the traditional decomposition of the pressure and gravitational terms, we now write

$$\nabla p + \hat{\boldsymbol{z}}\,g\,\rho = \nabla[p - \overline{p}(z) + \overline{p}(z)] + \hat{\boldsymbol{z}}\,g\,[\rho - \overline{\rho}(z) + \overline{\rho}(z)] \tag{25.118a}$$

$$= \nabla[p - \overline{p}(z)] + \hat{\boldsymbol{z}}\,g\,[\rho - \overline{\rho}(z)]. \tag{25.118b}$$

This step introduced the density, $\overline{\rho}(z)$, and the corresponding hydrostatically balanced pressure

$$\frac{\mathrm{d}\overline{p}(z)}{\mathrm{d}z} = -g\,\overline{\rho}(z). \tag{25.119}$$

(a) Show that the pressure is decomposed as

$$p(\boldsymbol{x}, t) = [p(\boldsymbol{x}, t) - \overline{p}(z)] + \overline{p}(z) = \rho_0\,\tilde{\varphi} + \overline{p}(z). \tag{25.120}$$

What is $\tilde{\varphi}$?

(b) Introduce a buoyancy

$$\tilde{b} = -\frac{g\,(\rho - \overline{\rho})}{\rho_0}, \tag{25.121}$$

defined relative to the depth-dependent background density, $\overline{\rho}(z)$, rather than the globally constant density ρ_0. Show that the momentum equation is given by

$$\frac{\mathrm{D}\boldsymbol{v}}{\mathrm{D}t} + 2\,\boldsymbol{\Omega} \wedge \boldsymbol{v} = -\nabla\tilde{\varphi} + \hat{\boldsymbol{z}}\,\tilde{b}. \tag{25.122}$$

(c) Show that the baroclinicity vector appearing in the Boussinesq vorticity equation takes the form

$$\tilde{\boldsymbol{B}} = \nabla\tilde{b} \wedge \hat{\boldsymbol{z}}, \tag{25.123}$$

which is mathematically the same as with the traditional Boussinesq approximation from Section 25.2.2. We have thus succeeded in generalizing the pressure decomposition and buoyancy field, yet without corrupting the familiar Boussinesq vorticity dynamics.

EXERCISE 25.2: POISSON EQUATION FOR PRESSURE

Consider a perfect Boussinesq fluid that is rotating on a β-plane and thus satisfies the governing equations

$$\frac{\mathrm{D}\boldsymbol{v}}{\mathrm{D}t} + f\,\hat{\boldsymbol{z}} \wedge \boldsymbol{v} = -\nabla\varphi + \hat{\boldsymbol{z}}\,b \tag{25.124a}$$

$$\frac{\mathrm{D}b}{\mathrm{D}t} = 0 \tag{25.124b}$$

$$\nabla \cdot \boldsymbol{v} = 0 \tag{25.124c}$$

$$f = f_0 + \beta\,y. \tag{25.124d}$$

(a) Show that the dynamic pressure φ satisfies the following elliptic boundary value problem (see Section 3.4 for discussion of elliptic partial differential equations)

$$-\nabla^2\varphi = \nabla \cdot \boldsymbol{G} \qquad \text{within the domain} \tag{25.125a}$$

$$-\hat{\boldsymbol{n}} \cdot \nabla\varphi = \hat{\boldsymbol{n}} \cdot \boldsymbol{G} \qquad \text{on the domain boundaries.} \tag{25.125b}$$

(b) Determine the vector $\boldsymbol{G}$.

(c) Show that the contribution to $\nabla \cdot \boldsymbol{G}$ from velocity self-advection [i.e., $(\boldsymbol{v} \cdot \nabla)\,\boldsymbol{v}$] is given by

$$[\nabla \cdot \boldsymbol{G}]_{\text{self-advection}} = \mathbb{S}_{mn}\,\mathbb{S}_{mn} - \mathbb{A}_{mn}\,\mathbb{A}_{mn}, \tag{25.126}$$

with $\mathbb{S}_{mn}$ components to the rate of strain tensor and $\mathbb{A}_{mn}$ components to the rotation tensor, both defined in Section 13.2.4. Hence, strain provides a positive source to $-\nabla^2\varphi$ whereas vorticity is a negative source.

Hint: a similar boundary value problem holds for the two-dimensional barotropic fluid examined in Section 39.4.

EXERCISE 25.3: 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,

$$\mathcal{K}_{\text{horz}} = \frac{\boldsymbol{u} \cdot \boldsymbol{u}}{2}, \tag{25.127}$$

satisfies the same equation as that for a non-hydrostatic fluid given by equation (25.38). Assume the simple form for the geopotential, $\Phi = g\,z$.

EXERCISE 25.4: ENERGETICS FOR A PERFECT BOUSSINESQ FLUID

In Section 25.4 we developed the energetic balances for an ideal Boussinesq fluid. We here rederive the same energetics but using the the momentum, buoyancy, and continuity equations in the form that exposes buoyancy

$$\frac{\mathrm{D}\boldsymbol{v}}{\mathrm{D}t} + 2\,\boldsymbol{\Omega} \wedge \boldsymbol{v} = -\nabla\varphi + \hat{\boldsymbol{z}}\,b \tag{25.128a}$$

$$\frac{\mathrm{D}b}{\mathrm{D}t} = 0 \tag{25.128b}$$

$$\nabla \cdot \boldsymbol{v} = 0 \tag{25.128c}$$

$$b = b(S, \Theta). \tag{25.128d}$$

Additionally, we here assume the simple form for the geopotential, $\Phi = g\,z$.

(a) Derive the kinetic energy budget.

(b) Derive the gravitational potential energy budget. Interpret the gravitational potential energy.

(c) Derive the mechanical energy budget. Discuss.

EXERCISE 25.5: 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 system are

$$\frac{\mathrm{D}\boldsymbol{v}}{\mathrm{D}t} + f\,\hat{\boldsymbol{z}} \wedge \boldsymbol{v} = -\nabla\varphi + \hat{\boldsymbol{z}}\,b + \nu\,\nabla^2\,\boldsymbol{v} \tag{25.129a}$$

$$\frac{\mathrm{D}b}{\mathrm{D}t} = Q^{\mathrm{b}} + \kappa\,\nabla^2 b \tag{25.129b}$$

$$\nabla\cdot\boldsymbol{v} = 0. \tag{25.129c}$$

In the above, ν is the viscosity leading to irreversible mixing of velocity, and κ is the diffusivity leading to irreversible mixing of buoyancy. Assume both 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 (25.17f) with haline contraction coefficient $\beta = 0$). Finally, Q^{b} is a boundary buoyancy forcing associated with irreversible changes in heat (e.g., boundary fluxes and/or penetrative shortwave radiation).

Hint: Elements of this exercise are discussed in Chapter 21 of *Vallis* (2017).

(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

$$\nabla^2\boldsymbol{v} = -\nabla\wedge\boldsymbol{\omega} \qquad \text{if } \nabla\cdot\boldsymbol{v} = 0 \tag{25.130}$$

to express the viscosity contributions in terms of the vorticity

$$\boldsymbol{\omega} = \nabla\wedge\boldsymbol{v}. \tag{25.131}$$

(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 $P^{\mathrm{b}} = -g^{-1}\,\Phi\,b$ 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 must occur at a lower level (lower gravitational geopotential) than cooling if a kinetic energy dissipating circulation is to be maintained.

EXERCISE 25.6: 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 25.2.6. We here assume the horizontal frictional acceleration is determined by vertical viscous friction in equation (25.28), and the stress boundary conditions are given by equations (25.30) and (25.32).

(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 (25.32).

EXERCISE 25.7: 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 25.2.6.

(a) Derive the Eulerian flux-form budget for gravitational potential energy written as $P^{\mathrm{b}} = -g^{-1}\,\Phi\,b$ 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 (25.24). 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 (25.24). 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 25.8: 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 25.2.6.

(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 25.9: 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 25.2.6. Namely, we consider the specific form for the parameterized eddy-induced velocity proposed by *Gent et al.* (1995)

$$\boldsymbol{u}^* = -\partial_z(B\,\boldsymbol{S}) \tag{25.132a}$$

$$w^* = \nabla_z \cdot (B\,\boldsymbol{S}) \tag{25.132b}$$

$$\boldsymbol{S} = -\frac{\nabla_z b}{N^2} \tag{25.132c}$$

$$\boldsymbol{v}^* \cdot \hat{\boldsymbol{n}} = 0 \qquad \text{at all ocean boundaries.} \tag{25.132d}$$

In this expression, $B > 0$ is an eddy diffusivity. To ensure $\boldsymbol{v}^* \cdot \hat{\boldsymbol{n}} = 0$ at all domain boundaries requires that $B = 0$ along these boundaries. The horizontal vector $\boldsymbol{S} = (S^{(x)}, S^{(y)}, 0)$ 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 $\boldsymbol{\Psi}^*$ such that

$$\boldsymbol{v}^* = \nabla \wedge \boldsymbol{\Psi}^*. \tag{25.133}$$

Choose the gauge with $\hat{\boldsymbol{z}} \cdot \boldsymbol{\Psi}^* = 0$.

(b) Show that

$$\int_{-H}^{\eta} \boldsymbol{u}^* \, \mathrm{d}z = 0. \tag{25.134}$$

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

$$\mathcal{B}^{(y)}(y,t) = \int_{x_1}^{x_2} \mathrm{d}x \int_{-H}^{\eta} b\,(v + v^*)\, \mathrm{d}z. \tag{25.135}$$

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 $\boldsymbol{v}^*$ to the buoyancy equation (25.17c) affect the global integrated *gravitational potential energy*? Discuss.

(e) How does the introduction of $\boldsymbol{v}^*$ to the buoyancy equation (25.17c) affect the global integrated *available potential energy*? Discuss.

26

Geostrophic mechanics

Fluid motion dominated by rotation is characterized by a small Rossby number. Inviscid flow with a small Rossby number maintains the geostrophic balance, which is a balance between the Coriolis acceleration and the pressure gradient acceleration. In this chapter, we introduce salient features of geostrophically balanced flow and the associated thermal wind shear. The associated diagnostic relations involve no time derivatives, and so cannot be used to predict the evolution of the fluid flow. However, their power for diagnostics is unquestioned as they provide a robust framework for interpreting the large-scale circulation of the atmosphere and ocean.

READER'S GUIDE TO THIS CHAPTER

This chapter assumes an understanding of the primitive equations from Chapter 24 and the Coriolis acceleration from Chapter 9. The material in this chapter is fundamental to understanding the nature and mechanisms of large-scale flow in the atmosphere and ocean. We make much use of this chapter in the remainder of this book. Throughout this discussion we are not explicitly concerned with sphericity, thus enabling the use of Cartesian coordinates.

26.1 Primitive equations . . . **460**
26.2 The Rossby number . . . **460**
26.2.1 Scaling for the Rossby number . . . 461
26.2.2 Ratio of material acceleration to Coriolis acceleration . . . 461
26.2.3 Ratio of local time tendency to Coriolis acceleration . . . 461
26.2.4 Rossby number for a kitchen sink . . . 462
26.2.5 Rossby number for a Gulf Stream ring . . . 462
26.3 Geostrophic balance . . . **462**
26.3.1 Geostrophic balance is distinctly fluid mechanical . . . 463
26.3.2 Geostrophic relation in geopotential coordinates . . . 463
26.3.3 Cyclonic and anti-cyclonic orientation . . . 463
26.3.4 Density gradients and thermal wind shear . . . 464
26.3.5 Geostrophic relation in pressure coordinates . . . 465
26.3.6 Further study . . . 465
26.4 Introducing planetary geostrophy . . . **465**
26.4.1 Planetary geostrophic equations . . . 465
26.4.2 Planetary geostrophic vorticity equation . . . 466
26.4.3 Taylor-Proudman and vertical stiffening . . . 467
26.4.4 Meridional motion in response to vortex stretching and wind curls . . . 468

26.4.5 Thermal wind balance for the ocean 469
26.4.6 Thermal wind balance for the atmosphere 471
26.4.7 Further study 472
26.5 Isopycnal form stress from geostrophic eddies **473**
26.5.1 Zonal mean zonal form stress on an isopycnal surface 473
26.5.2 Zonal mean zonal form stress acting on an isopycnal layer 475
26.5.3 Comments and further study 477
26.6 Exercises **477**

26.1 Primitive equations

Throughout this chapter, we make use of the hydrostatic primitive equations derived in Section 24.1

$$\rho \left[\frac{\partial}{\partial t} + \boldsymbol{v} \cdot \nabla + f \, \hat{\boldsymbol{z}} \wedge \right] \boldsymbol{u} = -\nabla_z p \tag{26.1a}$$

$$\frac{\partial p}{\partial z} = -g \, \rho \tag{26.1b}$$

$$\frac{\mathrm{D}\rho}{\mathrm{D}t} = \rho \, \nabla \cdot \boldsymbol{v}, \tag{26.1c}$$

where the velocity vector is written using Cartesian coordinates

$$\boldsymbol{v} = \boldsymbol{u} + \hat{\boldsymbol{z}} \, w = \hat{\boldsymbol{x}} \, u + \hat{\boldsymbol{y}} \, v + \hat{\boldsymbol{z}} \, w, \tag{26.2}$$

and the horizontal gradient operator is

$$\nabla_z = \hat{\boldsymbol{x}} \, \partial_x + \hat{\boldsymbol{y}} \, \partial_y. \tag{26.3}$$

For some of the scaling analysis we assume an incompressible fluid as per the Boussinesq approximation (Section 25.2). In this case, the mass continuity equation (26.1c) becomes the non-divergent condition on the velocity

$$\nabla \cdot \boldsymbol{v} = 0. \tag{26.4}$$

Furthermore, ρ in the horizontal momentum equation (26.1a) is converted to a constant reference density ρ_0. However, it remains the full density when appearing in the hydrostatic equation due to being multiplied by the gravitational acceleration.

26.2 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 motion of the ocean and atmosphere. There are two points to emphasize in this regard. First, much of the atmospheric motion is close to solid-body rotation, in which weather patterns are best viewed relative to the rotating sphere rather than relative to the "fixed" stars. Second, human scale horizontal length scales are generally far too small to take 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.

26.2.1 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 scales for the acceleration by writing

$$\frac{\partial \boldsymbol{u}}{\partial t} + (\boldsymbol{v} \cdot \nabla)\, \boldsymbol{u} \sim \frac{U}{T} + \frac{U^2}{L} + \frac{WU}{H}, \tag{26.5}$$

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 24.3). Likewise, the Coriolis acceleration scales as

$$f\, \hat{\mathbf{z}} \wedge \mathbf{u} \sim f_0\, U, \tag{26.6}$$

where $f_0 > 0$ is the scale for the Coriolis parameter. From the continuity equation for incompressible flow ($\nabla \cdot \boldsymbol{v} = 0$) we see that the vertical and horizontal velocity scales are related by

$$\frac{W}{H} \sim \frac{U}{L} \Rightarrow W \sim U\, \frac{H}{L}. \tag{26.7}$$

For compressible hydrostatic flows, we replace W with the scale for motion across pressure surfaces. In either the incompressible or compressible case, we assume the vertical to horizontal grid aspect ratio is small

$$\frac{H}{L} \ll 1, \tag{26.8}$$

as per the hydrostatic balance discussed in Section 24.3. Consequently, the vertical velocity scale is much less than the horizontal

$$W \ll U. \tag{26.9}$$

26.2.2 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

$$Ro = \frac{U^2/L}{f_0\, U} = \frac{U}{f_0\, L}. \tag{26.10}$$

As the length and velocity scales are not constant, the Rossby number is a function of more than just the latitude. Nonetheless, due to the latitudinal variation of the Coriolis parameter, the Rossby number is generally small near the poles and large in the tropics.

26.2.3 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

$$Ro = \frac{U/T}{U f_0} = \frac{1/T}{f_0}. \tag{26.11}$$

Thus, for motions that have a low frequency T^{-1} compared to the *rotational inertial frequency* f_0^{-1}, 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 on the dynamics.

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 $U \sim 0.01 - 0.1$ m s^{-1} whereas Gulf Stream rings have velocity scales on the order $U \sim 0.1 - 1.0$ m s^{-1}. However, their length scales are very different, with the scale for a sink $L \sim 1$m yet the Gulf Stream rings $L \sim 10^5$m. Taking the Coriolis parameter to be at 30° for both flows leads to $Ro_{sink} \sim 10^2 - 10^3$ and $Ro_{ring} \sim 10^{-2} - 10^{-1}$. The Coriolis acceleration is clearly important for the Gulf Stream rings whereas it is utterly negligible for the kitchen sink.*

26.2.4 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

$$Ro_{\text{sink}} \approx 10^2 - 10^3. \tag{26.12}$$

The Coriolis force is therefore negligible for kitchen sink fluid dynamics. This result explains the difficulty of experimentally correlating the hemisphere to a preferred rotational direction of water leaving a sink drain.

26.2.5 Rossby number for a Gulf Stream ring

For a Fulf 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

$$Ro_{\text{ring}} \approx 10^{-2} - 10^{-1}, \tag{26.13}$$

thus indicating the importance of the Coriolis acceleration for dynamics of Gulf Stream rings.

26.3 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 $-f\,\hat{\mathbf{z}} \wedge \mathbf{u}$. In a manner directly analogous to the Lorentz force in electrodynamics, the Coriolis acceleration acts perpendicular to the fluid motion

$$\boldsymbol{u} \cdot (\hat{\boldsymbol{z}} \wedge \boldsymbol{u}) = 0. \tag{26.14}$$

Hence, the Coriolis force effects the fluid motion but does not alter its kinetic energy; i.e., it does zero work on the fluid. In the northern hemisphere where $f > 0$, the Coriolis force 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 26.2). In the southern hemisphere, where $f < 0$, it acts in the opposite direction.

Figure 26.2: *Geostrophic motion around low and high pressure centers in the northern hemisphere and southern hemispheres ($f = 2\Omega \sin\phi > 0$ in the north and $f < 0$ in the south). Upper panel: the counter-clockwise motion around the low pressure center 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.*

26.3.1 Geostrophic balance is distinctly fluid mechanical

When pressure and Coriolis forces balance, parcel motion is said to be in *geostrophic* balance, whereby large-scale winds and currents generally follow isobars (lines of constant pressure). Recall from Chapter 9 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.

26.3.2 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 pressure gradient acceleration

$$f\,\hat{\mathbf{z}} \wedge \mathbf{u}_{\mathrm{g}} = -\rho^{-1}\,\nabla_z p, \tag{26.15}$$

or equivalently, we have the expression for the geostrophic velocity[1]

$$\mathbf{u}_{\mathrm{g}} = \frac{\hat{\boldsymbol{z}} \wedge \nabla p}{f\rho} \Longrightarrow u_{\mathrm{g}} = -\frac{1}{f\rho}\frac{\partial p}{\partial y} \qquad \text{and} \qquad v_{\mathrm{g}} = \frac{1}{f\rho}\frac{\partial p}{\partial x}. \tag{26.16}$$

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.

26.3.3 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

[1] We can write either ∇ or ∇_z in equation (26.16). The reason is that the $\hat{\boldsymbol{z}} \wedge$ operator selects only the horizontal portion of the gradient.

Figure 26.3: *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 24.1, which discusses how to compute horizontal pressure gradients in a hydrostatic fluid.*

example, geostrophic motion around a low pressure center in the northern hemisphere is counter-clockwise (Figure 26.2). Using the right hand rule, this motion represents a positively oriented rotation. Hence, with $f > 0$ in the north, counter-clockwise motion is cyclonic. Similarly in the south, 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 also represents cyclonic motion (Figure 26.2).

26.3.4 Density gradients and thermal wind shear

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

$$\frac{\partial(\nabla_z p)}{\partial z} = -g\nabla_z \rho. \tag{26.17}$$

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.

We illustrate this depth dependence in Figure 26.3 with a depth independent horizontal density gradient, $\partial\rho/\partial x = \text{constant} < 0$ (see also Figure 24.1), thus leading to a depth dependent horizontal gradient in the hydrostatic pressure. This figure also serves to illustrate how the horizontal pressure gradient can change sign 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

$$\boldsymbol{B} = \frac{\nabla\rho \wedge \nabla p}{\rho^2}. \tag{26.18}$$

As shown in Section 34.4, a nonzero baroclinicity imparts a torque on fluid elements that acts as a source of vorticity.

The depth dependence to the horizontal pressure gradient imparts a vertical shear to the horizontal geostrophic velocity

$$\frac{\partial(\rho\, f\, \boldsymbol{u}_{\mathrm{g}})}{\partial z} = \hat{\boldsymbol{z}} \wedge \nabla(\partial p/\partial z) = -g\, \hat{\boldsymbol{z}} \wedge \nabla\rho. \tag{26.19}$$

This connection between horizontal density gradients and vertical shears in the geostrophic velocity is known as the *thermal wind relation*, which we return to in Section 26.4.5.

26.3.5 Geostrophic relation in pressure coordinates

The hydrostatic balance

$$\frac{\partial p}{\partial z} = -\rho\, g \tag{26.20}$$

can be used to eliminate density from the geostrophic balance (26.15) to render

$$f\, \hat{\mathbf{z}} \wedge \mathbf{u}_{\mathrm{g}} = \frac{g \nabla_z p}{\partial p/\partial z}. \tag{26.21}$$

The right hand side is minus the gradient of the geopotential

$$\Phi = g\, z \tag{26.22}$$

along surfaces of constant pressure (see Appendix 8.11.2)

$$f\, \hat{\mathbf{z}} \wedge \mathbf{u}_{\mathrm{g}} = -\nabla_p\, \Phi \Rightarrow f\, \boldsymbol{u}_{\mathrm{g}} = \hat{\boldsymbol{z}} \wedge \nabla_p \Phi. \tag{26.23}$$

This is a useful expression of geostrophy for the compressible atmosphere.

26.3.6 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. Dave Fultz of the University of Chicago.

26.4 Introducing planetary geostrophy

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 here and discuss their implications, deferring a more systematic derivation for later, with shallow water PG derived in Section 40.3 and continuously stratified PG derived in Section 41.2.

26.4.1 Planetary geostrophic equations

The governing equations for PG are based on the Boussinesq equations stated in Section 25.2.7, with the assumption of a steady state linear and frictional/geostrophic balance for the horizontal momentum

$$\rho_o\, f\, (\hat{\boldsymbol{z}} \wedge \boldsymbol{u}) = -\nabla p - \rho\, g\, \hat{\boldsymbol{z}} + \frac{\partial \boldsymbol{\tau}}{\partial z} \tag{26.24a}$$

$$\nabla_z \cdot \boldsymbol{u} + \frac{\partial w}{\partial z} = 0 \tag{26.24b}$$

$$\frac{\mathrm{D}b}{\mathrm{D}t} = \dot{b}. \tag{26.24c}$$

Note that the stress $\boldsymbol{\tau}$ is assumed to act just in the horizontal directions. Furthermore, the vertical component of the momentum equation (26.24a) is the hydrostatic balance

$$\frac{\partial p}{\partial z} = -\rho\, g. \tag{26.25}$$

Velocity is slaved to buoyancy

The only time derivative appearing in the PG equations appears in the buoyancy equation (26.24c). 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. 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

$$\boldsymbol{\tau} = (\tau^x, \tau^y, 0). \tag{26.26}$$

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 the molecular value. In general, all models (analytic or numerical) of planetary scale circulations are too coarse to resolve the molecular scales. 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 49.2.4.

26.4.2 Planetary geostrophic vorticity equation

The vertical component of relative vorticity is given by

$$\zeta = \frac{\partial v}{\partial x} - \frac{\partial u}{\partial y}, \tag{26.27}$$

with a thorough discussion of vorticity given in Chapter 34. 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 (26.24a), and rearranging terms, leads to the planetary geostrophic vorticity equation

$$-\rho_o\, f\, \frac{\partial \boldsymbol{u}}{\partial z} + \hat{\boldsymbol{z}}\, \rho_o\, \nabla \cdot (f\, \boldsymbol{u}) = -g\, \nabla \wedge (\hat{\boldsymbol{z}}\, \rho) + \frac{\partial\, (\nabla \wedge \boldsymbol{\tau})}{\partial z}. \tag{26.28}$$

Note that $\nabla \cdot (f\,\boldsymbol{u}) = \nabla_z \cdot (f\,\boldsymbol{u})$ since $\boldsymbol{u}$ is the horizontal velocity vector. Introducing buoyancy (Section 25.2.2)

$$b = -g \left[\frac{\rho - \rho_o}{\rho_o} \right] \tag{26.29}$$

leads to

$$-f\,\frac{\partial \boldsymbol{u}}{\partial z} + \hat{\boldsymbol{z}}\,\nabla_z \cdot (f\,\boldsymbol{u}) = \nabla \wedge (\hat{\boldsymbol{z}}\,b) + \frac{1}{\rho_o}\frac{\partial\,(\nabla \wedge \boldsymbol{\tau})}{\partial z}. \tag{26.30}$$

Pressure gradients are removed from the vorticity equation

One of the key reasons to study vorticity is that its evolution equation is not explicity affected by pressure gradients, since the curl of the pressure gradient vanishes. Eliminating pressure gradients affords a simpler evolution equation for vorticity, and we pursue that evolution more thoroughly in Chapter 34. However, it is important to note that the effects on vorticity from pressure gradients appear indirectly through their effects on velocity, given that vorticity is the curl of velocity. So although pressure is not directly present in the vorticity equation, its effects are nonetheless felt.

Relative vorticity is absent from the PG vorticity equation

It is notable that there is no explicit appearance of the relative vorticity, ζ, in the planetary geostrophic vorticity equation (26.24a). The reason is that we dropped the material time derivative when forming the planetary geostrophic momentum equation (26.24a). By doing so, we drop all expressions of ζ in the vorticity equation. Planetary geostrophy is valid for those cases where

$$|\zeta| \ll |f|, \tag{26.31}$$

which means vorticity is dominated by the planetary vorticity. We encounter more complete versions of the vorticity equation in Chapter 34, 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 (26.30) by taking planetary geostrophic scaling in the full vorticity equation. We choose here the path through the PG momentum equation since we have yet to discuss the full vorticity equation (see Chapter 34).

26.4.3 Taylor-Proudman and vertical stiffening

Consider the vorticity equation (26.30) on an f-plane with zero friction, in which the horizontal geostrophic motion is horizontally non-divergent

$$\nabla_z \cdot \boldsymbol{u} = 0 \qquad f\text{-plane geostrophy.} \tag{26.32}$$

Use of continuity (equation (26.24b)) means there is no vertical stretching of a vertical material line element (Section 13.2.5)

$$\frac{\partial w}{\partial z} = 0. \tag{26.33}$$

As shown in Chapter 34, a vortex tube exhibits the same kinematics as a material line element described in Section 13.2. Hence, $\partial w/\partial z = 0$ means there is no vertical stretching of a vortex tube in the planetary geostrophic fluid.

Figure 26.4: *The Taylor-Proudman result (26.34) says that horizontal flow in a homogeneous rapidly rotating fluid 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 must coherently move around the obstacle. The result is a vertically stiffened motion known as Taylor columns.*

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 (26.30) implies that the horizontal velocity has zero vertical shear

$$\frac{\partial \boldsymbol{u}}{\partial z} = 0 \qquad f\text{-plane and homogeneous density.} \tag{26.34}$$

This result is known as the Taylor-Proudman theorem.

Relevance to the ocean and atmosphere

In the ocean and atmosphere, the assumptions leading to the Taylor-Proudman theorem 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 incompressible scaling $W/H \sim U/L$ would indicate.[2] Additionally, for unstratified or linearly stratified fluids, there is a tendency for geostrophic turbulence to cascade energy into the gravest (i.e., the largest scale) vertical mode. 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.

26.4.4 Meridional motion in response to vortex stretching and wind curls

The vertical component to the vorticity balance (26.30) leads to

$$\nabla_z \cdot (f\,\boldsymbol{u}) = \frac{1}{\rho_o} \frac{\hat{\boldsymbol{z}} \cdot \partial\,(\nabla \wedge \boldsymbol{\tau})}{\partial z} \tag{26.35}$$

[2]Incompressibility in the form $\partial_x u + \partial_y v + \partial_z w = 0$ leads to the relation $U/L \sim W/H$, where W is a typical vertical velocity scale, H is a typical vertical length scale, and U and L are the corresponding horizontal scales. See Section 24.3.2 for a discussion in the context of the hydrostatic balance.

which can be written

$$\beta\, v = -f\, \nabla_z \cdot \boldsymbol{u} + \frac{1}{\rho_o} \frac{\hat{\boldsymbol{z}} \cdot \partial\, (\nabla \wedge \boldsymbol{\tau})}{\partial z} \tag{26.36}$$

where

$$\beta = \frac{\partial f}{\partial y} \tag{26.37}$$

is the gradient of planetary vorticity. The continuity equation (26.24b) can be used to remove the horizontal divergence, which yields the vorticity balance

$$\beta\, v = f\, \frac{\partial w}{\partial z} + \frac{1}{\rho_o} \frac{\hat{\boldsymbol{z}} \cdot \partial\, (\nabla \wedge \boldsymbol{\tau})}{\partial z}. \tag{26.38}$$

The left hand side is the meridional advection of planetary vorticity. The first term on the right hand size arises from vortex stretching by planetary vorticity; i.e., planetary induction or the β-effect discussed in Section 34.6.2. The second term is the vertical divergence of the curl of the stress.

Reading the vorticity equation (26.38) 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 (26.38) 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.

26.4.5 Thermal wind balance for the ocean

Horizontal components to the inviscid vorticity equation (equation (26.30) with $\boldsymbol{\tau}$ set to zero) form the *thermal wind balance*

$$f\, \frac{\partial \boldsymbol{u}}{\partial z} = -\nabla \wedge (\hat{\boldsymbol{z}}\, b) = \hat{\boldsymbol{z}} \wedge \nabla b, \tag{26.39}$$

which takes on the component form

$$f\, \frac{\partial u}{\partial z} = -\frac{\partial b}{\partial y} \qquad \text{and} \qquad f\, \frac{\partial v}{\partial z} = \frac{\partial b}{\partial x}. \tag{26.40}$$

As seen already in Section 26.3.4, these relations can also be derived direction from taking the vertical derivative of the horizontal momentum equation (26.24a) and then using the horizontal gradient of the hydrostatic balance (26.25). In either case, the thermal wind balance (26.39) 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*. 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

Figure 26.5: *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 26.3, the meridional pressure gradient decreases when moving upward, $\partial(\partial p/\partial y)/\partial z < 0$. 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, $\partial u_g/\partial z > 0$.*

a poleward reduction in buoyancy. Also, for a stably stratified fluid, density increases with depth. Figure 26.5 illustrates this situation.

The zonal average removes all zonal variations, thus putting $\partial_x \rho = 0$ and so rendering the zonally averaged thermal wind relation

$$f \frac{\partial \overline{u}}{\partial z} = \frac{g}{\rho_o} \frac{\partial \overline{\rho}}{\partial y} = -\frac{\partial \overline{b}}{\partial y} > 0, \tag{26.41}$$

where $\overline{(\ )}$ is the zonal mean operator. In the northern hemisphere, $\partial_y \overline{b} < 0$, so that the zonal averaged thermal wind shear is positive, $\partial_z \overline{u} > 0$. 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, $\partial_y \overline{b} > 0$, 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 (26.39) between two constant depth surfaces leads to

$$\boldsymbol{u}(z) = \boldsymbol{u}(z_{\text{ref}}) - f^{-1} \nabla \wedge \hat{\boldsymbol{z}} \int_{z_{\text{ref}}}^{z} b \, \mathrm{d}z. \tag{26.42}$$

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 point along

the integration path, 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.

26.4.6 Thermal wind balance for the atmosphere

The large-scale atmosphere is compressible and predominantly in hydrostatic balance. The expression for geostrophic balance (26.23) 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 (26.23) to render

$$f \, \frac{\partial \boldsymbol{u}}{\partial p} = \hat{\boldsymbol{z}} \wedge \nabla_p \left[\frac{\partial \Phi}{\partial p} \right] . \tag{26.43}$$

The hydrostatic relation $\partial p/\partial z = -\rho \, g$ takes the form

$$\frac{\partial p}{\partial \Phi} = -\rho \Rightarrow \frac{\partial \Phi}{\partial p} = -1/\rho \tag{26.44}$$

in which case

$$f \, \frac{\partial \boldsymbol{u}}{\partial p} = -\hat{\boldsymbol{z}} \wedge \nabla_p (1/\rho). \tag{26.45}$$

Ideal gas atmosphere

The specific volume takes the following form for an ideal gas atmosphere (see Section 22.8.1)

$$\rho^{-1} = \frac{R^{\mathrm{M}} \, T}{p}. \tag{26.46}$$

Since the horizontal derivative in the thermal wind relation (26.45) is along pressure surfaces, we have

$$f \, \frac{\partial \boldsymbol{u}}{\partial p} = -\frac{R^{\mathrm{M}}}{p} \, \left[\hat{\boldsymbol{z}} \wedge \nabla_p T \right] . \tag{26.47}$$

This expression gives rise to the name "thermal wind", with vertical shears of the horizontal velocity generated by horizontal temperature gradients.

As for the ocean in equation (26.42), we vertically integrate the thermal wind expression (26.47), only now do so between two pressure levels

$$\boldsymbol{u}(p_{\mathrm{a}}) - \boldsymbol{u}_{\mathrm{g}}(p_{\mathrm{b}}) = f^{-1} \, R^{\mathrm{M}} \, \hat{\boldsymbol{z}} \wedge \nabla_p \left[\int_{p_{\mathrm{a}}}^{p_{\mathrm{b}}} \frac{T \, \mathrm{d}p}{p} \right] , \tag{26.48}$$

where $p_{\mathrm{a}} < p_{\mathrm{b}}$, so that p_{a} has 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)

$$\boldsymbol{u}_T = \boldsymbol{u}(p_{\mathrm{a}}) - \boldsymbol{u}(p_{\mathrm{b}}) \qquad \text{with } p_{\mathrm{a}} < p_{\mathrm{b}} \tag{26.49}$$

so that

$$\boldsymbol{u}_T = \frac{R^{\mathrm{M}}}{f} \, \hat{\boldsymbol{z}} \wedge \nabla_p \overline{T}^{\ln \mathrm{p}}, \tag{26.50}$$

where we introduced the log-pressure weighted temperature between the two pressure surfaces

$$\overline{T}^{\ln \mathrm{p}} = \int_{p_{\mathrm{a}}}^{p_{\mathrm{b}}} \frac{T \, \mathrm{d}p}{p}. \tag{26.51}$$

Figure 26.6: *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, u_g, 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 $f < 0$.*

The relation (26.50) 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, $\partial \overline{T}^{\ln \mathrm{p}}/\partial y < 0$. Hence, the zonal westerly winds increase in magnitude with height. Furthermore, the thermal wind shear points to the east. In general for 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, $f < 0$

Barotropic flow

Return to the thermal wind equation (26.45)

$$f \frac{\partial \boldsymbol{u}}{\partial p} = -\hat{\boldsymbol{z}} \wedge \nabla_p \alpha = \frac{\hat{\boldsymbol{z}} \wedge \nabla_p \rho}{\rho^2}. \tag{26.52}$$

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 34.2 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, $\boldsymbol{B} = \nabla p \wedge \nabla \alpha = 0$. The functional relation $\rho = \rho(p)$ (equivalently $p = p(\rho)$) is a sufficient condition for vanishing baroclinicity. As seen in Section 34.4, for a barotropic flow there is no generation of vorticity through the torques created when density isosurfaces (isopycnals) deviate from pressure isosurfaces (isobars).

26.4.7 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 result (26.34) 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 26.4 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. Dave Fultz of the University of Chicago 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, shows how Taylor curtains arise in rotating fluids.

26.5 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)

$$\mathbf{\Sigma}^{\text{form}} = p\,\nabla\eta, \tag{26.53}$$

with the opposite sign for the form stress on the bottom side of a surface. Here, $z = \eta(x, y, t)$ is the depth of the surface (see Figure 20.3 or Figure 26.7). 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 26.5.1) and on an isopycnal layer (Section 26.5.2), 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 deviations from the zonal mean; i.e., zonal fluctuations associated with waves and geostrophic turbulence, provides an eastward acceleration to the fluid while transporting 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 48.2 and the quasi-Stokes transport discussed in Section 49.2.

26.5.1 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

Figure 26.7: *Schematic of the zonal form stress, $\Sigma_x^{\rm form}$, acting on a surface whose zonal mean vertical position is $z = \overline{\eta}(y,t)$ and whose vertical position relative to the zonal mean is $z = \overline{\eta}(y,t) + \eta'(x,y,t)$.*

zonal mean (see Figure 26.7). 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

$$\overline{\eta} = \frac{1}{L} \int_0^L \eta \, \mathrm{d}x \tag{26.54}$$

and its corresponding zonal fluctuation is

$$\eta' = \eta - \overline{\eta}. \tag{26.55}$$

The zonal component of the form stress is thus given by

$$p \, \partial_x \eta = p(x, \overline{\eta} + \eta') \, \partial_x (\overline{\eta} + \eta') \tag{26.56a}$$
$$= p(x, \overline{\eta} + \eta') \, \partial_x \eta' \tag{26.56b}$$
$$\approx [p(x, \overline{\eta}) + \partial_z p(x, \overline{\eta}) \, \eta'] \, \partial_x \eta' \tag{26.56c}$$
$$= p(x, \overline{\eta}) \, \partial_x \eta' + \mathcal{O}(\eta')^2. \tag{26.56d}$$

Hence, to second order in fluctuations, η', the zonal form stress acting on the surface equals to $p(x, \overline{\eta}) \, \partial_x \eta'$, where it is important to note that pressure is evaluated at the zonal mean depth, $z = \overline{\eta}$.

To within the same accuracy, the zonal integral of the zonal form stress is given by

$$\int_0^L \Sigma_x^{\rm form} \, \mathrm{d}x \approx \int_0^L p(\overline{\eta}) \, (\partial \eta' / \partial x) \, \mathrm{d}x = - \int_0^L \eta' \, [\partial p(\overline{\eta}) / \partial x] \, \mathrm{d}x, \tag{26.57}$$

where the final equality follows from zonal periodicity. Now assume the zonal pressure gradient at $\overline{\eta}$ is balanced by a meridional geostrophic velocity

$$\partial p(\overline{\eta}) / \partial x = f \rho_0 \, v(\overline{\eta}), \tag{26.58}$$

where

$$v(\overline{\eta}) = \overline{v}(\overline{\eta}) + v'(\overline{\eta}) \tag{26.59}$$

so that

$$\int_0^L \Sigma_x^{\rm form} \, \mathrm{d}x = - \rho_0 \, f \int_0^L \eta' \, v' \, \mathrm{d}x, \tag{26.60}$$

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

$$\overline{\Sigma}_x^{\rm form} = - \rho_0 \, f \, \overline{v' \, \eta'}. \tag{26.61}$$

Zonal mean zonal form stress acting on an isopycnal surface

To further unpack the correlation appearing in equation (26.61), specialize to the case of an isopycnal surface in an adiabatic fluid. As shown in Sections 48.2.6 and 48.4.7, vertical fluctuations in the position of the isopycnal surfaces, relative to the zonal mean $\overline{\eta}$, are related to zonal fluctuations in the density

$$\eta' \approx -\frac{\rho'}{\partial\overline{\rho}/\partial z} = -\frac{b'}{N^2}, \tag{26.62}$$

where we introduced the squared buoyancy frequency of the zonal mean state as well as the fluctuating buoyancy

$$N^2 = -\frac{g}{\rho_0}\,\frac{\partial\overline{\rho}}{\partial z} \qquad \text{and} \qquad b' = -\frac{g\,\rho'}{\rho_0}. \tag{26.63}$$

The zonally averaged zonal form stress thus takes the form

$$\overline{\Sigma}_x^{\text{form}} = \frac{\rho_o\,f}{N^2}\,\overline{v'\,b'}. \tag{26.64}$$

Again, the assumptions rendering the result (26.64) 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

$$\overline{\Sigma}_x^{\text{form}} > 0. \tag{26.65}$$

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.

26.5.2 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 26.8. This layered/isopycnal analysis anticipates some of the development considered for the stacked shallow water model in Chapters 29 and 30 as well as for isopycnal models in Section 46.1.

The net form stress acting on the upper and lower layer interfaces in Figure 26.8 is given by

$$\begin{aligned}
\mathbf{\Sigma}^{\text{layer form}} &= p_1\nabla\eta_1 - p_2\nabla\eta_2 && \text{(26.66a)}\\
&= p(\eta + h/2)\,\nabla(\eta + h/2) - p(\eta - h/2)\,\nabla(\eta - h/2) && \text{(26.66b)}\\
&\approx [p(\eta) - \rho\,g\,h/2]\,\nabla(\eta + h/2) - [p(\eta) + \rho\,g\,h/2]\,\nabla(\eta - h/2) && \text{(26.66c)}\\
&= p\,\nabla h - \rho\,g\,h\,\nabla\eta && \text{(26.66d)}\\
&= \nabla(p\,h) - h\,\nabla(p + \rho\,g\,\eta) && \text{(26.66e)}\\
&= \nabla(p\,h) - \rho_0\,h\,\nabla M. && \text{(26.66f)}
\end{aligned}$$

Figure 26.8: *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 $\rho =$ 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.*

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

$$M\rho_0 = p + \rho\, g\, \eta, \tag{26.67}$$

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

$$p(\eta + h/2) \approx p(\eta) + \frac{\partial p}{\partial z}\frac{h}{2} = p(\eta) - \rho\, g\, h/2 \tag{26.68a}$$

$$p(\eta - h/2) \approx p(\eta) - \frac{\partial p}{\partial z}\frac{h}{2} = p(\eta) + \rho\, g\, h/2. \tag{26.68b}$$

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

$$\overline{\Sigma}_x^{\text{layer form}} = -\rho_0\, \overline{h'\, \partial M'/\partial x}, \tag{26.69}$$

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

$$f\, v' = \partial_x M'. \tag{26.70}$$

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

$$\overline{\Sigma}_x^{\text{layer form}} = -\rho_0\, f\, \overline{v'\, h'}. \tag{26.71}$$

Hence, as the geostrophic eddies provide a net eastward acceleration to the layer (equation (26.65)), 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 (26.71), parameterize the velocity-thickness eddy correlation, $\overline{v'\, h'}$, by downgradient diffusion of thickness

$$\overline{v'\, h'} = -\kappa\, \partial_y \overline{h}, \tag{26.72}$$

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 49.2.5. 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

$$\overline{\Sigma}_x^{\text{layer form}} = \rho_0 \, f \, \kappa \, \partial_y \overline{h}. \tag{26.73}$$

So in the northern hemisphere in regions where the zonal mean layer thickness increases to the north, $\partial_y \overline{h} > 0$, 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 26.5.1, 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.

26.5.3 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 48.2.

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 30.2. That discussion complements the presentation given here, there focusing exclusively on the case of 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 49.2.

26.6 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 Rosby 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 11.2.4.

(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

$$u = |\boldsymbol{u}| \cos \Delta \qquad v = |\boldsymbol{u}| \sin \Delta, \tag{26.74}$$

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

$$\rho = \rho(p, \theta). \tag{26.75}$$

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

$$\frac{\partial \mathbf{u}}{\partial z} = \left[\frac{N^2}{f\,\rho\,g}\right] (\hat{\mathbf{z}} \wedge \nabla_\theta p), \tag{26.76}$$

where

$$N^2 = -\frac{g}{\rho}\,\frac{\partial \rho}{\partial \theta}\,\frac{\partial \theta}{\partial z} = g\,\beta_\theta\,\frac{\partial \theta}{\partial z} > 0 \tag{26.77}$$

is the squared buoyancy frequency, assumed positive so that the fluid is gravitationally stable in the vertical (see Section 21.4.4). The term β_θ is the thermal expansion coefficient written in terms of potential temperature (Section 21.3.5). Finally, the horizontal gradient projected onto constant θ surfaces is given by (see Section 8.11.2)

$$\nabla_\theta = \hat{\boldsymbol{x}} \left[\frac{\partial}{\partial x}\right]_{y,\theta} + \hat{\boldsymbol{y}} \left[\frac{\partial}{\partial y}\right]_{x,\theta} \tag{26.78a}$$

$$= \nabla_z - \left[\frac{\nabla_z \theta}{\partial \theta/\partial z}\right] \frac{\partial}{\partial z}. \tag{26.78b}$$

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 $\rho = \rho(\theta, p)$.

27

Balanced inviscid horizontal flows

We here consider a variety of inviscid horizontal flow regimes characterized by a balance between a subset of terms appearing in the horizontal momentum equation. This discussion allows us to compare the geostrophic flow of Chapter 26 to a variety of *ageostrophic flows* such as gradient wind, inertial motion, and cyclostrophic balance. We offer a categorization of the flow following the *natural coordinates* used by *Holton* (1992). Natural coordinates provide a lucid means to compare the relative magnitudes of the Coriolis, pressure, and centrifugal accelerations acting on a fluid particle moving horizontally.

READER'S GUIDE TO THIS CHAPTER

This chapter assumes an understanding of geostrophic flow from Chapter 26. Some of this material is used in subsequent chapters, in particular Chapter 28 on Ekman layer dynamics as well as Chapters 40 and 42 on quasi-geostrophy. We make use of some geometry discussed in Chapter 4, though most of the salient points are revisited here so that Chapter 4 is an option rather than a requirement.

27.1 Horizontal flow described by natural coordinates . . . 480
- 27.1.1 Natural coordinates . . . 480
- 27.1.2 Material acceleration . . . 480
- 27.1.3 Centripetal and centrifugal accelerations . . . 482
- 27.1.4 Coriolis and pressure gradient . . . 482
- 27.1.5 Horizontal momentum equation . . . 483
- 27.1.6 An Eulerian decomposition of the acceleration . . . 483
- 27.1.7 Further study . . . 484

27.2 Exact geostrophic balance . . . 484
- 27.2.1 Steady f-plane flow . . . 485
- 27.2.2 Steady flow on a sphere . . . 485
- 27.2.3 What about geostrophic balance with curvature? . . . 485

27.3 Inertial motion of fluid particles . . . 486
- 27.3.1 Anti-cyclonic circular motion . . . 486
- 27.3.2 Balance between Coriolis and centrifugal . . . 486
- 27.3.3 Period for inertial motion . . . 486
- 27.3.4 Observing inertial motion . . . 487
- 27.3.5 Inertial motion is Lagrangian . . . 487
- 27.3.6 "Inertial" motion is not motion viewed from an inertial frame . . . 488

27.4 Cyclostrophic balance **488**
27.5 Gradient wind balance **488**
27.5.1 Constraints on gradient wind flow 489
27.5.2 The variety of gradient wind flows 489
27.5.3 Comments 491

27.1 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 particle. 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 geopotetial). Using the "natural coordinates" arising from this description we encounter the centripetal/centrifugal acceleration that arises from curvature in the trajectory. Note that this acceleration is distinct from the centrifugal acceleration that arises from planetary rotation and is contained within the effective gravitational acceleration acting in the local vertical (see Section 9.10.1). We also decompose the accelerations from pressure, friction, and Coriolis into natural coordinate directions. The kinematics afforded by natural coordinates succinctly summarizes the horizontal motion and suggests a fruitful categorization of balances arising in limiting cases.

27.1.1 Natural coordinates

Natural coordinates for horizontal motion are defined by a locally orthogonal set of unit vectors (see Figure 27.1)

$$\hat{\boldsymbol{z}} = \hat{\boldsymbol{u}} \wedge \hat{\boldsymbol{n}} = \text{vertical direction} \tag{27.1a}$$

$$\hat{\boldsymbol{s}} = \hat{\boldsymbol{n}} \wedge \hat{\boldsymbol{z}} = \text{tangent to velocity} \tag{27.1b}$$

$$\hat{\boldsymbol{n}} = \hat{\boldsymbol{z}} \wedge \hat{\boldsymbol{u}} = \text{normal direction.} \tag{27.1c}$$

The unit vector $\hat{\boldsymbol{u}}$ is tangent to the velocity vector (which is horizontal), so that

$$\boldsymbol{u} = |\boldsymbol{u}|\, \hat{\boldsymbol{u}} = \frac{\mathrm{D}s}{\mathrm{D}t}\, \hat{\boldsymbol{u}}, \tag{27.2}$$

where s is the arc-length measured along the trajectory as introduced in Section 2.4. The unit vector $\hat{\boldsymbol{n}}$ is perpendicular to the velocity and points to the left of the trajectory facing downstream.

27.1.2 Material acceleration

When writing the velocity according to equation (27.2), we decompose the acceleration into the change in speed and change in direction

$$\frac{\mathrm{D}\boldsymbol{u}}{\mathrm{D}t} = \frac{\mathrm{D}|\boldsymbol{u}|}{\mathrm{D}t}\, \hat{\boldsymbol{u}} + |\boldsymbol{u}|\, \frac{\mathrm{D}\hat{\boldsymbol{u}}}{\mathrm{D}t}. \tag{27.3}$$

Following our discussion of rotation in Section 9.4 (see Figure 9.3), the magnitude of the direction change can be written in terms of the infinitesimal angle swept out by the motion as the fluid particle moves along a trajectory

$$|\delta \hat{\boldsymbol{u}}| = \delta \vartheta. \tag{27.4}$$

Figure 27.1: *Illustrating the decomposition of horizontal motion of a fluid particle into natural coordinate directions. These directions are defined by a unit tagent vector, $\hat{\boldsymbol{u}}$, pointing in the direction of the fluid particle motion, and a unit normal vector, $\hat{\boldsymbol{n}}$, pointing to the left of the motion facing downstream.*

Figure 27.2: *Illustrating the radius of curvature associated with turning motion of a fluid particle. 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 $\hat{\boldsymbol{n}}$) and negative when turning in the opposite direction (to the right; convex as defined by $\hat{\boldsymbol{n}}$). See Section 4.2 for more details on curvature, with Figure 4.4 offering more details for how to determine the radius of curvature.*

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

$$\delta\vartheta = \frac{\delta s}{R}. \tag{27.5}$$

Finally, the infinitesimal change in tangent, $\delta\hat{\boldsymbol{u}}$, is directed normal to the motion, and is thus directed along the $\hat{\boldsymbol{n}}$ direction. We see this property by noting that

$$\hat{\boldsymbol{u}} \cdot \hat{\boldsymbol{u}} = 1 \Rightarrow \delta\hat{\boldsymbol{u}} \cdot \hat{\boldsymbol{u}} = 0. \tag{27.6}$$

That is, $\delta\hat{\boldsymbol{u}}$ is orthogonal to $\hat{\boldsymbol{u}}$, so that it points parallel or anti-parallel to $\hat{\boldsymbol{n}}$. We detailed this property of rotating unit vectors in Section 2.1.4 (see Figure 2.2). Our convention is that $\hat{\boldsymbol{n}}$ points to the left of $\hat{\boldsymbol{u}}$, so that if the trajectory turns to the left, then $\delta\hat{\boldsymbol{u}}$ points parallel to $\hat{\boldsymbol{n}}$, whereas if the trajectory turns to the right then $\delta\hat{\boldsymbol{u}}$ points anti-parallel to $\hat{\boldsymbol{n}}$. That is, $\delta\hat{\boldsymbol{u}}$ 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

$$\delta\hat{\boldsymbol{u}} = \hat{\boldsymbol{n}}\,\frac{\delta s}{R}. \tag{27.7}$$

Again, our sign convention takes $R > 0$ for a particle turning in the direction of $\hat{\boldsymbol{n}}$ (to the left facing downstream) and $R < 0$ for a particle turning opposite to $\hat{\boldsymbol{n}}$ (to the right facing downstream).

Figure 27.3: *Centripetal acceleration of a turning fluid particle, $\hat{\boldsymbol{n}}\,|\boldsymbol{u}|^2/R$, points to the concave side of the curve (towards the center) whereas the centrifugal acceleration, $-\hat{\boldsymbol{n}}\,|\boldsymbol{u}|^2/R$, points to the convex side (away from the center). The centripetal and centrifugal accelerations are paired action/reaction accelerations. For a left turning trajectory, the concave side is on the left and has positive radius of curvature, $R > 0$, whereas for the right turning trajectory 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.*

Hence, the material time change is

$$\frac{\mathrm{D}\hat{\boldsymbol{u}}}{\mathrm{D}t} = \frac{\mathrm{D}\hat{\boldsymbol{u}}}{\mathrm{D}s}\,\frac{\mathrm{D}s}{\mathrm{D}t} = \frac{\hat{\boldsymbol{n}}}{R}\,\frac{\mathrm{D}s}{\mathrm{D}t} = \frac{\hat{\boldsymbol{n}}}{R}\,|\boldsymbol{u}|, \tag{27.8}$$

where the speed is given by the time change of the arc-length along the trajectory

$$|\boldsymbol{u}| = \frac{\mathrm{D}s}{\mathrm{D}t}. \tag{27.9}$$

Combining these results renders the acceleration

$$\frac{\mathrm{D}\boldsymbol{u}}{\mathrm{D}t} = \frac{\mathrm{D}|\boldsymbol{u}|}{\mathrm{D}t}\,\hat{\boldsymbol{u}} + \frac{|\boldsymbol{u}|^2}{R}\,\hat{\boldsymbol{n}}. \tag{27.10}$$

The acceleration has thus been decomposed into the change in speed of the fluid particle along the direction of the motion, plus the centripetal acceleration due to curvature of the trajectory. In Section 27.1.3 we justify referring to $(|\boldsymbol{u}|^2/R)\,\hat{\boldsymbol{n}}$ as the *centripetal* acceleration.

27.1.3 Centripetal and centrifugal accelerations

The centripetal acceleration points towards the concave side of a turning particle 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 $\hat{\boldsymbol{n}}\,|\boldsymbol{u}|^2/R$? For motion turning to the left, towards $\hat{\boldsymbol{n}}$, the radius of curvature is positive, $R > 0$, so that $\hat{\boldsymbol{n}}\,|\boldsymbol{u}|^2/R$ points to the concave side of the trajectory (left side). If the particle is turning to the right then $R < 0$, which again means that $\hat{\boldsymbol{n}}\,|\boldsymbol{u}|^2/R$ points to the concave side (now on the right). We conclude that the acceleration $\hat{\boldsymbol{n}}\,|\boldsymbol{u}|^2/R$ indeed represents a centripetal acceleration and $-\hat{\boldsymbol{n}}\,|\boldsymbol{u}|^2/R$ is the centrifugal acceleration.

27.1.4 Coriolis and pressure gradient

The Coriolis acceleration takes the following form in natural coordinates

$$-f\,\hat{\boldsymbol{z}} \wedge \boldsymbol{u} = -(\hat{\boldsymbol{z}} \wedge \hat{\boldsymbol{u}})\,f\,|\boldsymbol{u}| \tag{27.11a}$$

$$= -\hat{\boldsymbol{n}}\,f\,|\boldsymbol{u}|, \tag{27.11b}$$

so that the Coriolis acceleration always points to the right of the flow direction. In contrast, the pressure gradient has two components

$$\nabla p = \hat{\boldsymbol{u}}\,(\hat{\boldsymbol{u}} \cdot \nabla p) + \hat{\boldsymbol{n}}\,(\hat{\boldsymbol{n}} \cdot \nabla p) \tag{27.12a}$$

$$= \hat{\boldsymbol{u}}\,\frac{\partial p}{\partial s} + \hat{\boldsymbol{n}}\,\frac{\partial p}{\partial n}. \tag{27.12b}$$

27.1.5 Horizontal momentum equation

Bringing the above results together leads to the horizontal momentum equation decompoesed into natural coordinates

$$\frac{\mathrm{D}|\boldsymbol{u}|}{\mathrm{D}t} = -\frac{1}{\rho_0}\,\frac{\partial p}{\partial s} + \boldsymbol{F} \cdot \hat{\boldsymbol{u}} \tag{27.13a}$$

$$\frac{|\boldsymbol{u}|^2}{R} + f\,|\boldsymbol{u}| = -\frac{1}{\rho_0}\,\frac{\partial p}{\partial n} + \boldsymbol{F} \cdot \hat{\boldsymbol{n}}, \tag{27.13b}$$

where $\boldsymbol{F}$ is the friction force per mass 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 layers in Chapter 28.

27.1.6 An Eulerian decomposition of the acceleration

The horizontal equations of motion (27.13a) and (27.13b) are relatively simple since we are focused on a Lagrangian description. In contrast, an Eulerian description requires us to expand the material time derivative to render

$$\frac{\partial \boldsymbol{u}}{\partial t} + (\boldsymbol{u} \cdot \nabla)\,\boldsymbol{u} = \frac{\mathrm{D}|\boldsymbol{u}|}{\mathrm{D}t}\,\hat{\boldsymbol{u}} + \frac{|\boldsymbol{u}|^2}{R}\,\hat{\boldsymbol{n}}, \tag{27.14}$$

so that

$$\hat{\boldsymbol{u}} \cdot \left[\frac{\partial \boldsymbol{u}}{\partial t} + (\boldsymbol{u} \cdot \nabla)\,\boldsymbol{u}\right] = \frac{\mathrm{D}|\boldsymbol{u}|}{\mathrm{D}t} = -\frac{1}{\rho_0}\,\frac{\partial p}{\partial s} + \boldsymbol{F} \cdot \hat{\boldsymbol{u}} \tag{27.15a}$$

$$\hat{\boldsymbol{n}} \cdot \left[\frac{\partial \boldsymbol{u}}{\partial t} + (\boldsymbol{u} \cdot \nabla)\,\boldsymbol{u}\right] = \frac{|\boldsymbol{u}|^2}{R} = -f\,|\boldsymbol{u}| - \frac{1}{\rho_0}\,\frac{\partial p}{\partial n} + \boldsymbol{F} \cdot \hat{\boldsymbol{n}}. \tag{27.15b}$$

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

$$R^{-1} = \frac{\hat{\boldsymbol{n}} \cdot [\partial \boldsymbol{u}/\partial t + (\boldsymbol{u} \cdot \nabla)\,\boldsymbol{u}]}{|\boldsymbol{u}|^2} = \frac{-f\,|\boldsymbol{u}| - \rho_0^{-1}\,\partial p/\partial n + \boldsymbol{F} \cdot \hat{\boldsymbol{n}}}{|\boldsymbol{u}|^2}. \tag{27.16}$$

We focus in the remainder of this chapter on the Lagrangian perspective given its much more concise formulation.

27.1.7 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 featuring Prof. A. Shapiro.

27.2 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 particle remains constant. If this motion furthermore 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. More precisely, exact geostrophic balance occurs under the following conditions:

- fluid particles move on a straight line so that the radius of curvature is infinite, $|R| = \infty$;
- fluid particles 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

$$\frac{\mathrm{D}|\boldsymbol{u}|}{\mathrm{D}t} = 0 \tag{27.17a}$$

$$f\,|\boldsymbol{u}| = -\frac{1}{\rho_0}\frac{\partial p}{\partial n}. \tag{27.17b}$$

Equation (27.17a) says that the speed of a fluid particle is constant. Equation (27.17b) 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.

Writing the horizontal advection of speed in the form

$$\boldsymbol{u}\cdot\nabla|\boldsymbol{u}| = |\boldsymbol{u}|\,\hat{\boldsymbol{u}}\cdot\nabla|\boldsymbol{u}| = |\boldsymbol{u}|\,\frac{\partial|\boldsymbol{u}|}{\partial s}, \tag{27.18}$$

allows us to write the material constancy of the flow speed as

$$\frac{\partial|\boldsymbol{u}|}{\partial t} + |\boldsymbol{u}|\,\frac{\partial|\boldsymbol{u}|}{\partial s} = 0. \tag{27.19}$$

Hence, a steady flow speed, with $\partial|\boldsymbol{u}|/\partial t = 0$, only holds for the exact geostrophic balance if the flow speed is fixed along each trajectory path

$$\frac{\partial|\boldsymbol{u}|}{\partial s} = 0 \Longrightarrow \frac{\partial|\boldsymbol{u}|}{\partial t} = 0. \tag{27.20}$$

What is required for this condition to hold?

Figure 27.4: *Exact geostrophic balance 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 sphere, however, steady exact geostrophic balance holds only for zonal flow.*

27.2.1 Steady f-plane flow

Geostrophic motion on an f-plane is horizontally non-divergent (Section 26.3)

$$\nabla \cdot \boldsymbol{u} = \nabla \cdot (\hat{\boldsymbol{u}}\,|\boldsymbol{u}|) = 0. \tag{27.21}$$

Flow in a straight line, with each trajectory parallel to one another, has the particle trajectory direction independent of space. Hence, the non-divergent condition means that

$$\nabla \cdot (\hat{\boldsymbol{u}}\,|\boldsymbol{u}|) = (\hat{\boldsymbol{u}} \cdot \nabla)\,|\boldsymbol{u}| = \frac{\partial |\boldsymbol{u}|}{\partial s} = 0, \tag{27.22}$$

which then proves that exact geostrophic flow on an f-plane is steady.

27.2.2 Steady flow on a sphere

The geostrophic velocity in the presence of a meridional gradient of the Coriolis parameter, $f = f(y)$, satisfies (Section 26.3)

$$\nabla \cdot (f\,\boldsymbol{u}) = 0. \tag{27.23}$$

Making use of $\nabla \cdot \hat{\boldsymbol{u}} = 0$ for straight line trajectories leads to

$$\nabla \cdot (f\,\boldsymbol{u}) = \frac{\partial (f\,|\boldsymbol{u}|)}{\partial s} = 0. \tag{27.24}$$

We conclude that $\partial|\boldsymbol{u}|/\partial s = 0$ 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 sphere is steady only for trajectories that follow constant latitude lines; i.e., zonal trajectories as depicted in Figure 27.4.

27.2.3 What about geostrophic balance with curvature?

The geostrophically balanced flows discussed in Chapter 26 generally have curvature, such as for the geostrophic motion around a pressure center as shown in Figure 26.2. 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 curved flow we are ignoring the centrifugal acceleration arising from curvature. This is a sensible approximation for large-scale flows. Even so, it is an approximation and it is identically zero only for straight line motion in which there is an *exact* balance between pressure and Coriolis.

27.3 Inertial motion of fluid particles

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

$$\frac{\mathrm{D}|\boldsymbol{u}|}{\mathrm{D}t} = 0 \tag{27.25a}$$

$$\frac{|\boldsymbol{u}|^2}{R} + f\,|\boldsymbol{u}| = 0. \tag{27.25b}$$

27.3.1 Anti-cyclonic circular motion

Equation (27.25a) says that inertial motion occurs with constant speed. Equation (27.25b) says that the motion maintains the balance between Coriolis and centrifugal accelerations

$$f\,|\boldsymbol{u}| = -\frac{|\boldsymbol{u}|^2}{R}. \tag{27.26}$$

To further understand the implications of this result, divide by the speed to render

$$f = -\frac{|\boldsymbol{u}|}{R} \tag{27.27}$$

so that the radius for the inertial circle is

$$R = -|\boldsymbol{u}|/f. \tag{27.28}$$

Equation (27.27) 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. Hence, inertial motion is oriented anti-cyclonically (opposite orientation to the earth's rotation). If the Coriolis parameter is constant, then the motion is circular, as depicted in Figure 27.5.

27.3.2 Balance between Coriolis and centrifugal

Return to the balance given by equation (27.26) and recall that the Coriolis acceleration in the northern hemisphere points to the right when facing downstream, as per equation (27.11b). Hence, we interpret the balance (27.26) as between the Coriolis acceleration pointing to the right and the centrifugal acceleration pointing to the left. The same balance occurs in the southern hemisphere betweeen Coriolis and centrifugal, though with the opposite orientation of the motion.

27.3.3 Period for inertial motion

Equation (27.28) says that the speed of the fluid particle is given by the radius of curvature times the magnitude of the Coriolis parameter

$$|\boldsymbol{u}| = R\,|f|. \tag{27.29}$$

Figure 27.5: *Inertial motion of a fluid particle 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 $|R| = ||\boldsymbol{u}|/|f|$.*

The time for a fluid particle 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

$$T_{\text{inertial}} = \frac{2\pi R}{|\boldsymbol{u}|} = \frac{2\pi}{|f|}. \tag{27.30}$$

We encountered this inertial period in Section 10.3 when considering inertial oscillations for a point particle.

27.3.4 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 so that there can readily be a balance between centrifugal and Coriolis for the moving seawater fluid particle. 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 current speed of $|\boldsymbol{u}| \sim 0.1$ m s^{-1}, which is not atypical of current speeds outside of strong boundary currents or mesoscale eddies, and assume the Coriolis parameter $f = 10^{-4}$ s^{-1}. In this case the inertial radius is

$$R_{\text{inertial}} \approx 10^3 \text{ m}. \tag{27.31}$$

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.

27.3.5 Inertial motion is Lagrangian

The analysis in the current section concerns a fluid particle moving without feeling the impacts from pressure forces. The fluid particle thus exhibits an identical force balance to the point particle discussed in Section 10.3. So although we can measure inertial oscillations at a fixed point in space, the present considerations are Lagrangian or material in nature, focusing on motion of a fluid particle. Furthermore, the inertial period refers to the time it takes for a fluid particle 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 inertia-gravity waves that are have periods close to the inertial period (see Section 31.3).

27.3.6 "Inertial" motion is not motion viewed from 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 inertia. Hence, as noted in Section 10.3.4, "inertial motion" in this context does *not* refer to motion as viewed in an inertial reference frame, such as that discussed in Section 9.3.

27.4 Cyclostrophic balance

Cyclostrophic balance occurs under the following conditions:

- fluid particles 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

$$\frac{\mathrm{D}|\boldsymbol{u}|}{\mathrm{D}t} = 0 \tag{27.32a}$$

$$\frac{|\boldsymbol{u}|^2}{R} = -\frac{1}{\rho_0}\frac{\partial p}{\partial n}. \tag{27.32b}$$

Again, equation (27.32a) says that the speed is constant following a material fluid particle. Equation (27.32b) says that cyclostrophic flow occurs when the centrifugal acceleration balances the pressure gradient, with the squared speed given by

$$|\boldsymbol{u}|^2 = -\frac{R}{\rho_0}\frac{\partial p}{\partial n}. \tag{27.33}$$

This equation can be satisfied for either clockwise or counter-clockwise motion around a low pressure center, as shown in Figure 27.6. For clockwise flow, the radius of curvature is negative, $R < 0$, 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 radiuses 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 our neglect of 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.

27.5 Gradient wind balance

Gradient wind balance occurs under the following conditions:

- fluid particles move along lines of constant pressure so that $\partial p/\partial s = 0$;

Figure 27.6: *Cyclostrophic motion of a fluid particle on a plane occurs when the flow is horizontal, 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. 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.*

- vanishing friction.

The resulting equations of motion (27.13a) and (27.13b) take the form

$$\frac{\mathrm{D}|\boldsymbol{u}|}{\mathrm{D}t} = 0 \tag{27.34a}$$

$$\frac{|\boldsymbol{u}|^2}{R} + f\,|\boldsymbol{u}| = -\frac{1}{\rho_0}\frac{\partial p}{\partial n}, \tag{27.34b}$$

Again, equation (27.34a) says that the speed is constant following a material fluid particle. Equation (27.34b) says that gradient wind balanced flow occurs when the centrifugal and Coriolis accelerations balance the pressure gradient acting normal to the motion.

27.5.1 Constraints on gradient wind flow

The quadratic formula leads to the following expression for the speed of gradient wind flow

$$|\boldsymbol{u}| = \frac{R}{2}\left[f \pm \sqrt{f^2 - \frac{4}{R}\frac{1}{\rho_0}\frac{\partial p}{\partial n}}\right]. \tag{27.35}$$

The speed is a real number if the pressure gradient, Coriolis parameter, and radius of curvature satisfy

$$f^2 > \frac{4}{R}\frac{1}{\rho_0}\frac{\partial p}{\partial n} \Longrightarrow \frac{1}{\rho_0}\frac{\partial p}{\partial n} < \frac{f^2 R}{4}. \tag{27.36}$$

This constraint means 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 assymetry 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).

27.5.2 The variety of gradient wind flows

Section 3.2 of *Holton* (1992) identifies the following force balances available with a gradient wind balance. We illustrate these cases in Figure 27.7.

Figure 27.7: *The variety of gradient wind balances available in the northern hemisphere ($f > 0$). Gradient wind motion 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: $\boldsymbol{P} = -\rho_0^{-1}\,\partial p/\partial n$, $\boldsymbol{C}o = f\,|\boldsymbol{u}|$, and $\boldsymbol{C}e = |\boldsymbol{u}|^2/R$. 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.*

- REGULAR HIGH PRESSURE CENTER (RIGHT TURN WITH HIGH PRESSURE ON RIGHT): This flow occurs withIn $R < 0$ and $\partial p/\partial n < 0$. 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 27.7). Furthermore, the inequality (27.36) provides a bound to the size of the pressure gradient so that

$$\frac{1}{\rho_0}\frac{\partial p}{\partial n} \leq \frac{R\,f^2}{4}. \tag{27.37}$$

  That is, the high pressure center cannot be larger than this bound in order for there to be a real solution. Note the *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 27.7).

- REGULAR LOW (LEFT TURN WITH LOW PRESSURE ON LEFT): This flow occurs with $R > 0$ and $\partial p/\partial n > 0$. This case occurs with the Coriolis and centrifugal accelerations pointing away from the low pressure center, and these two accelerations balance the pressure acceleration

pointing toward the center (upper right panel of Figure 27.7). The inequality (27.36) provides no bound to the magnitude of the low pressure.

- ANOMALOUS LOW (RIGHT TURN WITH LOW PRESSURE ON RIGHT): This flow occurs with $R < 0$ and $\partial p/\partial n > 0$. 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 27.7). The inequality (27.36) 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.

27.5.3 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. However, 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) makes use of the analysis in this section for the study of vortices in ocean flows. The deviations from geostrophy become important when considering relatively small vortices and/or vorticies in the tropics.

28

Ekman mechanics

Ekman mechanics explains key features of how momentum and matter are transferred across the turbulent boundary layers into the quasi-adiabatic geostrophic interior of the ocean and atmosphere. The Ekman boundary layer is affected by rotation and as such it exhibits many behaviors that are distinct relative to non-rotating boundary layers. For example, the Ekman layer imparts a stretching and squeezing of interior fluid columns that strongly couples to vorticity and circulation of the interior fluid regime. Ekman layers thus play a fundamental role in oceanic and atmospheric general circulation.

In this chapter we explore the rudimentary mechanics of Ekman layers. In Ekman layers we find a balance between the horizontal presure gradient, Coriolis acceleration , and vertical frictional acceleration. These layers realize a horizontally divergent/convergent mass transport that leads to the vertical exchange of mass, momentum, and vorticity with the fluid interior.

READER'S GUIDE TO THIS CHAPTER

We here assume an understanding of geostrophic flow from Chapter 26 and the description of flow using natural coordinates in Chapter 27. Ekman mechanics is a primary topic of geophysical fluid mechanics and applications to general circulation of the ocean and atmosphere. However, we do not make much use of Ekman layer mechanics in subsequent chapters, thus making this chapter largely self-contained.

28.1 Dynamical balance within Ekman layers **494**
28.2 Natural coordinates according to isobars **494**
28.3 Ekman balance in natural coordinates **495**
28.4 Cross isobar flow driven by Rayleigh drag **495**
28.4.1 What is Rayleigh drag? 496
28.4.2 Ekman velocity driven by Rayleigh drag 496
28.4.3 Horizontal spiral plus vertical rising/sinking 497
28.5 Laplacian vertical friction **497**
28.6 Ekman number and Ekman layer thickness **498**
28.6.1 Non-dimensionalization 498
28.6.2 The Ekman number 499
28.6.3 Ekman layer thickness 500
28.6.4 Estimates for the vertical eddy viscosity 500
28.6.5 Ekman layer thickness with Rayleigh drag 500
28.7 Ekman layer integrated mass transport **501**

28.7.1 Horizontal Ekman mass transport 501
28.7.2 Vertical transport into or out of the Ekman layer 502
28.7.3 Ekman and geostrophic mass transports 503
28.8 Further study **504**

28.1 Dynamical balance within Ekman layers

To understand the key features of Ekman layers it is sufficient to study a Boussinesq hydrostatic fluid within a turbulent boundary layer. The balance of accelerations in the momentum equation is between Coriolis, pressure, and friction

$$f\,\hat{\boldsymbol{z}} \wedge \boldsymbol{u} = -\frac{1}{\rho_o}\nabla_z p + \boldsymbol{F}. \tag{28.1}$$

In this chapter we focus on a frictional acceleration, $\boldsymbol{F}$, arising from the vertical exchange of horizontal momentum between fluid layers. We ignore the centrifugal acceleration described in Section 27.1.3, even though the motion generally has curvature.

It is useful to separate velocity into a geostrophic contribution defined by a balance between pressure gradient and Coriolis accelerations

$$f\,\hat{\boldsymbol{z}} \wedge \boldsymbol{u}_{\mathrm{g}} = -\frac{1}{\rho_o}\nabla_z p \Longrightarrow \boldsymbol{u}_{\mathrm{g}} = \frac{1}{f\rho_0}\,\hat{\boldsymbol{z}} \wedge \nabla p \tag{28.2}$$

plus an ageostrophic or Ekman contribution defined by a balance between the frictional and Coriolis accelerations

$$f\,\hat{\boldsymbol{z}} \wedge \boldsymbol{u}_{\mathrm{e}} = \boldsymbol{F} \Longrightarrow \boldsymbol{u}_{\mathrm{e}} = -f^{-1}\,\hat{\boldsymbol{z}} \wedge \boldsymbol{F}. \tag{28.3}$$

This decomposition has the appearance of superposing linearly independent flows, one geostrophic and one ageostrophic. And yet the flows are quite coupled. Namely, ageostrophic motions alter the pressure field which in turn affects the geostrophic flow. So the presence of friction and the associated ageostrophic flows leads to a geostrophic flow that differs from the inviscid case. Hence, the above decomposition is motivated by mathematical convenience and does not reflect a physical decoupling of the geostrophic and ageostrophic flows.

28.2 Natural coordinates according to isobars

As per the definition (28.2), geostrophic motion occurs along lines of constant pressure. We are here interested in deviations from geostrophic motion realized by adding friction: how does friction support horizontal motion across isobars? To help answer this question it is useful to decompose the horizontal flow into natural coordinates along an arbitrary geopotential surface. But instead of defining the natural coordinates according to the flow direction, as done in Section 27.1, we here define the directions according to pressure contours (isobars)

$$\hat{\boldsymbol{z}} = \hat{\boldsymbol{s}} \wedge \hat{\boldsymbol{n}} = \text{vertical direction} \tag{28.4a}$$

$$\hat{\boldsymbol{n}} = \hat{\boldsymbol{z}} \wedge \hat{\boldsymbol{s}} = \text{down pressure gradient direction} \tag{28.4b}$$

$$\hat{\boldsymbol{s}} = \hat{\boldsymbol{n}} \wedge \hat{\boldsymbol{z}} = \text{tangent to isobar.} \tag{28.4c}$$

The unit vector $\hat{\boldsymbol{s}}$ is tangent to isobars in the horizontal plane, whereas $\hat{\boldsymbol{n}}$ is perpendicular to isobars and oriented down the horizontal pressure gradient

$$\hat{\boldsymbol{n}} = -\frac{\nabla_z p}{|\nabla_z p|}. \tag{28.5}$$

Figure 28.1: *Natural coordinates defined along an arbitrary geopotential surface according to isobars in the horizontal plane. The normal direction, $\hat{\boldsymbol{n}} = -\nabla_z p/|\nabla_z p|$, 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, $\hat{\boldsymbol{s}}$, points along the isobar so that $\hat{\boldsymbol{s}} \wedge \hat{\boldsymbol{n}} = \hat{\boldsymbol{z}}$.*

These natural coordinates are illustrated in Figure 28.1.

28.3 Ekman balance in natural coordinates

We now represent the geostrophic and Ekman velocities in natural coordinates. As found in studying the geostrophic balance in Section 27.2, the geostrophic velocity flows along isobars and so only has a component in the $\hat{\boldsymbol{s}}$ direction

$$\hat{\boldsymbol{s}} \cdot \boldsymbol{u}_{\mathrm{g}} = -\frac{1}{f\rho_0}\frac{\partial p}{\partial n} \tag{28.6a}$$

$$\hat{\boldsymbol{n}} \cdot \boldsymbol{u}_{\mathrm{g}} = 0. \tag{28.6b}$$

In contrast, the Ekman velocity has a component both along and across isobars

$$\hat{\boldsymbol{s}} \cdot \boldsymbol{u}_{\mathrm{e}} = f^{-1}\,\hat{\boldsymbol{n}} \cdot \boldsymbol{F} \tag{28.7a}$$

$$\hat{\boldsymbol{n}} \cdot \boldsymbol{u}_{\mathrm{e}} = -f^{-1}\,\hat{\boldsymbol{s}} \cdot \boldsymbol{F}. \tag{28.7b}$$

As expected, the Ekman velocity vanishes when the frictional acceleration vanishes. These equations show that friction aligned along and across isobars drive Ekman velocities across and along isobars, respectively. If there is no friction component along the isobar, $\hat{\boldsymbol{s}} \cdot \boldsymbol{F} = 0$, 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.

28.4 Cross isobar flow driven by Rayleigh drag

We here assume the frictional acceleration takes the form of Rayleigh drag. As noted in Section 7.4 of *Marshall and Plumb* (2008), the relative simplicity of Rayleigh drag affords analytical expresions for the Ekman velocity in terms of the geostrophic velocity, offering an explicit illustration of how the Ekman flow provides a cross-isobar component to the flow in the direction down the pressure gradient.

Figure 28.2: *Illustrating the flow associated with Ekman dynamics in the northern hemisphere ($f > 0$). Left panel: geostrophic flow, $\boldsymbol{u}_g$, around a low pressure center is counter-clockwise and aligned with pressure isosurfaces. Friction aligned drives Ekman 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.*

28.4.1 What is Rayleigh drag?

We take the frictional acceleration in the form of a Rayleigh drag acting on the velocity field

$$\boldsymbol{F}^{\text{drag}} = -\frac{U_{\text{fric}}\,\boldsymbol{u}}{\delta} = -\gamma\,\boldsymbol{u}, \tag{28.8}$$

where δ is a depth scale and U_{fric} is a friction velocity scale with dimensions L/T. The ratio

$$\gamma = \frac{U_{\text{fric}}}{\delta} \tag{28.9}$$

has dimensions T^{-1} and can be thought of as an inverse spin-down time. That is, if only Rayleigh drag affected changes to the horizontal momentum, $\partial_t \boldsymbol{u} = -\gamma\,\boldsymbol{u}$, 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.

Rayleigh drag dissipates all flow features regardless of their structure; i.e., mathematically it does not prefer any particular length scales in the fluid flow. This lack of scale selectivity contrasts to the Laplacian friction discussed in Section 28.5, with Laplacian friction dissipating small scales more rapidly than the large scales.

28.4.2 Ekman velocity driven by Rayleigh drag

Making use of the Rayleigh drag (28.8) brings the expressions (28.7a) and (28.7b) for the Ekman velocity into the form

$$\hat{\boldsymbol{s}} \cdot \boldsymbol{u}_{\text{e}} = -(\gamma/f)\,\hat{\boldsymbol{n}} \cdot \boldsymbol{u}_{\text{e}} \tag{28.10a}$$

$$\hat{\boldsymbol{n}} \cdot \boldsymbol{u}_{\text{e}} = (\gamma/f)\,\hat{\boldsymbol{s}} \cdot (\boldsymbol{u}_{\text{e}} + \boldsymbol{u}_{\text{g}}). \tag{28.10b}$$

These equations allow us to express the Ekman velocity in terms of the geostrophic velocity

$$\hat{\boldsymbol{s}} \cdot \boldsymbol{u}_{\text{e}} = -(\gamma/f)\,\hat{\boldsymbol{n}} \cdot \boldsymbol{u}_{\text{e}} \tag{28.11a}$$

$$\hat{\boldsymbol{n}} \cdot \boldsymbol{u}_{\text{e}} = \hat{\boldsymbol{s}} \cdot \boldsymbol{u}_{\text{g}} \left[\frac{(\gamma/f)}{1 + (\gamma/f)^2}\right], \tag{28.11b}$$

so that the velocity components for the total velocity, $\boldsymbol{u} = \boldsymbol{u}_{\mathrm{g}} + \boldsymbol{u}_{\mathrm{e}}$, and its squared magnitude, are

$$\hat{\boldsymbol{s}} \cdot \boldsymbol{u} = \hat{\boldsymbol{s}} \cdot \boldsymbol{u}_{\mathrm{g}} \left[\frac{1}{1 + (\gamma/f)^2} \right] \tag{28.12a}$$

$$\hat{\boldsymbol{n}} \cdot \boldsymbol{u} = (\gamma/f)\, \hat{\boldsymbol{s}} \cdot \boldsymbol{u} \tag{28.12b}$$

$$(\hat{\boldsymbol{s}} \cdot \boldsymbol{u})^2 + (\hat{\boldsymbol{n}} \cdot \boldsymbol{u})^2 = \frac{(\hat{\boldsymbol{s}} \cdot \boldsymbol{u}_{\mathrm{g}})^2}{1 + (\gamma/f)^2}. \tag{28.12c}$$

As expected, the cross-isobar flow 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, $\hat{\boldsymbol{s}} \cdot \boldsymbol{u} > 0$. Rayleigh drag thus causes the fluid to spiral into low pressure centers and away from a high pressure centers, such as that depicted in Figure 28.2. Furthermore, the magnitude of the total flow is reduced relative to the geostrophic flow, thus reflecting the dissipation of kinetic energy from drag.

28.4.3 Horizontal spiral plus vertical rising/sinking

A robust feature of Ekman flow is the spiraling motion of the fluid in the direction down the pressure gradient. This feature holds for other forms of friction, though it is more difficult to show analytically than in the case of Rayleigh drag. So far, we have focused on the horizontal spiral motion as shown in Figure 28.2. However, there is a corresponding vertical motion as well, with the 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 an atmospheric boundary layer whereby mass rises above a low pressure center Ekman layer in response to the horizontal convergence. Conversely, mass diverges from the high pressure Ekman layer, thus induces a sinking motion over the high to replace the diverging mass.

In the remaining sections of this chapter we develop the formalism needed to compute the mass transport into and out of the Ekman layer. This mass transport plays an especially important role in how mechanical forcing drives the ocean general circulation.

28.5 Laplacian vertical friction

Frictional acceleration is generally not well defined since it depends on details of the turbulence and how it is parameterized. We made use of Rayleigh drag in Section 28.4 for mathematical expediency. Even so, it is useful to here offer a few words about the more conventional Laplacian friction operator given by

$$\boldsymbol{F}^{\text{viscous}} = \frac{1}{\rho_o} \frac{\partial \boldsymbol{\tau}}{\partial z} = \frac{\partial}{\partial z} \left[A \frac{\partial \boldsymbol{u}}{\partial z} \right], \tag{28.13}$$

with $A > 0$ a vertical turbulent kinematic viscosity with dimensions L^2/T. This form of the friction operator emulates the Laplacian operator representing molecular viscous friction (equation (19.37)). It is also that form most commonly used in numerical models and it was used by Ekman in his investigations leading to the development of Ekman theory.

Expanding the derivative reveals that the Laplacian friction operator 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

$$\boldsymbol{F}^{\text{viscous}} = \frac{\partial A}{\partial z} \frac{\partial \boldsymbol{u}}{\partial z} + A \frac{\partial^2 \boldsymbol{u}}{\partial z^2}. \tag{28.14}$$

Figure 28.3: *Illustrating the three-dimensional flow associated with Ekman layers in the northern hemisphere ($f > 0$) atmosphere. 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 entering the interior atmosphere above. Right panel: flow spiralling away from a high pressure center creates divergence of mass out 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 28.6). An analogous picture holds for the surface Ekman layer in the ocean, yet with the Ekman layer at the top of the column.*

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 47.3 we discuss the mathematical properties of Laplacian friction.

28.6 Ekman number and Ekman layer thickness

The Ekman layer thickness is a function of the vertical friction acting within the boundary layer. The more friction the thicker the boundary layer. However, there is no deductive theory for the friction arising from turbulence nor even for its appropriate mathematical form. Even so, we find it useful to here consider the Rayleigh drag and Laplacian viscosity forms introduced earlier and develop expressions for the boundary layer thickness based on these operators. Doing so offers intuition and some level of quantitative accuracy.

As per the usual process in fluid mechanics, we find it useful to non-dimensionalize the equations to isolate important non-dimensional numbers affecting the flow regime. We here isolate the Ekman number for boundary layers affected by friction and rotation. Friction is important where the Ekman number is order unity. Note that flows unaffected by rotation, it is the Reynolds number that proves to be the key non-dimensional number (Section 19.7) for boundary layers, rather than the Ekman number arising when rotation is important.

28.6.1 Non-dimensionalization

We make use of the following scaling

$$(x, y) = L\,(\hat{x}, \hat{y}) \qquad z = L_z\,\hat{z} \qquad (u, v) = U\,(\hat{u}, \hat{v}) \qquad f = f_0\,\hat{f} \qquad p = P\,\hat{p} \tag{28.15}$$

where the variables with hats are non-dimensional,[1] and we introduced typical scales for length, velocity, Coriolis, and pressure. As defined below, the vertical scale, L_z, is the thickness over which

[1] Be careful not to confuse $\hat{\boldsymbol{u}}$ used here for non-dimensional velocity with the $\hat{\boldsymbol{u}}$ unit vector pointing along the fluid particle trajectory as defined in Section 27.1.1.

friction is nontrivial, thus offering a scale for the Ekman layer thickness. The pressure scale, P, is assumed to follow geostrophic scaling

$$P = f_0 \, \rho_0 \, U \, L. \tag{28.16}$$

Inserting the relations (28.15) into equation (28.1) leads to the non-dimensional frictional geostrophic equation

$$\hat{\boldsymbol{f}} \wedge \hat{\boldsymbol{u}} = -\hat{\nabla}\hat{p} + \frac{\boldsymbol{F}}{f_0 \, U}. \tag{28.17}$$

28.6.2 The Ekman number

We here introduce a non-dimensional number, the *Ekman number*, that measures the importance of accelerations due to vertical friction versus the Coriolis acceleration.

The Ekman number for Laplacian vertical friction

The ratio of scales for the frictional acceleration to the Coriolis acceleration defines the Ekman number

$$Ek = \frac{\text{frictional acceleration}}{\text{Coriolis acceleration}}. \tag{28.18}$$

For the viscous stress form of Laplacian vertical frictional acceleration (equation (28.13))

$$\boldsymbol{F}^{\text{viscous}} = \frac{A \, U}{L_z^2} \, \frac{\partial^2 \hat{\boldsymbol{u}}}{\partial \hat{z}^2}, \tag{28.19}$$

so that the Ekman number is

$$Ek = \frac{A}{f_0 \, L_z^2}. \tag{28.20}$$

If we take the vertical scale, L_z, equal to the depth scale over which interior flow processes occur, then the Ekman number will generally be very small, in which case friction is negligible. However, this reasoning is flawed when closer to the boundary since the Ekman number multiplies the highest derivative in equation (28.17). Setting the Ekman number to zero thus represents a singular limit in the sense of perturbation theory. Hence, we expect a boundary layer to form, inside of which the Ekman number is order unity and the role of Laplacian vertical friction is nontrivial.

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

$$Ro = \frac{U}{f_0 \, L} \qquad \text{and} \qquad Ek = \frac{A}{f_0 \, L_z^2} \qquad \text{and} \qquad Re = \frac{W \, L_z}{A}, \tag{28.21}$$

where

$$L_z = \text{vertical length scale} \tag{28.22a}$$
$$L = \text{horizonal length scale} \tag{28.22b}$$
$$U = \text{horizonal velocity scale} \tag{28.22c}$$
$$W = \text{vertical velocity scale.} \tag{28.22d}$$

and we defined the Reynolds number (Section 19.7) in terms of the vertical viscosity, vertical length scale, and vertical velocity scale. The ratio of the Rossby to Reynolds number is thus given by

$$\frac{Ro}{Re} = \frac{U}{f_0\, L}\, \frac{A}{W\, L_z} = Ek\, \frac{L_z}{L}\, \frac{U}{W}. \tag{28.23}$$

If we now assume the fluid is incompressible, then the continuity equation $\nabla \cdot \boldsymbol{v} = 0$ means that the vertical velocity scales according to $W = U\, L_z/L$, in which case $Ek = Ro/Re$. However, as shown in Section 42.2.1, the vertical velocity for a small Rossby number flow scales like

$$W = Ro\, U\, \frac{L_z}{L} \Longrightarrow Ek = \frac{Ro^2}{Re}. \tag{28.24}$$

28.6.3 Ekman layer thickness

The Ekman number increases when there is more boundary layer turbulence, in which case the eddy viscosity, A, is large relative to its small values in the interior. Additionally, the Ekman number increases when moving towards the equator, where the Coriolis parameter reduces.[2] The viscous friction acceleration, (28.19), is order unity over a depth scale where the Ekman number is order unity, which occurs within a boundary layer. This vertical scale defines the viscous Ekman boundary layer thickness

$$Ek = 1 \Rightarrow \delta^{\text{viscous}} = \left[\frac{A}{f_0}\right]^{1/2}. \tag{28.25}$$

28.6.4 Estimates for the vertical eddy viscosity

The eddy viscosity is not readily available from measurements or theory. However, if we can observe the boundary layer thickness, then we have a means to infer the scale for the viscosity

$$A = f_0\, (\delta^{\text{viscous}})^2. \tag{28.26}$$

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

$$A^{\text{atmos}} \sim 10^2\ \text{m}^2\ \text{s}^{-1}. \tag{28.27}$$

In the ocean, the upper ocean boundary layer depth, outside of the deep convection regions, is roughly 50 m, in which case

$$A^{\text{ocean}} \sim 0.25\ \text{m}^2\ \text{s}^{-1}. \tag{28.28}$$

28.6.5 Ekman layer thickness with Rayleigh drag

For the Rayleigh drag form of friction given by

$$\boldsymbol{F}^{\text{drag}} = -\left[\frac{U_{\text{fric}}}{\delta}\right] U\, \hat{\boldsymbol{u}}, \tag{28.29}$$

the Ekman number takes the form

$$Ek = \frac{U_{\text{fric}}}{\delta\, f_0}. \tag{28.30}$$

[2] When getting too 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.

Figure 28.4: *Ekman layer at the ocean surface, defined for vertical position $-\delta(x,y,t) \leq z \leq \eta(x,y,t)$, with δ specifying the Ekman layer bottom and η the free surface vertical position. Wind stress imparts horizontal momentum to the upper ocean that is transmitted throughout the Ekman layer via turbulent stresses. In addition, mass flux, Q_m, crosses the ocean free surface thus affecting the mass budget in the Ekman layer ($Q_m > 0$ means mass enters the ocean). There is also a bottom Ekman layer created by bottom boundary layer turbulence.*

The Ekman number is unity, and hence friction important, if the thickness δ takes on the Ekman value

$$\delta^{\text{drag}} = \frac{U_{\text{fric}}}{f_0}. \tag{28.31}$$

The Laplacian and Rayleigh expressions for the Ekman layer thicknesses are the same if the frictional velocity scale and viscosity are related by

$$\delta^{\text{drag}} = \delta^{\text{viscous}} \Longleftrightarrow U_{\text{fric}} = \sqrt{A\, f_0}. \tag{28.32}$$

28.7 Ekman layer integrated mass transport

It is possible to establish some integral properties of the Ekman layer even without knowing details of the friction. A particularly important quantity is the integrated mass transport crossing the base of the ocean surface Ekman layer sitting between the Ekman layer bottom and the ocean free surface

$$-\delta(x,y,t) \leq z \leq \eta(x,y,t), \tag{28.33}$$

as depicted in Figure 28.4. Knowledge of this mass transport has important implications for how mechanical energy imparted to the boundary layer drives circulation within the interior of the fluid.

28.7.1 Horizontal Ekman mass transport

Integrating the horizontal Ekman balance (28.3) over the depth of the Ekman layer leads to

$$\boldsymbol{f} \wedge \boldsymbol{M}_{\text{e}} = \int_{-\delta}^{\eta} \rho_o\, \boldsymbol{F}\, \mathrm{d}z, \tag{28.34}$$

where

$$\boldsymbol{M}_{\text{e}} = \int_{-\delta}^{\eta} \rho_o\, \boldsymbol{u}_{\text{e}}\, \mathrm{d}z \tag{28.35}$$

Figure 28.5: *Horizontal transport integrated over the depth of the surface ocean Ekman layer is directed at right angles to the wind stress in the northern hemisphere (to the left in the southern). Here, the wind stress, $\boldsymbol{\tau}$, is shown blowing to the north so that in the northern hemisphere ($f > 0$), the depth integrated horizontal Ekman transport, $\boldsymbol{M}_e$, is to the east. This result from Ekman dynamics is independent of the assumptions made about friction within the boundary layer.*

is the depth integrated ageostrophic mass transport within the Ekman boundary layer. Assume friction in the form of a vertical viscous stress (equation (28.13)), so that

$$\boldsymbol{M}_{\mathrm{e}} = -f^{-1}\,\hat{\boldsymbol{z}} \wedge [\boldsymbol{\tau}(\eta) - \boldsymbol{\tau}(-\delta)]. \tag{28.36}$$

Stress at the bottom of the Ekman layer, $\boldsymbol{\tau}(-\delta)$, matches to the stress in the ocean interior, which is much smaller than stress at the ocean surface, $\boldsymbol{\tau}(\eta)$. Neglecting this interior stress then leads to the ageostrophic Ekman layer horizontal mass transport

$$\boldsymbol{M}_{\mathrm{e}} = -f^{-1}\,\hat{\boldsymbol{z}} \wedge \boldsymbol{\tau}(\eta). \tag{28.37}$$

Mass transport in the Ekman layer is thus at right angles to the surface stress. This orientation is a consequence of the Coriolis force. It is also a major result from Ekman's original studies of ice motion in the Arctic.

28.7.2 Vertical transport into or out of the Ekman layer

As seen in Figure 28.3, the horizontal Ekman transport into or out of the Ekman layer induces a vertical transport into or out of the Ekman layer. To obtain an expression for the vertical transport, integrate the continuity equation $\nabla \cdot \boldsymbol{v} = 0$ over the vertical extent of the Ekman layer

$$\frac{\partial}{\partial x}\left[\int_{-\delta}^{\eta} u\,\mathrm{d}z\right] + \frac{\partial}{\partial y}\left[\int_{-\delta}^{\eta} v\,\mathrm{d}z\right] + [w(\eta) - \boldsymbol{u}(\eta)\cdot\nabla\eta] - [w(-\delta) + \boldsymbol{u}(-\delta)\cdot\nabla\delta] = 0. \tag{28.38}$$

For an incompressible fluid, the kinematic boundary condition at the ocean free surface is given by equation (16.5)

$$w + \frac{Q_{\mathrm{m}}}{\rho_0} = \frac{\partial\eta}{\partial t} + \boldsymbol{u}\cdot\nabla\eta \qquad z = \eta(x,y,t). \tag{28.39}$$

Similarly, at the bottom of the Ekman layer we measure the volume transport through this layer by computing the dia-surface transport, $w^{(\dot{\delta})}$, according to equation (17.37)

$$w^{(\dot{\delta})} = w - (\partial_t + \boldsymbol{u}\cdot\nabla)\,z \qquad z = -\delta(x,y,t). \tag{28.40}$$

The sign convention is such that $w^{(\dot{\delta})} > 0$ means that mass enters the Ekman layer through its base, whereas mass leaves through the base when $w^{(\dot{\delta})} < 0$.

Figure 28.6: *Mass budget over the surface Ekman layer of the ocean, with impacts from surface mass transport Q_m, transport through the bottom of the layer, $w^{(\dot{\delta})}$, and horizontal transport, $\boldsymbol{M}$. If these transports static. But if there are any imbalances then the layer thickness will have a nonzero time tendency, $\partial h_e/\partial y \neq 0$.*

Using the kinematic boundary conditions (28.39) and (28.40) in the depth integrated volume budget (28.38), and rearranging, leads to the expression for the volume transport crossing the surface and bottom of the Ekman layer

$$w^{(\dot{\delta})} = \frac{\partial h_{\mathrm{e}}}{\partial t} + \frac{-Q_{\mathrm{m}} + \nabla \cdot \boldsymbol{M}}{\rho_0} \tag{28.41}$$

where

$$h_{\mathrm{e}} = \eta + \delta \tag{28.42}$$

is the thickness of the Ekman layer and

$$\boldsymbol{M} = \int_{-\delta}^{\eta} \rho_0 \, \boldsymbol{u} \, \mathrm{d}z \tag{28.43}$$

is the Ekman layer integrated horizontal mass transport. It is useful as a sanity check to note that letting the Ekman layer go to the ocean bottom (so that $w^{(\dot{\delta})} = 0$) reduces equation (28.41) to the kinematic free surface equation for the full ocean column (equation (16.8))

$$\frac{\partial \eta}{\partial t} = \frac{Q_{\mathrm{m}}}{\rho_0} - \nabla \cdot \boldsymbol{U}. \tag{28.44}$$

Bringing the surface and bottom boundary terms together in equation (28.41) helps with its interpretation

$$\rho_0 \, w^{(\dot{\delta})} + Q_{\mathrm{m}} = \rho_0 \, \frac{\partial h_{\mathrm{e}}}{\partial t} + \nabla \cdot \boldsymbol{M}. \tag{28.45}$$

The left hand side measures the mass transport crossing the bottom of the Ekman layer plus the free surface. This transport balances a time change in the Ekman layer thickness plus the divergence of mass from the layer. A steady state layer thickness ($\partial h_{\mathrm{e}}/\partial t = 0$) is realized if the divergence of mass through horizontal transport is exactly balanced by mass entering the boundary layer through the top and/or bottom of the layer. We illustrate this budget in Figure 28.6.

28.7.3 Ekman and geostrophic mass transports

The horizontal mass transport given by equation (28.43) has a contribution from both the geostrophic and Ekman velocities

$$\boldsymbol{M} = \boldsymbol{M}_{\mathrm{g}} + \boldsymbol{M}_{\mathrm{e}}. \tag{28.46}$$

The geostrophic velocity is horizontally non-divergent on an f-plane (Section 26.4.3), yet generally has a nonzero divergence on the sphere. The horizontal Ekman transport is determined largely by the wind stress according to equation (28.37), with its divergence given by

$$\nabla \cdot \boldsymbol{M}_{\mathrm{e}} = \hat{\boldsymbol{z}} \cdot [\nabla \wedge (\boldsymbol{\tau}/f)]. \tag{28.47}$$

This result brings the Ekman layer mass budget (28.45) into the form

$$\rho_0 \, w^{(\dot{\delta})} + Q_{\mathrm{m}} = \rho_0 \, \frac{\partial h_{\mathrm{e}}}{\partial t} + \nabla \cdot \boldsymbol{M}_{\mathrm{g}} + \hat{\boldsymbol{z}} \cdot [\nabla \wedge (\boldsymbol{\tau}/f)]. \tag{28.48}$$

The effects from wind stress in equation 28.48 warrant particular attention. The wind 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 nearly inviscid interior. These interior fluid columns are stiffened through the effects of Taylor-Proudman (Section 26.4.3), thus coupling the Ekman layer to a vast portion of the interior flow. From our understanding of vorticity (studied in Chapter 34), particularly the notions of vortex stretching, we see that the Ekman induced streching/compression of interior fluid columns leads to a change in vorticity of the fluid interior, thus driving motion that affects a broad scale circulation. This coupling between the Ekman layer to the interior fluid is a fundamental feature of rotating fluids that is missing from non-rotating boundary layer flows. These general ideas of Ekman layer mechanics form the basis for how mechanical forcing from the winds drives ocean vorticity and the large-scale ocean general circulation.

28.8 Further study

This chapter borrows from Section 5.7 of *Vallis* (2017) and Section 7.4 of *Marshall and Plumb* (2008). The student is also encouraged to read Chapter 8 of *Cushman-Roisin and Beckers* (2011) for further insights.

For visualization, the following videos are recommended:

- 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.

Part V

Shallow water mechanics

Shallow water models consist of constant density 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 purely on the dynamics of perfect fluid layers. Momentum is transferred between layers through pressure forces that act on sloping layer interfaces. Furthermore, dynamical motion occurs in columns, with horizontal velocity independent of vertical position within a layer.

The shallow water model provides us with a suite of versatile theoretical models of use to deduce fluid dynamical impacts from both rotation and stratification. It is among the most popular fluid models for theorists. Consequently, the shallow water model features heavily in many areas of geophysical fluid mechanics as well as in applications to the ocean and atmosphere.

29

Shallow water models

We here formulate the kinematic and dynamic equations for a suite of shallow water models by developing equations for a single shallow water layer; multiple shallow water layers (stacked shallow water); and reduced gravity models (models with one layer that is dynamically inactive).

READER'S GUIDE TO THIS CHAPTER

This chapter is the mathematical and physical basis for subsequent discussions that involve the shallow water model. It relies on elements of fluid kinematics and dynamics described in earlier chapters.

29.1 A single shallow water layer . . . **508**
- 29.1.1 Pressure within the fluid layer . . . 508
- 29.1.2 Momentum equation . . . 509
- 29.1.3 Thickness equation . . . 510
- 29.1.4 Kinematic boundary conditions . . . 512
- 29.1.5 Stretching and vertical velocity . . . 513
- 29.1.6 Further study . . . 514

29.2 Emphasizing the hydrostatic approximation . . . **514**
- 29.2.1 Hydrostatic balance with respect to background density . . . 515
- 29.2.2 Density is a uniform constant . . . 515
- 29.2.3 Fluid is incompressible . . . 515
- 29.2.4 Non-hydrostatic and hydrostatic forces in a homogeneous layer . . . 516
- 29.2.5 Comments . . . 516

29.3 Shallow water fluid in a rotating tank . . . **516**
- 29.3.1 Equations of motion . . . 516
- 29.3.2 Free surface shape in solid-body rotation . . . 517
- 29.3.3 Further study . . . 517

29.4 Reduced gravity model for the upper ocean . . . **517**
- 29.4.1 Momentum and thickness equations for the active layer . . . 517
- 29.4.2 Relating undulations of the top and bottom layer interfaces . . . 518
- 29.4.3 Momentum equation with reduced gravity . . . 519
- 29.4.4 Further study . . . 519

29.5 Stacked shallow water equations . . . **520**
- 29.5.1 Model formulation . . . 520
- 29.5.2 Further study . . . 522

29.6 Exercises . . . **522**

29.1 A single shallow water layer

Consider a homogeneous layer of fluid in a uniform effective gravitational field (gravity plus planetary centrifugal), contained on its side boundaries by vertical walls. 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 (volume in a uniform density layer) means that thicker fluid regions must come at the cost of thinner fluid regions. Furthermore, layer thickness gradients create pressure differences (thicker water has larger hydrostatic bottom 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 motion.

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 that ensures that the accumulation of fluid in one region is balanced by 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.

Figure 29.1: *A single layer of shallow water fluid with h the thickness of the water column, H the averaged fluid depth (if the area integral of η_b is zero), $z = \eta$ the free surface height measured with respect to $z = 0$, and $z = \eta_b$ the height of the bottom boundary measured with respect to $z = 0$. Note that $z = 0$ is usefully chosen so that the area average of η_b is zero. Furthermore, with $\Delta\eta$ the deviation of the free surface relative to the averaged depth, H, volume conservation means that the area integral of $\Delta\eta$ vanishes. In summary, we have $\eta(x, y, t) = \eta_b(x, y) + h(x, y, t) = H + \Delta\eta(x, y, t)$ as well as $\int \eta_b \, \mathrm{d}x \, \mathrm{d}y = \int \Delta\eta \, \mathrm{d}x \, \mathrm{d}y = 0$. Additionally, we are concerned with fluctuations that leave the free surface monotonic; i.e., we do not consider overturns or breaking shallow water waves.*

29.1.1 Pressure within the fluid layer

Figure 29.1 shows a single shallow water layer with a generally non-flat bottom and an undulating free surface height. We assume the fluid to be in hydrostatic balance, so that the vertical momentum equation reduces to

$$\frac{\partial p}{\partial z} = -\rho \, g. \tag{29.1}$$

Recall from Section 24.3 that the hydrostatic balance is consistent with lateral length scales much larger than vertical (small vertical to horizontal aspect ratio). Hence, a shallow water fluid is a relevant idealization if we are considering large horizontal scales relative to the vertical. This configuration is common for large-scale geophysical fluids.

Figure 29.2: *The horizontal acceleration from pressure within a single shallow water fluid layer is determined by the surface height, $-\nabla p = -g\,\nabla\eta$. The acceleration is uniform throughout the layer and points from sea level highs towards sea level lows.*

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

$$p(x,y,z,t) = p_o(x,y,t) + g\,\rho \int_z^\eta \mathrm{d}z \tag{29.2a}$$

$$= p_o(x,y,t) + g\,\rho\,[\eta(x,y,t) - z], \tag{29.2b}$$

where $p_o(x,y,t)$ is the pressure applied to the layer free surface, say from the overlying atmosphere. Furthermore, the horizontal pressure gradient thus takes the form

$$\nabla_z p = \nabla_z p_o + g\,\rho\,\nabla_z \eta. \tag{29.3}$$

Since p_o 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. We generaly ignore the applied surface pressure, p_o, since the fluid above the layer is assumed to have zero inertia. However, that assumption can be relaxed in order to study the effects of atmospheric pressure on a single layer ocean, for example. With $p_o = 0$, horizontal pressure forces within the fluid layer are determined solely by undulations in the free surface

$$\nabla_z p = g\,\rho\,\nabla\eta \qquad \text{if } p_o = 0. \tag{29.4}$$

As depicted in Figure 29.2, acceleration from the pressure force is uniform throughout the layer and it points from sea level highs to sea level lows.

29.1.2 Momentum equation

If there is no friction anywhere in the fluid, then the horizontal momentum is effected only by the Coriolis and pressure forces. In this case, the horizontal momentum equation takes the form

$$\frac{\mathrm{D}\boldsymbol{u}}{\mathrm{D}t} + \boldsymbol{f} \wedge \boldsymbol{u} = -g\nabla\eta, \tag{29.5}$$

where

$$\boldsymbol{v} = (\boldsymbol{u}, w) \tag{29.6}$$

splits out the horizontal velocity vector, $\boldsymbol{u}$, from the vertical velocity component, w.

The Coriolis parameter, $\boldsymbol{f}$, is independent of depth, as is the horizontal pressure force. 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

$$\frac{\mathrm{D}\boldsymbol{u}}{\mathrm{D}t} = \left[\frac{\partial}{\partial t} + u\,\frac{\partial}{\partial x} + v\,\frac{\partial}{\partial y}\right]\boldsymbol{u} \tag{29.7}$$

so that the shallow water momentum equation (29.5) takes on the form

$$\left[\frac{\partial}{\partial t} + u\,\frac{\partial}{\partial x} + v\,\frac{\partial}{\partial y} + f\,\hat{\boldsymbol{z}}\wedge\right]\boldsymbol{u} = -g\nabla\eta. \tag{29.8}$$

29.1.3 Thickness equation

The mass of a shallow water layer is constant in the absence of mixing, sources, or sinks. 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 Exercise 29.2. 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.

Figure 29.3: *Mass budget for a column of shallow water fluid with cross-sectional area* $\mathrm{d}A$*, constant density* ρ*, and thickness* h*. In the absence of boundary mass fluxes, the column mass is affected only by horizontal transport.*

Consider an infinitesimal vertical column of shallow water fluid that is fixed in space. Let the horizontal cross-sectional area be written as $\mathrm{d}A$ and the thickness be

$$h = \eta - \eta_b \tag{29.9}$$

(see Figure 29.3). The total mass of fluid in this column is given by

$$M = \int \mathrm{d}A \int_{\eta_b}^{\eta} \rho\,\mathrm{d}z = \rho \int (\eta - \eta_b)\,\mathrm{d}A = \rho \int h\,\mathrm{d}A. \tag{29.10}$$

Time changes in the column mass thus arise from time changes in the layer thickness integrated over the horizontal area of the column

$$\frac{\mathrm{d}M}{\mathrm{d}t} = \rho \int \frac{\partial h}{\partial t}\,\mathrm{d}A, \tag{29.11}$$

where

$$\frac{\partial h}{\partial t} = \frac{\partial\,(\eta - \eta_b)}{\partial t} = \frac{\partial \eta}{\partial t}, \tag{29.12}$$

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, no mass crosses the top or bottom interfaces). The mass flux penetrating the vertical boundary is given by

$$\text{mass per time entering column} = -\rho \oint \boldsymbol{u} \cdot \hat{\boldsymbol{n}}\, \mathrm{d}\mathcal{S}, \tag{29.13}$$

where $\hat{\boldsymbol{n}}$ is the outward normal at the column boundary, and $\mathrm{d}\mathcal{S}$ 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 integral

$$\text{mass per time entering column} = -\rho \oint \mathrm{d}l \int \boldsymbol{u} \cdot \hat{\boldsymbol{n}}\, \mathrm{d}z, \tag{29.14}$$

where $\mathrm{d}l$ is the infinitesimal line element around the column circumference. Since $\hat{\boldsymbol{n}} \cdot \boldsymbol{u}$ is depth independent, we can perform the vertical integral to render

$$-\rho \oint \mathrm{d}l \int \boldsymbol{u} \cdot \hat{\boldsymbol{n}}\, \mathrm{d}z = -\rho \oint h\, \boldsymbol{u} \cdot \hat{\boldsymbol{n}}\, \mathrm{d}l \tag{29.15a}$$

$$= -\rho \int \nabla \cdot (h\, \boldsymbol{u})\, \mathrm{d}A, \tag{29.15b}$$

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 (29.11), and noting that the horizontal cross-sectional area is arbitrary, yields an equation for the layer thickness

$$\frac{\partial h}{\partial t} + \nabla \cdot (h\, \boldsymbol{u}) = 0. \tag{29.16}$$

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 (29.16) in the material form

$$\frac{\mathrm{D}h}{\mathrm{D}t} = -h\, \nabla \cdot \boldsymbol{u}. \tag{29.17}$$

Hence, thickness of a material fluid column increases in regions where the horizontal flow converges.

Special case with a rectangular column

To help solidify our understanding of the step (29.15b) in the above derivation, consider the special case of a rectangular column, for which the mass per time of fluid entering the column is given by

$$\text{mass per time entering column} = -\rho \int [(u\, h)_{\text{east}} - (u\, h)_{\text{west}}]\, \mathrm{d}y - \rho \int [(v\, h)_{\text{north}} - (v\, h)_{\text{south}}]\, \mathrm{d}x. \tag{29.18}$$

Taking the limit as the column becomes infinitesimal leads to

$$\text{mass per time entering column} = -\rho \int \left[\frac{\partial (u\, h)}{\partial x} + \frac{\partial (v\, h)}{\partial y}\right] \mathrm{d}x\, \mathrm{d}y = -\rho \int \nabla \cdot (h\, \boldsymbol{u})\, \mathrm{d}A, \tag{29.19}$$

thus recovering the result (29.15b).

29.1.4 Kinematic boundary conditions

Kinematic boundary conditions arise from geometric constraints placed on the fluid system. We consider here the kinematic boundary conditions at the ocean surface and bottom in the case where there is no flow through either interface. Recall from our discussion of fluid kinematics in Part III, we use the term *material surface* for any continuous surface or interface that is impenetrable to mass flow. The kinematics of such material surfaces is found throughout geophysical fluid mechanics. In Section 14.4 we derived the kinematic boundary conditions for a fluid, and we here apply those ideas to the shallow water system.

Bottom kinematic boundary condition

The ocean bottom is located at a vertical position

$$z = \eta_b(x, y). \tag{29.20}$$

This location can equivalently be specified mathematically by the surface

$$s(x, y, z) = \eta_b(x, y) - z = 0. \tag{29.21}$$

The outward normal at this surface is thus given by

$$\hat{\boldsymbol{n}} = \frac{\nabla s}{|\nabla s|} = \frac{\nabla \eta_b - \hat{\boldsymbol{z}}}{\sqrt{1 + \nabla \eta_b \cdot \nabla \eta_b}}. \tag{29.22}$$

If the bottom is impenetrable to flow, the velocity field is constrained to satisfy the no-normal flow boundary condition

$$\boldsymbol{v} \cdot \hat{\boldsymbol{n}} = 0 \qquad \text{at } z = \eta_b. \tag{29.23}$$

That is, fluid can move tangentially to the bottom, but not normal to the bottom. Making use of the bottom outward normal (29.22) leads to

$$w = \boldsymbol{u} \cdot \nabla \eta_b \qquad \text{at } z = \eta_b. \tag{29.24}$$

For a flat bottom, with $\nabla \eta_b = 0$, the no-normal flow condition means that $w = 0$ at the bottom interface. But more generally, sloping bottoms lead to a nonzero vertical velocity component whose value depends on the projection of the horizontal velocity onto the bottom slope.

The kinematic result (29.24) is written in an Eulerian sense, with the velocity constrained to satisfy this relation at each point along the bottom interface. It has a complementary material interpretation based on acknowledging that the bottom interface is a material surface. A parcel 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. We can ensure this constraint by setting

$$\frac{\mathrm{D}s}{\mathrm{D}t} = \frac{\mathrm{D}\left(z - \eta_b\right)}{\mathrm{D}t} = 0 \qquad \text{at } z = \eta_b. \tag{29.25}$$

Rearrangement of this result leads to the Eulerian constraint (29.24). Equivalently, we can write this boundary condition in the form

$$w = \frac{\mathrm{D}\eta_b}{\mathrm{D}t} \qquad \text{at } z = \eta_b. \tag{29.26}$$

Surface kinematic boundary condition

We again assume the surface boundary is a material surface. Consequently, the surface kinematic boundary condition follows analogously to the bottom. However, there is a fundamentally new feature in that the fluid free surface is a time dependent moving boundary. The free surface is located at a vertical position $z = \eta$. Equivalently, the free surface can be specified by a surface of constant s, where

$$s(x, y, z, t) = z - \eta(x, y, t) = 0. \tag{29.27}$$

The outward normal to the free surface is thus given by

$$\hat{\boldsymbol{n}} = \frac{\nabla s}{|\nabla s|} = \frac{\hat{\boldsymbol{z}} - \nabla\eta}{\sqrt{1 + \nabla\eta \cdot \nabla\eta}}. \tag{29.28}$$

We must account for motion of the surface when formulating the no-flow condition. To do so, we write this constraint as

$$(\boldsymbol{v} - \boldsymbol{v}^{(s)}) \cdot \hat{\boldsymbol{n}} = 0 \qquad \text{at } z = \eta(x, y, t), \tag{29.29}$$

where $\boldsymbol{v}^{(s)}$ is the velocity of a point on the ocean surface. The velocity of a point fixed on an arbitrary surface satisfies

$$\frac{\partial s}{\partial t} + \boldsymbol{v}^{(s)} \cdot \nabla s = 0. \tag{29.30}$$

As defined, $\boldsymbol{v}^{(s)}$ advects a fluid parcel in a manner to always keep the parcel fixed on the constant s surface. With $\hat{\boldsymbol{n}} = \nabla s/|\nabla s|$, we have

$$\boldsymbol{v}^{(s)} \cdot \hat{\boldsymbol{n}} = -\frac{\partial_t s}{|\nabla s|} = \frac{\partial_t \eta}{\sqrt{1 + \nabla\eta \cdot \nabla\eta}}. \tag{29.31}$$

Hence, if the surface remains static, then $\boldsymbol{v}^{(s)} \cdot \hat{\boldsymbol{n}} = 0$. But more generally, the surface is moving, and that movement is fundamental to the surface kinematic boundary condition.

Making use of the result (29.31) in the no-normal flow constraint (29.29) then leads to the surface kinematic boundary condition

$$w - \boldsymbol{u} \cdot \nabla\eta = \frac{\partial \eta}{\partial t} \qquad \text{at } z = \eta. \tag{29.32}$$

As for the bottom kinematic boundary condition written as (29.25), we can interpret this result materially, in which case

$$\frac{\mathrm{D}s}{\mathrm{D}t} = \frac{\mathrm{D}\,(z - \eta)}{\mathrm{D}t} = 0 \qquad \text{at } z = \eta. \tag{29.33}$$

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

$$w = \frac{\mathrm{D}\eta}{\mathrm{D}t} \qquad \text{at } z = \eta. \tag{29.34}$$

29.1.5 Stretching and vertical velocity

Since the fluid has constant density, we know that the velocity has zero divergence

$$\nabla \cdot \boldsymbol{v} = 0 \Rightarrow \frac{\partial w}{\partial z} = -\nabla \cdot \boldsymbol{u}. \tag{29.35}$$

This result also follows since material parcels in the constant density shallow water layer maintain a constant volume (see Section 16.1). 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

$$w(z) = w(\eta_b) - (z - \eta_b)\, \nabla \cdot \boldsymbol{u}, \tag{29.36}$$

so that the vertical velocity is a linear function of depth. Applying this equation at the ocean surface yields

$$w(\eta) = w(\eta_b) - (\eta - \eta_b)\, \nabla \cdot \boldsymbol{u}. \tag{29.37}$$

Eliminating the horizontal convergence between equations (29.36) and (29.37) leads to

$$w(z) - w(\eta_b) = \left(\frac{z - \eta_b}{\eta - \eta_b}\right) [w(\eta) - w(\eta_b)]. \tag{29.38}$$

Making use of the surface kinematic boundary condition (29.34) and bottom kinematic boundary condition (29.26) renders the material form

$$\frac{1}{z - \eta_b}\left(\frac{\mathrm{D}(z - \eta_b)}{\mathrm{D}t}\right) = \frac{1}{\eta - \eta_b}\left(\frac{\mathrm{D}(\eta - \eta_b)}{\mathrm{D}t}\right). \tag{29.39}$$

Finally, introducing the layer thickness $h = \eta - \eta_b$ yields the material conservation law

$$\frac{\mathrm{D}}{\mathrm{D}t}\left(\frac{z - \eta_b}{h}\right) = 0. \tag{29.40}$$

Again, $h = \eta - \eta_b$ is the layer thickness and $z - \eta_b$ is the height of a fluid parcel from the bottom inteface (see Figure 29.1). Consequently, equation (29.40) means that the ratio of the parcel height above the bottom to the layer thickness remains constant as the parcel moves through the shallow water fluid. That is, a column of shallow water fluid stretches or squeezes uniformly within a shallow water fluid. Shallow water dynamics thus comprises the dynamics of moving coherent fluid columns within a layer. This constrained behaviour results from the linear dependence with depth of the vertical velocity, which itself is a result of the depth independence of the horizontal velocity.

29.1.6 Further study

The shallow water model is ubiquitous in the geophysical fluid literature. An early application of the single shallow water layer was Laplace's studies of tides on a sphere (see exercise 29.4). Our presentation largely follows Chapter 3 of *Vallis* (2017) with another lucid discussion given by *Salmon* (1998).

29.2 Emphasizing the hydrostatic approximation

We revisit the above formulation to emphasize the hydrostatic approximation as the fundamental assumption leading to the shallow water equations. For this purpose, start from the compressible equations of motion for a perfect fluid in a rotating reference frame

$$\frac{\mathrm{D}\rho}{\mathrm{D}t} + \rho \nabla \cdot \boldsymbol{v} = 0 \tag{29.41a}$$

$$\rho\, \frac{\mathrm{D}\boldsymbol{v}}{\mathrm{D}t} + 2\, \boldsymbol{\Omega} \wedge \rho\, \boldsymbol{v} = -\hat{\boldsymbol{z}}\, g\, \rho - \nabla p. \tag{29.41b}$$

In a manner analogous to our discussion of the Boussinesq approximation in Section 25.2, decompose the density and pressure into a depth dependent term and a deviation

$$\rho(x,y,z,t) = \rho_r(z) + \rho'(x,y,z,t) \tag{29.42a}$$
$$p(x,y,z,t) = p_r(z) + p'(x,y,z,t) \tag{29.42b}$$

where the reference pressure is in hydrostatic balance with the reference density

$$\frac{\mathrm{d}p_r}{\mathrm{d}z} = -\rho_r\, g. \tag{29.43}$$

Inserting this decomposition into the equation of motion leads to

$$\rho\,\frac{\mathrm{D}\boldsymbol{v}}{\mathrm{D}t} + 2\,\boldsymbol{\Omega}\wedge\rho\,\boldsymbol{v} = -\hat{\boldsymbol{z}}\,g\,\rho' - \nabla p'. \tag{29.44}$$

29.2.1 Hydrostatic balance with respect to background density

Now introduce the following rather strict approximation: *The full fluid is in hydrostatic balance with respect to the background density, so that* $\rho' = 0$. As shown below, this assumption means that the density is a uniform constant and that the fluid is in turn incompressible.

29.2.2 Density is a uniform constant

With $\rho = \rho_r(z)$, the mass continuity equation implies

$$w\,\frac{\mathrm{d}\rho_r}{\mathrm{d}z} = 0. \tag{29.45}$$

For a nonzero vertical velocity, this constraint means that the density is itself a constant in space and time

$$\rho = \text{constant}. \tag{29.46}$$

The perturbation hydrostatic pressure is thus given by

$$p'(x,y,z,t) = g\,\rho\,(\eta - z), \tag{29.47}$$

where $\eta = \eta(x,y,t)$ is the free surface height. The horizontal momentum equation thus takes the form

$$\frac{\mathrm{D}\boldsymbol{u}}{\mathrm{D}t} + 2\,\boldsymbol{\Omega}\wedge\boldsymbol{v} = -g\,\nabla\eta. \tag{29.48}$$

An initial horizontal velocity that is depth independent will thus remain such.

29.2.3 Fluid is incompressible

With a uniform density, the mass continuity implies that the fluid is incompressible

$$\nabla\cdot\boldsymbol{v} = 0. \tag{29.49}$$

With $\boldsymbol{u}$ depth independent, we can depth integrate $\nabla\cdot\boldsymbol{v} = 0$ as well as the kinematic boundary conditions in Section 29.1.4 to render the thickness equation

$$\frac{1}{h}\,\frac{\mathrm{D}h}{\mathrm{D}t} = -\nabla\cdot\boldsymbol{u}. \tag{29.50}$$

29.2.4 Non-hydrostatic and hydrostatic forces in a homogeneous layer

In Section 45.1, we study surface gravity waves in a homogeneous (constant density) fluid layer. Such waves have an amplitude that exponentially decays with depth so that the horizontal and vertical fluid motion associated with the waves have a non-zero vertical shear. Now the hydrostatic pressure within a homogeneous layer has a depth-independent horizontal gradient

$$\partial_z \left(\nabla_z p_{\text{hydrostatic}} \right) = -g \, \nabla_z \rho = 0, \tag{29.51}$$

so the hydrostatic pressure gradients can only drive a horizontal flow that is depth independent. Hence, a vertically sheared horizontal flow within a homogeneous fluid must arise from a non-hydrostatic pressure force. As discussed in Section 45.1.3, surface gravity waves indeed involve non-hydrostatic pressure forces that drive the vertical dependence to the wave amplitude.

In contrast, as shown in Section 24.3.3, a non-homogeneous density field, in particular density with a horizontal gradient, leads to a vertically dependent hydrostatic pressure gradient. This hydrostatic pressure force can in turn drive vertically sheared flow, with thermal wind the canonical example in rotating fluids (Section 26.3.4).

29.2.5 Comments

The basic assumption leading to the shallow water model is the hydrostatic approximation for the full fluid state relative to the background density ρ_r. Mass continuity then implies ρ is a uniform constant, further implying the fluid is incompressible. The shallow water momentum and thickness equations then follow.

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 24.3). 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 25.2).

29.3 Shallow water fluid in a rotating tank

We introduced the geopotential in Section 9.12.2, which are surfaces where the effective gravitational force (sum of central gravity plus planetary centrifugal) is constant. Correspondingly, we introduced geopotential vertical coordinates in Section 9.13.3. We here revisit that discussion in relation to a shallow water fluid undergoing constant rotation about the vertical ($\boldsymbol{f} = 2\,\Omega\,\hat{\boldsymbol{z}}$). In this case, we see how the centrifugal acceleration (due to rotation of the tank) leads to a parabolic shape for the surface of a rotating tank fluid in solid-body rotation.

29.3.1 Equations of motion

The equation of motion for a fluid in a rotating tank is given by (see also Exercise 24.4)

$$\frac{\mathrm{D}\boldsymbol{u}}{\mathrm{D}t} + f\,\hat{\boldsymbol{z}} \wedge \boldsymbol{u} = -\nabla \left(p/\rho + g_{\mathrm{e}}\, z - \Omega^2\, r^2/2 \right). \tag{29.52}$$

Note the use of both a gravitational potential and centrifugal potential. As for planetary motion we know these two potentials can be combined into a geopotential. However, we here keep them separate since we are interested in details of the motion as seen from the laboratory. We make use

of polar coordinates as defined by

$$x = r\,\cos\theta \tag{29.53}$$

$$y = r\,\sin\theta, \tag{29.54}$$

with r the radial distance from the rotation axis and θ the angle made in a counter-clockwise direction from the positive x-axis.

29.3.2 Free surface shape in solid-body rotation

Consider a fluid at rest in the rotating frame, thus undergoing solid-body rotation. Static equlibrium in the rotating frame is realized when the forcing on the right hand side vanishes so that

$$p/\rho + g_{\mathrm{e}}\,z - \Omega^2\,r^2/2 = p_0/\rho, \tag{29.55}$$

where p_0 is a constant pressure to be specified below. At the free surface where $z = \eta = h$ (recall flat bottom), the pressure equals to that applied to the free surface by the overlying media, $p = p_{\mathrm{a}}$ (e.g., atmospheric pressure). The equilibrium layer thickness is thus given by

$$h = \frac{(\Omega\,r)^2}{2\,g_{\mathrm{e}}} + \left(\frac{p_0 - p_{\mathrm{a}}}{\rho\,g_{\mathrm{e}}}\right). \tag{29.56}$$

For simplicity, assume the applied pressure is spatially constant. Hence, we specify p_0 according to the thickness at $r = 0$, so that

$$h(r) = \frac{(\Omega\,r)^2}{2\,g_{\mathrm{e}}} + h(0). \tag{29.57}$$

This column thickness is therefore parabolic in shape, with increasing thickness moving away from the rotation axis.

29.3.3 Further study

The discussion in this section parallels that in Exercise 24.4. We consider the angular momentum for this system in Section 30.5. Refer to Section 6.6.4 of *Marshall and Plumb* (2008) for more discussion of rotating tank laboratory experiments.

29.4 Reduced gravity model for the upper ocean

The reduced gravity model describes an active layer of uniform density, ρ_1, above an 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.

29.4.1 Momentum and thickness equations for the active layer

We develop the momentum equations for the reduced gravity model by assuming hydrostatic balance, in which pressure at a depth, z, in the upper layer is computed as (see Figure 29.4)

$$p_1(x, y, z, t) = p_0(x, y, t) + g\,\rho_1\,(\eta_0 - z). \tag{29.58}$$

Figure 29.4: *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 and density ρ_1. The dynamically active layer is bounded above by a zero density layer. The lower inactive layer is assumed to be infinitely deep so that its continuity equation can be ignored; i.e., even though there is zero currents within this layer it can still undulate.*

Since the fluid above the upper layer has zero density, we set

$$p_0 = 0. \tag{29.59}$$

We immediately see that the horizontal momentum equation for the upper layer is written

$$\frac{\mathrm{D}\boldsymbol{u}}{\mathrm{D}t} + \boldsymbol{f} \wedge \boldsymbol{u} = -g\,\nabla\eta_0. \tag{29.60}$$

The equations for the upper layer are completed by mass (volume) conservation that yields the thickness equation

$$\frac{\mathrm{D}h}{\mathrm{D}t} = -h\,\nabla\cdot\boldsymbol{u}. \tag{29.61}$$

29.4.2 Relating undulations of the top and bottom layer interfaces

The pressure in the lower stagnent layer is given by the weight per horizontal area of fluid above it

$$p_2(x,y,z,t) = g\,\rho_1\,(\eta_0 - \eta_1) + g\,\rho_2\,(\eta_1 - z). \tag{29.62}$$

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

$$\nabla_z p_2 = g\,\rho_1\,\nabla(\eta_0 - \eta_1) + g\,\rho_2\,\nabla\eta_1 = 0. \tag{29.63}$$

This relation provides a constraint that links undulations of the top and bottom interfaces of the dynamically active layer

$$\eta_0 = -\eta_1\left[\frac{\rho_2 - \rho_1}{\rho_1}\right] + \text{constant}. \tag{29.64}$$

The density ratio on the right hand side is positive but typically much smaller than unity. Hence, the relation (29.64) means that undulations of the free surface, η_0, are of opposite sign and of much smaller amplitude than undulations in the lower interface, η_1. This behaviour is typical of undulations of the thermocline region of the ocean and the free surface (see Figure 29.5).

29.4.3 Momentum equation with reduced gravity

Relation (29.64) can be used to write the momentum equation (29.60) in the form

$$\frac{\mathrm{D}\boldsymbol{u}}{\mathrm{D}t} + \boldsymbol{f} \wedge \boldsymbol{u} = +g_1' \nabla \eta_1, \tag{29.65}$$

where

$$g_1' = g \left[\frac{\rho_2 - \rho_1}{\rho_1} \right] \ll g \tag{29.66}$$

defines the *reduced gravity*. It is more typical to make use of the momentum equation in the form (29.65), than the original form (29.60). The reason is that historically, ocean hydrography measurements[1] have allowed for an estimate of the pycnocline slope, $\nabla \eta_1$, whereas it was not until satellite altimetry measurements (post-1993) that we could estimate the sea level slope, $\nabla \eta_0$.

Figure 29.5: *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 η_0 in Figure 29.4, whereas the pycnocline (heavy black line) corresponds to the lower interface η_1 of Figure 29.4. The sea surface experiences an applied pressure $p = p_a$, assumed to be uniform for this idealized situation. Isolines of hydrostatic pressure are shown, with a slight upward bow to the isobars within the light water region, and flat isobars in the deeper region of zero motion. Note how sea level is a maximum above the pycnocline minimum, with this geometry reflected in equation (29.64). 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 m. Note that the essential hydrostatic features of this example are contained in Figure 24.1 when considering the pressure difference between two columns of fluid with unequal density.*

29.4.4 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 1.5-layer reduced gravity model for interpretating aspects of the

[1] In oceanography, hydrography refers to measurements of temperature, salinity, and pressure; see *Talley et al.* (2011).

observed ocean. Additional use is made by *Griffies et al.* (2014) for interpreting patterns of sea level in the ocean. Figure 29.5 is based on Figure 3.3 from *Tomczak and Godfrey* (1994) as well as Figure 37 from *Griffies et al.* (2014).

29.5 Stacked shallow water equations

Figure 29.6: *Two dynamically active layers of stacked shallow water fluid. The notation corresponds to that for the reduced gravity model of Figure 29.4, with two dynamically active layers shown in the present figure. The "atmosphere" above the layers is assumed to have zero mass, and the total thickness of the column is* $h_1 + h_2 = \eta_0 - \eta_b$.

We here consider two dynamically active shallow water layers as shown in Figure 29.6. This model offers the canonical tool for theoretical studies of baroclinic behaviour. The equations for more than two layers follow by induction from the two-layer case.

29.5.1 Model formulation

Each shallow water layer satisfies its own independent thickness equation, representing the conservation of mass for each layer

$$\frac{\partial h_1}{\partial t} + \nabla \cdot (h_1 \, \boldsymbol{u}_1) = 0 \tag{29.67}$$

$$\frac{\partial h_2}{\partial t} + \nabla \cdot (h_2 \, \boldsymbol{u}_2) = 0. \tag{29.68}$$

We emphasize that there is no coupling between these equations, as each layer separately must satisfy volume conservation.

We next need the pressure in each layer to formulate the pressure forces for driving flow. As before, make use of the hydrostatic balance and integrate down from the surface (assuming zero

mass layer above), which results in the pressure fields

$$p_1 = \rho_1 \, g \, (\eta_0 - z) \tag{29.69}$$
$$p_2 = \rho_1 \, g \, (\eta_0 - \eta_1) + \rho_2 \, g \, (\eta_1 - z). \tag{29.70}$$

It is convenient to write pressure in layer-two using the reduced gravity, which leads to

$$p_2 = \rho_1 \, g \, (\eta_0 - \eta_1) + \rho_2 \, g \, (\eta_1 - z) \tag{29.71a}$$
$$= g \, \eta_1 \, (\rho_2 - \rho_1) + g \, \rho_1 \, \eta_0 - g \, \rho_2 \, z \tag{29.71b}$$
$$= \rho_1 \left[g \, \eta_0 + g \, \eta_1 \, \frac{\rho_2 - \rho_1}{\rho_1} \right] - g \, \rho_2 \, z \tag{29.71c}$$
$$= \rho_1 \, (g \, \eta_0 + g_1' \, \eta_1) - g \, \rho_2 \, z, \tag{29.71d}$$

where we introduced the reduced gravity between layers one and two

$$g_1' = g \left[\frac{\rho_2 - \rho_1}{\rho_1} \right]. \tag{29.72}$$

The terms $g \, \rho_i \, z$ appearing in the layer pressures p_i have zero horizontal gradient. Hence, they play no dynamical role in determining the horizontal velocity.

The horizontal momentum equations for the two layers take the general form

$$\rho_1 \left[\frac{\mathrm{D}_1 \boldsymbol{u}_1}{\mathrm{D}t} + \boldsymbol{f} \wedge \boldsymbol{u}_1 \right] = -\nabla p_1 \tag{29.73a}$$
$$\rho_2 \left[\frac{\mathrm{D}_2 \boldsymbol{u}_2}{\mathrm{D}t} + \boldsymbol{f} \wedge \boldsymbol{u}_2 \right] = -\nabla p_2, \tag{29.73b}$$

where we introduced the layer material time derivatives

$$\frac{\mathrm{D}_n}{\mathrm{D}t} = \frac{\partial}{\partial t} + \boldsymbol{u}_n \cdot \nabla. \tag{29.74}$$

Making use of expressions (29.69) and (29.71d) for layer pressures leads to

$$\frac{\mathrm{D}_1 \boldsymbol{u}_1}{\mathrm{D}t} + \boldsymbol{f} \wedge \boldsymbol{u}_1 = -g \, \nabla \eta_0 \tag{29.75}$$
$$\frac{\mathrm{D}_2 \boldsymbol{u}_2}{\mathrm{D}t} + \boldsymbol{f} \wedge \boldsymbol{u}_2 = -\frac{\rho_1}{\rho_2} \left(g \, \nabla \eta_0 + g_1' \, \nabla \eta_1 \right). \tag{29.76}$$

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 (29.67) and (29.68). We thus write

$$\eta_0 = \eta_b + h_1 + h_2 \tag{29.77a}$$
$$\eta_1 = \eta_b + h_2, \tag{29.77b}$$

so that

$$p_1 = \rho_1 \, g \, (\eta_b + h_1 + h_2) \tag{29.78a}$$
$$p_2 = \rho_1 \, \left[g \, (\eta_b + h_1 + h_2) + g_1' \, (\eta_b + h_2) \right], \tag{29.78b}$$

thus resulting in the momentum equations

$$\frac{\mathrm{D}_1 \boldsymbol{u}_1}{\mathrm{D}t} + \boldsymbol{f} \wedge \boldsymbol{u}_1 = -g\,\nabla(\eta_b + h_1 + h_2) \tag{29.79a}$$

$$\frac{\mathrm{D}_2 \boldsymbol{u}_2}{\mathrm{D}t} + \boldsymbol{f} \wedge \boldsymbol{u}_2 = -\frac{\rho_1}{\rho_2}\,\nabla\left[g\,(\eta_b + h_1 + h_2) + g_1'\,(\eta_b + h_2)\right]. \tag{29.79b}$$

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

$$\frac{\mathrm{D}_1 \boldsymbol{u}_1}{\mathrm{D}t} + \boldsymbol{f} \wedge \boldsymbol{u}_1 = -g\,\nabla(\eta_b + h_1 + h_2) \tag{29.80a}$$

$$\frac{\mathrm{D}_2 \boldsymbol{u}_2}{\mathrm{D}t} + \boldsymbol{f} \wedge \boldsymbol{u}_2 = -\nabla\left[g\,(\eta_b + h_1 + h_2) + g_1'\,(\eta_b + h_2)\right]. \tag{29.80b}$$

Notice how the difference in layer velocities, $\boldsymbol{u}_1 - \boldsymbol{u}_2$, is affected by a pressure gradient arising just from bottom topography and the interior layer thickness, h_2

$$\frac{\mathrm{D}_1 \boldsymbol{u}_1}{\mathrm{D}t} - \frac{\mathrm{D}_2 \boldsymbol{u}_2}{\mathrm{D}t} + \boldsymbol{f} \wedge (\boldsymbol{u}_1 - \boldsymbol{u}_2) = g_1'\nabla\,(\eta_b + h_2). \tag{29.81}$$

This vertical "shear" does not directly feel undulations of the free surface. Rather, it feels these surface undulations only indirectly via nonlinear terms appearing in the advection terms on the left hand side. We further discuss this result in Section 30.1.2 by introducing thermal wind and the Margules' Relation.

29.5.2 Further study

We generalize the stacked shallow water equations in Section 46.1 by developing the equations for a Boussinesq isopycnal ocean model.

29.6 Exercises

EXERCISE 29.1: RELATIONS FOR VERTICAL VELOCITY (EXERCISE (3.2) OF *Vallis* (2006))
Show that the vertical velocity within a shallow water system is given by

$$w = \left(\frac{z - \eta_b}{h}\right)\frac{\mathrm{D}h}{\mathrm{D}t} + \frac{\mathrm{D}\eta_b}{\mathrm{D}t}. \tag{29.82}$$

Interpret the result, showing that it gives sensible answers at the top and bottom of the fluid layer.

EXERCISE 29.2: SHALLOW WATER LAYER WITH SURFACE VOLUME SOURCES
In Section 29.1 we assumed a zero volume crossing the boundary of the shallow water fluid. Consequently, both the surface and bottom boundaries are material surfaces. For this exercise we introduce a surface volume source as occurs across the ocean surface through evaporation, precipitation, and river runoff. This surface volume transfer in turn means the surface boundary is no longer a material surface. For this problem, let $\mathbb{p}$ be the volume per time per horizontal area of fluid entering across the surface of the shallow water layer ($\mathbb{p}$ has dimensions of length per time and $\mathbb{p} > 0$ means volume enters the shallow water layer). Assume the water in $\mathbb{p}$ has the same density and same velocity as the shallow water layer. Hence, there is no modification to the layer stratification (i.e., it remains a homogeneous layer).

(a) Equations (29.33) and (29.34) offer equivalent expressions for the surface kinematic boundary conditions in the absence of a volume transfer across the surface. How are these expressions modified in the presence of $\mathbb{p} \neq 0$?

(b) What is the layer thickness equation in the presence of $\mathbb{p} \neq 0$?

(c) Equation (29.40) shows that in the absence of volume sources, a column of shallow water fluid stretches or squeezes uniformly. How is this relation modified in the presence of $\mathbb{p} \neq 0$?

(d) Equation (33.25) shows that the potential vorticity $Q = (\zeta + f)/h$ is materially conserved for an inviscid shallow water fluid layer in the absence of volume sources, $\mathrm{D}Q/\mathrm{D}t = 0$. How is this material conservation equation modified in the presence of $\mathbb{p} \neq 0$? Answering this question requires knowledge of the shallow water potential vorticity derivation given in Section 33.3.

EXERCISE 29.3: 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

$$\partial_x h \left[1 - \mathrm{Fr}^2\right] = \partial_x \eta_b \tag{29.83}$$

where the Froude number is given by

$$\mathrm{Fr} = \frac{u}{\sqrt{g\,h}}. \tag{29.84}$$

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 29.4: SHALLOW WATER EQUATIONS WITH TIDES

In Chapter 43 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 43.4, assume the perturbation geopotential is depth independent.

EXERCISE 29.5: 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 $\rho_1/\rho_2 \to 0$ the single layer shallow water equations emerge. Make use of notation from the two-layer system shown in Figure 29.6. This model might be used to study flow in the atmosphere well above the boundary layer, or the abyssal ocean well below the pycnocline.

30

Shallow water dynamics

We focus in this chapter on developing dynamical features for the shallow water system. This study includes geostrophy, thermal wind (in the form of Margules' relation), momentum budgets, form stress, mechanical energy, and available potential energy. As we will see, the shallow water fluid provides a particularly relevant and analytically tractable means to study aspects of geophysical fluid motion. Although a highly idealized system, many of the intuitions developed by studing the shallow water fluid carry over to realistic geophysical fluids.

READER'S GUIDE TO THIS CHAPTER

This chapter builds from the shallow water formulations in Chapter 29; 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.

30.1 Geostrophic balance and thermal wind **526**
30.1.1 Geostrophy for a single shallow water layer 526
30.1.2 Geostrophy and thermal wind for two shallow water layers 527
30.1.3 Geostrophic transport within layers 528
30.1.4 Comments 529
30.2 Contact pressure forces and form stress **529**
30.2.1 Contact pressure force along vertical sides 530
30.2.2 Contact pressure force along the top and bottom interfaces 531
30.2.3 Form stress 532
30.2.4 Net integrated contact pressure force on the column 532
30.2.5 Contact pressure force on a single layer 533
30.3 Thickness weighted momentum equation **533**
30.3.1 Basic manipulations 533
30.3.2 Geostrophic and ageostrophic contributions 534
30.3.3 Exposing the contact force version of pressure 534
30.4 Force balances in a zonally re-entrant channel **535**
30.4.1 Volume transport for steady flow 535
30.4.2 Steady meridional balances 536
30.4.3 Steady zonal balance 537
30.4.4 A minor role for frictional bottom stresses 537
30.4.5 Correlation between surface height and topography slope 538
30.4.6 Connection to meridional geostrophic transport 539
30.4.7 Sinusoidal example 539

30.4.8 Comments 541
30.5 Angular momentum in a rotating cylindrical tank **542**
30.5.1 Angular momentum for a column of shallow water fluid 542
30.5.2 Material time evolution of the angular momentum 543
30.5.3 Materially invariant angular momentum 544
30.5.4 Comments 544
30.6 Mechanical energy budget for a shallow water layer **544**
30.6.1 Gravitational potential energy 544
30.6.2 Kinetic energy 545
30.6.3 Mechanical energy 545
30.7 Available potential energy of a shallow water layer **546**
30.8 Exercises **547**

30.1 Geostrophic balance and thermal wind

As described in Chapter 26, geostrophic balance arises from dropping the material time derivative in the inviscid horizontal momentum equation. 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.

30.1.1 Geostrophy for a single shallow water layer

The geostrophic balance for a single shallow water fluid layer takes the form

$$f\,\hat{\boldsymbol{z}} \wedge \boldsymbol{u}_{\mathrm{g}} = -g\,\nabla\eta \Longrightarrow f\,\boldsymbol{u}_{\mathrm{g}} = g\,\hat{\boldsymbol{z}} \wedge \nabla\eta, \tag{30.1}$$

or in component form

$$u_{\mathrm{g}} = -\frac{g}{f}\frac{\partial\eta}{\partial y} \qquad \text{and} \qquad v_{\mathrm{g}} = \frac{g}{f}\frac{\partial\eta}{\partial x}. \tag{30.2}$$

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 30.1 shows a schematic of the geostrophic balance for a single shallow water layer.

Figure 30.1: *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.*

30.1.2 Geostrophy and thermal wind for two shallow water layers

Now consider two shallow water layers as in Figure 29.6. Recall the layer pressure equations (29.69) and (29.70), which leads to the pressure difference

$$p_1 - p_2 = g\,\eta_1\,(\rho_1 - \rho_2) + g\,z\,(\rho_2 - \rho_1) \tag{30.3a}$$
$$= -g_1'\,\rho_1\,\eta_1 + g_1'\,\rho_1\,z, \tag{30.3b}$$
$$= g_1'\,\rho_1\,(-\eta_1 + z), \tag{30.3c}$$

where the reduced gravity is given by

$$g_1' = g\left[\frac{\rho_2 - \rho_1}{\rho_1}\right] \ll g. \tag{30.4}$$

The density difference $\rho_2 - \rho_1$ is generally much smaller than either density, so that $g_1' \ll g$. For a Boussinesq shallow water system, the momentum equations are given by

$$\frac{\mathrm{D}_1\boldsymbol{u}_1}{\mathrm{D}t} + \boldsymbol{f} \wedge \boldsymbol{u}_1 = -\rho_1^{-1}\,\nabla p_1 \tag{30.5a}$$
$$\frac{\mathrm{D}_2\boldsymbol{u}_2}{\mathrm{D}t} + \boldsymbol{f} \wedge \boldsymbol{u}_2 = -\rho_1^{-1}\,\nabla p_2, \tag{30.5b}$$

where we used the top layer density as the Boussinesq reference density, and where we introduced the material time derivatives for each layer

$$\frac{\mathrm{D}_n}{\mathrm{D}t} = \frac{\partial}{\partial t} + \boldsymbol{u}_n \cdot \nabla. \tag{30.6}$$

Making use of the pressure difference (30.3c) renders

$$\frac{\mathrm{D}_1\boldsymbol{u}_1}{\mathrm{D}t} - \frac{\mathrm{D}_2\boldsymbol{u}_2}{\mathrm{D}t} + \boldsymbol{f} \wedge \Delta\boldsymbol{u} = -\rho_1^{-1}\,\nabla(p_1 - p_2) \tag{30.7a}$$
$$= g_1'\,\nabla\eta_1, \tag{30.7b}$$

where

$$\Delta\boldsymbol{u} = \boldsymbol{u}_1 - \boldsymbol{u}_2 \tag{30.8}$$

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

$$\boldsymbol{f} \wedge \Delta\boldsymbol{u}_{\mathrm{g}} = g_1'\,\nabla\eta_1, \tag{30.9}$$

which is equivalent to

$$f\,\Delta\boldsymbol{u}_{\mathrm{g}} = -\hat{\boldsymbol{z}} \wedge g_1'\nabla\eta_1 \Longrightarrow \Delta u_{\mathrm{g}} = +\frac{g_1'}{f}\,\frac{\partial\eta_1}{\partial y} \qquad \text{and} \qquad \Delta v_{\mathrm{g}} = -\frac{g_1'}{f}\,\frac{\partial\eta_1}{\partial x}. \tag{30.10}$$

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. We illustrate this relation in Figure 30.2. The Margules relation is a discrete (two-layer) version of the thermal wind relation discussed in Section 26.4.5.

Figure 30.2: *Illustrating Margule's relation for the northern hemisphere ($f > 0$). Here we show the interface between a cold/heavy layer to the right and a warm/light layer to the left. The slope of the interface is positive, $\partial \eta_1 / \partial y > 0$, 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.*

30.1.3 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 finite region, such as an ocean basin. For an N-layer shallow water fluid this transport is written

$$\boldsymbol{U} = \int \boldsymbol{u} \, \mathrm{d}z = \sum_{k=1}^{N} \boldsymbol{u}_k \, h_k, \tag{30.11}$$

where $\boldsymbol{u}_k$ 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

$$\boldsymbol{u}_k = (\rho_1 \, f)^{-1} \, \hat{\boldsymbol{z}} \wedge \nabla p_k, \tag{30.12}$$

where ρ_1 is the layer-1 density, which is also used as the reference density for the Boussinesq approximation.

As an exercise, let us work through the geostrophic transport for three layers, in which the layer pressures are written

$$p_1 = \rho_1 \, g \, (\eta_0 - z) \tag{30.13a}$$

$$p_2 = \rho_1 \, g \, (\eta_0 - \eta_1) + \rho_2 \, g \, (\eta_1 - z) = \rho_1 \, (g \, \eta_0 + g_1' \, \eta_1) - \rho_2 \, g \, z \tag{30.13b}$$

$$p_2 = \rho_1 \, g \, (\eta_0 - \eta_1) + \rho_2 \, g \, (\eta_1 - \eta_2) + \rho_3 \, g \, (\eta_2 - z) = \rho_1 \, (g \, \eta_0 + g_1' \, \eta_1 + g_2' \, \eta_2) - \rho_3 \, g \, z, \tag{30.13c}$$

where the reduced gravities are

$$g_1' = \frac{\rho_2 - \rho_1}{\rho_1} \qquad \text{and} \qquad g_2' = \frac{\rho_3 - \rho_2}{\rho_1}. \tag{30.14}$$

We are thus able to express the geostrophic transport in terms of the thickness weighted slopes of the layer interfaces

$$h_1 \, \boldsymbol{u}_1 = \frac{h_1}{f} \, \hat{\boldsymbol{z}} \wedge \nabla (g \, \eta_0) \tag{30.15a}$$

$$h_2 \, \boldsymbol{u}_2 = \frac{h_2}{f} \, \hat{\boldsymbol{z}} \wedge (g \, \nabla \eta_0 + g_1' \, \nabla \eta_1) \tag{30.15b}$$

$$h_3 \, \boldsymbol{u}_3 = \frac{h_3}{f} \, \hat{\boldsymbol{z}} \wedge (g \, \nabla \eta_0 + g_1' \, \nabla \eta_1 + g_2' \, \nabla \eta_2), \tag{30.15c}$$

and to write the depth integrated geostrophic transport as

$$\boldsymbol{U} = \frac{1}{f}\left[g\,(h_1 + h_2 + h_3)\,\nabla\eta_0 + g_1'\,(h_1 + h_2)\,\nabla\eta_1 + g_2'\,h_3\,\nabla\eta_2\right], \tag{30.16}$$

where $h_1 + h_2 + h_3 = \eta_0 - \eta_{\mathrm{b}}$ is the total depth of the column.

30.1.4 Comments

An alternative definition of reduced gravity for the 2-layer system uses the average density,

$$\overline{\rho} = (\rho_1 + \rho_2)/2 \tag{30.17}$$

for the denominator, in which case the modified reduced gravity is

$$\overline{g}_1' = g\left[\frac{\rho_2 - \rho_1}{\overline{\rho}}\right]. \tag{30.18}$$

Correspondingly, the pressure gradient on the right hand side of the momentum equations (30.5a) and (30.5b) have $1/\rho_1$ replaced by $1/\overline{\rho}$. Even so, the equation for the vertical shear evolution, (30.7b), 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 30.1.3.

30.2 Contact pressure forces and form stress

Thus far in our discussion of shallow water models, we have considered the pressure force in the mathematical form of a pressure gradient acting within the fluid volume. In this way, the pressure force appears as a body force just like gravity, centrifugal, and Coriolis. Alternatively, we may consider pressure to be a contact force per area acting at the interface between arbitrary fluid regions. If the contact pressure force integrates to a nonzero value over the region boundaries, then pressure accelerates the region.

As already discussed in Sections 19.1.3 and 38.2, the connection between the two forms for the pressure force arise through an application of Gauss's divergence theorem to scalar fields (see Section 2.7.2)

$$\boldsymbol{F}_{\mathcal{R}}^{\mathrm{press}} = -\int_{\mathcal{R}} \nabla p \,\mathrm{d}V = -\oint_{\partial\mathcal{R}} p\,\hat{\boldsymbol{n}}\,\mathrm{d}\mathcal{S}. \tag{30.19}$$

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 $\hat{\boldsymbol{n}}$. 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.

We discussed this dual representation of the pressure force per area in Section 18.1. Here, we pursue the contact force perspective as a means to understand the *form stress* acting between layers of shallow water fluid. There is also a form stress acting between a fluid layer and the solid earth, as well as between a fluid layer and the overlying atmosphere (when that atmosphere has a non-zero mass). Much of this material was discussed more generally in Chapter 20. It is such an important topic for understanding how forces act in fluids that we find it useful to revisit the subject here in the context of a shallow water fluid.

Figure 30.3: *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. 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 $\eta_{k\pm1/2}$ 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. This figure shares much in common with Figure 20.6.*

30.2.1 Contact pressure force along vertical sides

The pressure at a vertical position within a shallow water layer is given by (see Figure 30.3)

$$p(z) = \rho_k\, g\, (\eta_{k-1/2} - z) + p_{k-1/2}. \tag{30.20}$$

Integrating this pressure over the layer thickness yields

$$\int_{\eta_{k+1/2}}^{\eta_{k-1/2}} p(z)\, \mathrm{d}z = \rho_k\, g \left[\eta_{k-1/2}\,(\eta_{k-1/2} - \eta_{k+1/2}) - (1/2)\,(\eta_{k-1/2}^2 - \eta_{k+1/2}^2)\right] + p_{k-1/2}\, h_k \tag{30.21a}$$

$$= \rho_k\, g\, h_k^2/2 + h_k\, p_{k-1/2}. \tag{30.21b}$$

As seen in Section 30.6.1, $\rho_k\, g\, h_k^2/2$ is related to the gravitational potential energy for the column. Now assume the column to be infinitesimally thin so that it can approximated as rectangular. The zonal pressure force, in the limit that the column becomes thin, takes the form

$$\mathrm{d}y \int_{\eta_{k+1/2}}^{\eta_{k-1/2}} (p_L - p_R)\, \mathrm{d}z = -\mathrm{d}x\, \mathrm{d}y \left[(1/2)\, \rho_k\, g\, \frac{\partial h_k^2}{\partial x} + \frac{\partial (h_k\, p_{k-1/2})}{\partial x}\right] \tag{30.22a}$$

$$= -\mathrm{d}x\, \mathrm{d}y\, \frac{\partial}{\partial x} \left[(1/2)\, \rho_k\, g\, h_k^2 + h_k\, p_{k-1/2}\right], \tag{30.22b}$$

where $\mathrm{d}x\,\mathrm{d}y$ is the horizontal cross-sectional area of the column. A similar result holds for the meridional direction, thus rendering the net contact pressure force acting on the vertical sides

$$\boldsymbol{F}^{\text{press}}_{\text{sides}} = -\mathrm{d}x\,\mathrm{d}y\,\nabla\left[(1/2)\,\rho_k\,g\,h_k^2 + h_k\,p_{k-1/2}\right]. \tag{30.23}$$

That is, the contact force on the sides reduces, as the column becomes thin, to a gradient force. This exercise reveals no more than the integral theorem (30.19). Nonetheless, it is revealing to see how the integral theorem manifests within the shallow water system.

30.2.2 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

$$\hat{\boldsymbol{n}}_{k-1/2} = \frac{\nabla\left(z - \eta_{k-1/2}\right)}{\left|\nabla\left(z - \eta_{k-1/2}\right)\right|} = \frac{\hat{\boldsymbol{z}} - \nabla\eta_{k-1/2}}{\sqrt{1 + (\nabla\eta_{k-1/2})^2}}. \tag{30.24}$$

For mathematical simplicity, temporarily assume the interface slope to have a zero projection in the $\hat{\boldsymbol{y}}$ direction. In this case, the outward normal is

$$\hat{\boldsymbol{n}}_{k-1/2} = \frac{\hat{\boldsymbol{z}} - \hat{\boldsymbol{x}}\,\partial_x\eta_{k-1/2}}{\sqrt{1 + (\partial_x\eta_{k-1/2})^2}} \tag{30.25a}$$

$$= \frac{\hat{\boldsymbol{z}} - \hat{\boldsymbol{x}}\,\tan\varphi_{k-1/2}}{\sqrt{1 + \tan^2\varphi_{k-1/2}}} \tag{30.25b}$$

$$= (\hat{\boldsymbol{z}} - \hat{\boldsymbol{x}}\,\tan\varphi_{k-1/2})\,\cos\varphi_{k-1/2}, \tag{30.25c}$$

where we defined the interface slope as

$$\frac{\partial\eta_{k-1/2}}{\partial x} = \tan\varphi_{k-1/2}, \tag{30.26}$$

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 column[1]

$$\mathrm{d}\mathcal{S}_{k-1/2} = \frac{\mathrm{d}x\,\mathrm{d}y}{\cos\varphi_{k-1/2}}. \tag{30.27}$$

Hence, the product of the area and the outward normal is given by

$$\hat{\boldsymbol{n}}_{k-1/2}\,\mathrm{d}\mathcal{S}_{k-1/2} = \mathrm{d}x\,\mathrm{d}y\,(\hat{\boldsymbol{z}} - \hat{\boldsymbol{x}}\,\partial_x\eta_{k-1/2}). \tag{30.28}$$

[1]Equation (30.27) was also found in Section 14.4.3 when developing the kinematic boundary condition for a material interface.

This result generalizes to an interface slope that projects into both horizontal directions

$$\hat{\boldsymbol{n}}_{k-1/2}\,\mathrm{d}\mathcal{S}_{k-1/2} = \mathrm{d}x\,\mathrm{d}y\,(\hat{\boldsymbol{z}} - \nabla\eta_{k-1/2}). \tag{30.29}$$

Hence, the contact pressure force acting on layer-k at the top of the column is given by

$$\boldsymbol{F}^{\text{press}}_{\text{top}} = -\mathrm{d}x\,\mathrm{d}y\,(\hat{\boldsymbol{z}} - \nabla\eta_{k-1/2})\,p_{k-1/2}. \tag{30.30}$$

Analogous considerations lead to the contact pressure force acting on layer-k at the bottom of the column

$$\boldsymbol{F}^{\text{press}}_{\text{bot}} = \mathrm{d}x\,\mathrm{d}y\,(\hat{\boldsymbol{z}} - \nabla\eta_{k+1/2})\,p_{k+1/2}. \tag{30.31}$$

30.2.3 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

$$\boldsymbol{F}^{\text{form stress}}_{\text{top}} = \mathrm{d}x\,\mathrm{d}y\,(p_{k-1}\nabla\eta_{k-1}) \tag{30.32}$$

$$\boldsymbol{F}^{\text{form stress}}_{\text{bott}} = -\mathrm{d}x\,\mathrm{d}y\,(p_{k}\nabla\eta_{k}). \tag{30.33}$$

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 $\hat{\boldsymbol{x}}$ direction (e.g., see Figure 30.3). Form stress acting at the interface provides a $+\hat{\boldsymbol{x}}$ directed acceleration on the column. For the $k+1/2$ interface, a negatively sloped interface also experiences a $+\hat{\boldsymbol{x}}$ directed acceleration.

30.2.4 Net integrated contact pressure force on the column

Summing the contact pressure forces (30.23), (30.30), and (30.31) leads to the net pressure force acting on the column

$$\boldsymbol{F}^{\text{press}}_{\text{net}} = -\mathrm{d}x\,\mathrm{d}y\,\left[\nabla\left((1/2)\,\rho_k\,g\,h_k^2 + h_k\,p_{k-1/2}\right) + (\hat{\boldsymbol{z}} - \nabla\eta_{k-1/2})\,p_{k-1/2} - (\hat{\boldsymbol{z}} - \nabla\eta_{k+1/2})\,p_{k+1/2}\right] \tag{30.34a}$$

$$= -M_k\,g\,\hat{\boldsymbol{z}} - \mathrm{d}x\,\mathrm{d}y\,\left[\nabla\left[(1/2)\,\rho_k\,g\,h_k^2 + h_k\,p_{k-1/2}\right] - p_{k-1/2}\nabla\eta_{k-1/2} + p_{k+1/2}\nabla\eta_{k+1/2}\right]. \tag{30.34b}$$

To reach this result, we used the hydrostatic relation for the vertical force

$$p_{k-1/2} - p_{k+1/2} = -\rho_k\,g\,h_k, \tag{30.35}$$

and introduced the column mass

$$M_k = \rho_k\,h_k\,\mathrm{d}x\,\mathrm{d}y. \tag{30.36}$$

The vertical component of the net contact pressure force equals to the weight of the column, which is expected since the fluid is assumed to be in hydrostatic balance. The horizontal contact pressure force arises from a total horizontal gradient plus the form stress at the surface and bottom interfaces. In the ocean, the gradient term is removed when integrating horizontally over the full domain, assuming the thickness of the layer vanishes upon reaching the coastlines. 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.

30.2.5 Contact pressure force on a single layer

As a check on our calculation of the contact pressure force (30.34b), consider a single shallow water layer under a massless atmosphere. In this case, the contact pressure force per mass is given by

$$\frac{\boldsymbol{F}^{\text{press}}_{\text{net}}}{M} = -g\,\hat{\boldsymbol{z}} - \frac{1}{\rho\,h}\left[(1/2)\,\nabla(\rho\,g\,h^2) + p_b\nabla\eta_b\right] \tag{30.37a}$$

$$= -g\,\hat{\boldsymbol{z}} - \frac{1}{\rho\,h}\left[\rho\,g\,h\,\nabla h + \rho\,g\,h\nabla\eta_b\right] \tag{30.37b}$$

$$= -g\,\hat{\boldsymbol{z}} - g\nabla\eta. \tag{30.37c}$$

The final equality made use of the identity

$$\eta = h + \eta_b \tag{30.38}$$

as shown in Figure 29.1. As expected, the horizontal component of this force equals to the pressure gradient body force per mass detailed in Section 29.1.1.

30.3 Thickness weighted momentum equation

In this section we derive momentum equation for the depth integrated velocity for a single shallow water layer, thus allowing for a measure of the momentum budget in a column of shallow water fluid extending from the surface to the bottom of the layer (see Figure 30.4)

$$\int_{\eta_b}^{\eta} \boldsymbol{u}\,\mathrm{d}z = h\,\boldsymbol{u}. \tag{30.39}$$

The resulting *thickness weighted momentum equation* is written in an Eulerian flux form. Doing so facilitates analysis of forces acting on the layer such as for the zonal channel considered in Section 30.4.

Figure 30.4: *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).*

30.3.1 Basic manipulations

Recall the velocity and thickness equations written using the material time operator

$$\frac{\mathrm{D}\boldsymbol{u}}{\mathrm{D}t} + \boldsymbol{f}\wedge\boldsymbol{u} = -g\nabla\eta \qquad \text{and} \qquad \frac{\mathrm{D}h}{\mathrm{D}t} = -h\,\nabla\cdot\boldsymbol{u}. \tag{30.40}$$

Combining these two equations allows us to write the thickness weighted material acceleration as

$$h\,\frac{\mathrm{D}\boldsymbol{u}}{\mathrm{D}t} = h\,\frac{\mathrm{D}\boldsymbol{u}}{\mathrm{D}t} + \boldsymbol{u}\left[\frac{\mathrm{D}h}{\mathrm{D}t} + h\,\nabla\cdot\boldsymbol{u}\right] = \frac{\partial(h\,\boldsymbol{u})}{\partial t} + \nabla\cdot[\boldsymbol{u}\,(h\,\boldsymbol{u})], \tag{30.41}$$

so that the thickness weighted equation takes the vector form

$$\frac{\partial(h\,\boldsymbol{u})}{\partial t} + \nabla\cdot[\boldsymbol{u}\,(h\,\boldsymbol{u})] + \boldsymbol{f}\wedge(h\,\boldsymbol{u}) = -g\,h\,\nabla\eta, \tag{30.42}$$

and the component form

$$\frac{\partial(h\,u)}{\partial t} + \partial_x(h\,u^2) + \partial_y(h\,u\,v) - v\,h\,f = -\,g\,h\,\partial_x\eta \tag{30.43a}$$

$$\frac{\partial(h\,v)}{\partial t} + \partial_x(h\,u\,v) + \partial_y(h\,v^2) + u\,h\,f = -\,g\,h\,\partial_y\eta. \tag{30.43b}$$

These forms of the equations serve many uses. Additionally, we offer two slight modifications in the following that add some insight into the formulations.

30.3.2 Geostrophic and ageostrophic contributions

Bringing the Coriolis terms to the right hand side of equations (30.43a) and (30.43b) renders

$$\frac{\partial(h\,u)}{\partial t} + \partial_x(h\,u^2) + \partial_y(h\,u\,v) = h\,(-g\,\partial_x\eta + v\,f) \tag{30.44a}$$

$$\frac{\partial(h\,v)}{\partial t} + \partial_x(h\,u\,v) + \partial_y(h\,v^2) + u\,h\,f = h\,(-g\,\partial_y\eta - u\,f). \tag{30.44b}$$

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

$$f\,u = f\,(u_{\mathrm{a}} + u_{\mathrm{g}}) = f\,u_{\mathrm{a}} - g\,\partial\eta/\partial y \tag{30.45a}$$

$$f\,v = f\,(v_{\mathrm{a}} + v_{\mathrm{g}}) = f\,v_{\mathrm{a}} + g\,\partial\eta/\partial x, \tag{30.45b}$$

in which case equations (30.44a) and (30.44b) become

$$\frac{\partial(h\,u)}{\partial t} + \partial_x(h\,u^2) + \partial_y(h\,u\,v) = h\,f\,v_{\mathrm{a}} \tag{30.46a}$$

$$\frac{\partial(h\,v)}{\partial t} + \partial_x(h\,u\,v) + \partial_y(h\,v^2) = -h\,f\,u_{\mathrm{a}}. \tag{30.46b}$$

One should be careful *not* to take $f = 0$ limit of these equations since there is the chance of spuriously concluding there is no pressure contribution. Instead, equations (30.44a) and (30.44b) should be the basis for that limit.

30.3.3 Exposing the contact force version of pressure

Writing $\eta = \eta_b + h$ allows us to expose the contributions from bottom topography

$$\frac{\partial(h\,\boldsymbol{u})}{\partial t} + \nabla\cdot[\boldsymbol{u}\,(h\,\boldsymbol{u})] + \boldsymbol{f}\wedge(h\,\boldsymbol{u}) = -g\,[(1/2)\nabla h^2 + h\,\nabla\eta_b], \tag{30.47}$$

which takes on the component form

$$\frac{\partial(h\,u)}{\partial t} + \partial_x(h\,u^2 + g\,h^2/2) + \partial_y(h\,u\,v) - v\,h\,f = -\,g\,h\,\partial_x\eta_b \tag{30.48a}$$

$$\frac{\partial(h\,v)}{\partial t} + \partial_x(h\,u\,v) + \partial_y(h\,v^2 + g\,h^2/2) + u\,h\,f = -\,g\,h\,\partial_y\eta_b. \tag{30.48b}$$

The right hand side term,

$$-\,g\,h\,\nabla\eta_b = -(p_{\rm bot}/\rho)\,\nabla\eta_b \tag{30.49}$$

exposes the forcing by the topographic form stress acting on the layer (Section 30.2.5), where

$$p_{\rm bot} = \rho\,g\,h \tag{30.50}$$

is the bottom pressure. Likewise, the term $-\nabla(g\,h^2/2)$ is the gradient of the depth integrated pressure across the layer. We make use of equations (30.48a) and (30.48b) when discussing force balances in a zonally periodic channel in Section 30.4.

30.4 Force balances in a zonally re-entrant channel

In this section we study the depth integated momentum of a shallow water layer within a zonally re-entrant channel such as depicted in Figure 30.5. In particular, we examine the force balances required to reach a steady state in the presence of a wind stress acceleration, $\boldsymbol{\tau}/\rho$. 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. A similar analysis was presented in Section 20.4 for the axial angular momentum budget in a continuously stratified fluid. Following from that analysis, for simplicity we here ignore the role of internal viscous friction. However, we do allow for bottom frictional stresses in the form of quadratic bottom drag

$$\boldsymbol{F}^{\rm drag} = -C_{\rm d}\,\boldsymbol{u}\,|\boldsymbol{u}|, \tag{30.51}$$

where $C_{\rm d} > 0$ 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 thickness weighted equations. That is, all thickness weighted fields vanish at the shoreline edge merely since the thickness vanishes there. We also made use of this property in Section 20.4. We prefer to work with the sloping sides since they are more realistic than the vertical sides commonly used for similar analyses.

30.4.1 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

$$\nabla\cdot(h\,\boldsymbol{u}) = 0. \tag{30.52}$$

As discussed for incompressible flow in Chapter 16, this non-divergence condition means that there is zero net steady transport crossing any simply connected closed contour in the fluid. A particularly

Figure 30.5: *A zonally periodic channel with northern and southern shelves and arbitrary topography. 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 lime-green contour extends across the zonal extent of the channel and is periodic, with a unit vector, $\hat{\boldsymbol{n}}$, depicted normal to a point along the contour. In a steady state, the net fluid transport crossing this contour vanishes: $\oint h\,\boldsymbol{u}\cdot\hat{\boldsymbol{n}}\,\mathrm{d}s = 0$, 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, the vanishing layer thickness found at the edge renders a simple treatment of boundary conditions for the thickness weighted equations.*

interesting closed contour is one that is periodic and extends across the full zonal extent of the channel (see Figure 30.5), in which case

$$\oint h\,\boldsymbol{u}\cdot\hat{\boldsymbol{n}}\,\mathrm{d}s = 0, \tag{30.53}$$

where $\oint$ denotes a periodic line integral across the zonal extent of the channel, $\hat{\boldsymbol{n}}$ is a unit vector normal to the contour, and $\mathrm{d}s$ is the arc-length line element along the contour. The constraint (30.53) 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, then we see that for a steady state there is no net meridional transport across a latitude circle

$$\oint h\,\boldsymbol{u}\cdot\hat{\boldsymbol{y}}\,\mathrm{d}x = \oint h\,v\,\mathrm{d}x = 0 \qquad \text{steady flow.} \tag{30.54}$$

In particular, this result means that the depth weighted Coriolis acceleration appearing in the zonal momentum equation vanishes when integrated zonally

$$\oint f\,h\,v\,\mathrm{d}x = f\oint h\,v\,\mathrm{d}x = 0 \qquad \text{steady flow.} \tag{30.55}$$

We make use of this result in Section 30.4.3.

30.4.2 Steady meridional balances

Consider the meridional thickness weighted momentum equation (30.48b) in the presence of a wind stress

$$\frac{\partial(h\,v)}{\partial t} + \partial_x(h\,u\,v) + \partial_y(h\,v^2 + g\,h^2/2) + u\,h\,f = -g\,h\,\partial_y\eta_b + \tau^y/\rho - C_{\mathrm{d}}\,v\,|\boldsymbol{u}|. \tag{30.56}$$

Integrating zonally over the channel removes the zonal transport term, $\partial_x(h\,u\,v)$, due to periodicity

$$\oint \left[\frac{\partial(h\,v)}{\partial t} + \partial_y(h\,v^2 + g\,h^2/2) + u\,h\,f\right] \mathrm{d}x = \oint \left[-g\,h\,\partial_y\eta_b + \tau^y/\rho - C_\mathrm{d}\,v\,|\boldsymbol{u}|\right] \mathrm{d}x. \tag{30.57}$$

We can pull the time derivative outside of the area 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

$$\frac{\mathrm{d}}{\mathrm{d}t}\oint v\,h\,\mathrm{d}x = \oint \left[-\partial_y(h\,v^2 + g\,h^2/2) - u\,h\,f - g\,h\,\partial_y\eta_b + \tau^y/\rho - C_\mathrm{d}\,v\,|\boldsymbol{u}|\right] \mathrm{d}x. \tag{30.58}$$

Correspondingly, a steady state along a latitude circle is realized by the balance

$$\oint \left[\partial_y(h\,v^2 + g\,h^2/2) + u\,h\,f\right] \mathrm{d}x = \oint \left[-g\,h\,\partial_y\eta_b + \tau^y/\rho - C_\mathrm{d}\,v\,|\boldsymbol{u}|\right] \mathrm{d}x. \tag{30.59}$$

The right hand side represents forcing by the topographic form stress, winds, and bottom drag, with that forcing along a latitude circle balancing the zonal Coriolis acceleration and the meridional divergence of meridional momentum advection plus potential energy (see Section 30.6.1 for potential energy in a shallow water fluid).

30.4.3 Steady zonal balance

Now consider the zonal momentum equation written in its flux-form (see equation (30.48a))

$$\frac{\partial(h\,u)}{\partial t} + \partial_x(h\,u^2 + g\,h^2/2) + \partial_y(h\,u\,v) - v\,h\,f = -g\,h\,\partial_x\eta_b + \tau^x/\rho - C_\mathrm{d}\,u\,|\boldsymbol{u}|. \tag{30.60}$$

Assuming a steady state and integrating along a latitude circle leads to

$$\rho\oint \partial_y(h\,u\,v)\mathrm{d}x = \oint \left[-p_\mathrm{bot}\,\partial_x\eta_b + \tau^x - C_\mathrm{d}\,\rho\,u\,|\boldsymbol{u}|\right] \mathrm{d}x, \tag{30.61}$$

where we dropped the Coriolis acceleration as per volume conservation in equation (30.55), and introduced the bottom pressure $p_\mathrm{bot} = \rho\,g\,h$. 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

$$\oint p_\mathrm{bot}\,\partial_x\eta_b\,\mathrm{d}x \approx \oint \left[\tau^x - C_\mathrm{d}\,\rho\,u\,|\boldsymbol{u}|\right] \mathrm{d}x \qquad \text{nonlinear term small.} \tag{30.62}$$

We realize an exact balance over the full channel domain by integrating the latitude balance (30.61) 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

$$\int \left[\oint p_\mathrm{bot}\,\partial_x\eta_b\,\mathrm{d}x\right] \mathrm{d}y = \int \left[\oint (\tau^x - C_\mathrm{d}\,\rho\,u\,|\boldsymbol{u}|)\,\mathrm{d}x\right] \mathrm{d}y. \tag{30.63}$$

30.4.4 A minor role for frictional bottom stresses

Consider a flat bottom channel, in which case the area integrated balance (30.63) is between winds and bottom drag

$$\int \left[\oint \tau^x\,\mathrm{d}x\right] \mathrm{d}y = \int \left[\oint C_\mathrm{d}\,\rho\,u\,|\boldsymbol{u}|\,\mathrm{d}x\right] \mathrm{d}y \Longrightarrow \langle\tau^x\rangle = C_\mathrm{d}\,\rho\,\langle u\,|\boldsymbol{u}|\rangle, \tag{30.64}$$

where the angle brackets denote a channel area mean. Typical values for bottom drag coefficient are $C_{\rm d} \approx 2 \times 10^{-3}$. An area mean eastward zonal wind stress of $\langle \tau^x \rangle = 0.1$ N m^{-2} leads to a root-mean-square zonal velocity of

$$\sqrt{\langle u\,|\boldsymbol{u}|\rangle} \approx \sqrt{\langle u^2 \rangle} \approx 0.2 \text{ m s}^{-1}, \tag{30.65}$$

where the first approximation follows from assuming the zonal velocity dominates over the meridional velocity. How realistic is this number? 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, then there would be a zonal volume transport of roughly 1500×10^6 m^3 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. They hence inferred that topographic form stress is the dominant term that balances wind stress, and they supported that inference by 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 this conclusion. We observe that this balance is is distinct from that occurring in mid-latitude ocean gyres, where frictional bottom drag at western boundaries provides the dominant sink for vorticity imparted by wind stress curls. The Southern Ocean, and more generally zonally re-entrant channels, are fundamentally distinct from gyres. Given these considerations, we dispense with bottom drag for the remainder of this section.

30.4.5 Correlation between surface height and topography slope

Given the minor role for bottom drag in establishing a steady channel flow, the balance (30.63) says that an eastward wind stress must be balanced, in the area integral over the channel, by a westward topographic form stress

$$\int \left[\oint \tau^x \, \mathrm{d}x \right] \mathrm{d}y = \int \left[\oint p_{\rm bot} \, \partial_x \eta_b \mathrm{d}x \right] \mathrm{d}y. \tag{30.66}$$

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 30.6 described in Section 30.4.7.

To further reveal the correlation between surface height and bottom topography, write $p_{\rm bot} = \rho\, g\, h$ and use $\eta = h + \eta_b$ so that

$$p_{\rm bot}\, \partial_x \eta_b = \rho\, g\, (\eta - \eta_b)\, \partial_x \eta_b = \rho\, g\, \eta\, \partial_x \eta_b - (g/2)\, \partial_x \eta_b^2. \tag{30.67}$$

The balance (30.66) thus becomes

$$\langle \tau^x \rangle = \rho\, g \langle \eta\, \partial_x \eta_b \rangle. \tag{30.68}$$

Furthermore, due to zonal periodicity, it is only the zonal anomalies in η and η_b that contribute so that

$$\langle \tau^x \rangle = \rho\, g \langle \eta'\, \partial_x \eta_b' \rangle, \tag{30.69}$$

where

$$\eta = \eta' + L^{-1} \oint \eta \, \mathrm{d}x \qquad \text{and} \qquad \eta_b = \eta_b' + L^{-1} \oint \eta_b \, \mathrm{d}x, \tag{30.70}$$

with $L = \oint \mathrm{d}x$ 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, $\int \eta' \, \partial_x \eta_b' \, \mathrm{d}A > 0$. 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 (30.69). 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 $\oint \eta' \, \partial_x \eta_b' \, \mathrm{d}x = 0$, 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 of the topography. We consider an explicit example in Section 30.4.7 to help in our understanding.

30.4.6 Connection to meridional geostrophic transport

Zonal periodicity allows us to write the balance (30.69) in the form

$$\langle \tau^x \rangle = -\rho \, g \langle \partial_x \eta' \, \eta_b' \rangle \tag{30.71}$$

Furthermore, $g \, \partial_x \eta'$ is associated with an anomalous meridional geostrophic velocity

$$g \, \partial_x \eta' = f \, v_g' \tag{30.72}$$

so that the balance (30.71) is

$$\langle \tau^x \rangle = -\rho \, \langle f \, v_g' \, \eta_b' \rangle. \tag{30.73}$$

Hence, the steady balance is realized with anomalous meridional geostrophic transport correlated to topographic anomalies.

30.4.7 Sinusoidal example

To help further our understanding of the balance (30.69), 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

$$\eta' = \eta_o \sin(\kappa \, x + \varphi) \qquad \text{and} \qquad \eta_b' = D \sin(\kappa \, x) \tag{30.74}$$

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

$$f \, v_{\mathrm{g}}' = g \, \partial_x \eta' = g \, \kappa \, \eta_o \cos(\kappa \, x + \varphi). \tag{30.75}$$

Plugging into the balance (30.69) leads to

$$\sin \varphi = \left[\frac{L \, \langle \tau^x \rangle}{D \, \rho \, g \, \eta_o \, \pi \, n} \right]. \tag{30.76}$$

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

$$L = 10^7 \text{ m} \qquad D = 10^3 \text{ m} \qquad \eta_o = 1 \text{ m} \qquad \tau^x = 1 \text{ N m}^{-2}, \tag{30.77}$$

in which case

$$\sin \varphi \approx (\pi\, n)^{-1}. \tag{30.78}$$

For $n = 1$, which corresponds to just one topographic bump, then we have a phase shift of $\varphi \approx 18°$, and this case is depicted in Figure 30.6 for the southern hemisphere. If there are two bumps, then the phase shift is reduced to $\varphi \approx 9°$ since each bump shares half the burden of balancing the wind stress.

Figure 30.6: *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 $+\hat{\boldsymbol{x}}$ 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 (30.76). 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 stresss 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 20.9.*

Summarizing the dependencies

We here highlight the various dependencies in the phase equation (30.76).

- 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 in the Southern Ocean that provide the dominant balance for the zonal wind stress.

- What if the parameters are such that the right hand side of equation (30.76) 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 topograhic 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 30.4.4).

Distinguishing steady motion from zero motion

Why is the phase shift (30.76) 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.

30.4.8 Comments

We studied to some extent the role of topographic form stress in balancing wind stress over a zonal channel. Although there are huge idealizations involved with a single shallow water layer, this analysis has direct relevance to a realistic 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 30.2. 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.

30.5 Angular momentum in a rotating cylindrical tank

We here study angular momentum for a layer of inviscid shallow water fluid in a rotating cylindrical tank. This system was first discussed in Section 29.3, where we developed the horizontal equation of motion

$$\frac{\mathrm{D}\boldsymbol{u}}{\mathrm{D}t} + f\,\hat{\boldsymbol{z}} \wedge \boldsymbol{u} = -\nabla\left(p/\rho + g_{\mathrm{e}}\,z - \Omega^2\,r^2/2\right), \tag{30.79}$$

where $r^2 = x^2 + y^2$ is the radial distance from the rotational axis,

$$\Omega = f/2 \tag{30.80}$$

is the constant angular rotation rate, and the vertical component to the right hand side is the hydrostatic balance, $\partial p/\partial z = -\rho\,g_{\mathrm{e}}$. Where convenient, we make use of the polar coordinates (see Chapter 7.3) in the following, in which case

$$x = r\,\cos\theta \tag{30.81a}$$

$$y = r\,\sin\theta, \tag{30.81b}$$

with the polar angle θ measured counter-clockwise from the positive x-axis.

30.5.1 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 10.6 and 18.4)

$$L^z = \delta M\,[\boldsymbol{x} \wedge (\boldsymbol{u} + \boldsymbol{U}_{\mathrm{solid}})] \cdot \hat{\boldsymbol{z}}, \tag{30.82}$$

where $\boldsymbol{x} = x\,\hat{\boldsymbol{x}} + y\,\hat{\boldsymbol{y}} = r\,\hat{\boldsymbol{r}}$ 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

$$\boldsymbol{U}_{\mathrm{solid}} = (f/2)\,\hat{\boldsymbol{z}} \wedge \boldsymbol{x} = r\,\Omega\,\hat{\boldsymbol{\theta}}\ , \tag{30.83}$$

where $\hat{\boldsymbol{z}} \wedge \hat{\boldsymbol{r}} = \hat{\boldsymbol{\theta}}$ is the azimuthal unit vector pointing counter-clockwise around the origin.

We can further massage the expression for the angular momentum by writing

$$\boldsymbol{x} \wedge \boldsymbol{u} = (x\,v - y\,u)\,\hat{\boldsymbol{z}} = r^2\dot{\theta}\,\hat{\boldsymbol{z}}, \tag{30.84}$$

where $\dot{\theta} = \mathrm{D}\theta/\mathrm{D}t$ is the angular velocity. Likewise, we have

$$\boldsymbol{x} \wedge \boldsymbol{U}_{\mathrm{solid}} = r^2\,\Omega\,\hat{\boldsymbol{z}}, \tag{30.85}$$

so that the angular momentum can be written

$$L^z = \delta M\,[\boldsymbol{x} \wedge (\boldsymbol{u} + \boldsymbol{U}_{\mathrm{solid}})] \cdot \hat{\boldsymbol{z}} = \delta M\,r^2\,(\dot{\theta} + \Omega). \tag{30.86}$$

30.5.2 Material time evolution of the angular momentum

The material time evolution for the angular momentum is given by

$$\frac{\mathrm{D}L^z}{\mathrm{D}t} = \delta M \left[\boldsymbol{u} \wedge (\boldsymbol{u} + \boldsymbol{U}_{\text{solid}})\right] \cdot \hat{\boldsymbol{z}} + \delta M \left[\boldsymbol{x} \wedge \left(\frac{\mathrm{D}\boldsymbol{u}}{\mathrm{D}t} + \frac{\mathrm{D}\boldsymbol{U}_{\text{solid}}}{\mathrm{D}t}\right)\right] \cdot \hat{\boldsymbol{z}}. \tag{30.87}$$

Using the solid-body rotation velocity given by equation (30.83), and with a constant rotation rate, yields

$$\boldsymbol{u} \wedge \boldsymbol{U}_{\text{solid}} + \boldsymbol{x} \wedge \frac{\mathrm{D}\boldsymbol{U}_{\text{solid}}}{\mathrm{D}t} = \boldsymbol{u} \wedge (\boldsymbol{\Omega} \wedge \boldsymbol{x}) + \boldsymbol{x} \wedge (\boldsymbol{\Omega} \wedge \boldsymbol{u}) \tag{30.88a}$$

$$= (\boldsymbol{x} \cdot \boldsymbol{u})\, f\, \hat{\boldsymbol{z}}. \tag{30.88b}$$

Making use of the material evolution of the horizontal velocity given by equation (30.79) renders

$$\left[\boldsymbol{x} \wedge \frac{\mathrm{D}\boldsymbol{u}}{\mathrm{D}t}\right] \cdot \hat{\boldsymbol{z}} = \left(\boldsymbol{x} \wedge \left[-f\, \hat{\boldsymbol{z}} \wedge \boldsymbol{u} - \nabla\left(p/\rho + g_{\mathrm{e}}\, z - \Omega^2\, r^2/2\right)\right]\right) \cdot \hat{\boldsymbol{z}} \tag{30.89a}$$

$$= -f\,(\boldsymbol{x} \cdot \boldsymbol{u}) - (\boldsymbol{x} \wedge g\, \nabla\eta) \cdot \hat{\boldsymbol{z}}. \tag{30.89b}$$

The centrifugal term dropped out since

$$\boldsymbol{x} \wedge \nabla r^2 = 2\, \boldsymbol{x} \wedge r\, \hat{\boldsymbol{r}} = 2\, \boldsymbol{x} \wedge \boldsymbol{x} = 0. \tag{30.90}$$

The gravitational term dropped out since

$$(\boldsymbol{x} \wedge \nabla z) \cdot \hat{\boldsymbol{z}} = (\boldsymbol{x} \wedge \hat{\boldsymbol{z}}) \cdot \hat{\boldsymbol{z}} = 0, \tag{30.91}$$

as does the vertical component to the pressure gradient. We are thus left with

$$\frac{1}{\delta M}\, \frac{\mathrm{D}L^z}{\mathrm{D}t} = -g\, (\boldsymbol{x} \wedge \nabla\eta) \cdot \hat{\boldsymbol{z}}. \tag{30.92}$$

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.

The evolution of angular momentum is the same regardless the frame of reference. Hence, there can be no dependence on the Coriolis parameter, which indeed is the case for equation (30.92). That is, the angular momentum is a frame invariant property, so that its evolution is the same whether measured in an inertial or a non-inertial reference frame.

We can bring the expression (30.92) into a more transparent form by switching to polar coordinates

$$\boldsymbol{x} \wedge \nabla\eta = r\, \hat{\boldsymbol{r}} \wedge \left[\hat{\boldsymbol{r}}\, \frac{\partial\eta}{\partial r} + \hat{\boldsymbol{\theta}}\, \frac{1}{r}\, \frac{\partial\eta}{\partial\theta}\right] = \frac{\partial\eta}{\partial\theta}\, \hat{\boldsymbol{z}}, \tag{30.93}$$

so that

$$\frac{1}{\delta M}\, \frac{\mathrm{D}L^z}{\mathrm{D}t} = -g\, \frac{\partial\eta}{\partial\theta}. \tag{30.94}$$

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.

30.5.3 Materially invariant angular momentum

The angular momentum for a fluid column is materially invariant (i.e., a constant on a material fluid parcel) if

$$\frac{\mathrm{D}L^z}{\mathrm{D}t} = 0 \Longleftrightarrow \frac{\partial \eta}{\partial \theta} = 0. \tag{30.95}$$

For a flat bottom, equation (29.57) 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.

30.5.4 Comments

The material evolution equation (30.92) 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 9.12.2). 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.

30.6 Mechanical energy budget for a shallow water layer

We here derive the mechanical energy budget for a single shallow water layer sitting on top of a generally non-flat bottom.

30.6.1 Gravitational potential energy

The gravitational potential energy per horizontal area of a shallow water fluid of constant density is given by

$$\mathscr{P} = g\,\rho \int_{\eta_\mathrm{b}}^{\eta} z\,\mathrm{d}z \tag{30.96a}$$

$$= \frac{g\,\rho}{2}\,(\eta^2 - \eta_\mathrm{b}^2) \tag{30.96b}$$

$$= g\,\rho\,h\,(\eta - h/2). \tag{30.96c}$$

Note how the gravitational potential energy vanishes when $\eta^2 = \eta_\mathrm{b}^2$. For the case $\eta = \eta_\mathrm{b}$, there is no fluid since the free surface sits on top of the bottom, so we expect the potential energy to vanish. For the case $\eta = -\eta_\mathrm{b} > 0$, there is the same amount of fluid above $z = 0$ as below. Since potential energy is computed with respect to a reference state $z = 0$, potential energy vanishes for the case where the same mass of fluid sits beneath $z = 0$ as above. Furthermore, note that in the flat bottom case, $\eta_\mathrm{b} = 0$ so that $h = \eta - \eta_\mathrm{b} = \eta$, in which case the potential energy (30.96c) reduces to

$$\mathscr{P}_\mathrm{flat} = g\,\rho\,h^2/2. \tag{30.97}$$

The material time tendency of the potential energy written in the form (30.96b) is

$$\frac{\mathrm{D}\mathscr{P}}{\mathrm{D}t} = g\,\rho\left[\eta\,\frac{\mathrm{D}\eta}{\mathrm{D}t} - \eta_\mathrm{b}\,\frac{\mathrm{D}\eta_\mathrm{b}}{\mathrm{D}t}\right] \tag{30.98a}$$

$$= g\,\rho\,(\eta\,w_\eta - \eta_\mathrm{b}\,w_\mathrm{b}), \tag{30.98b}$$

where we used equations for the vertical velocity component from Section 29.1.5. 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

$$\frac{\mathrm{D}\mathscr{P}}{\mathrm{D}t} = g\,\rho\,(\eta\,w_\eta - \eta_\mathrm{b}\,w_\mathrm{b}) \tag{30.99a}$$

$$= g\,\rho\,[\eta\,(w_\mathrm{b} - h\,\nabla\cdot\boldsymbol{u}) - \eta_\mathrm{b}\,w_\mathrm{b}] \tag{30.99b}$$

$$= g\,\rho\,[w_\mathrm{b}\,(\eta - \eta_\mathrm{b}) - \eta\,h\,\nabla\cdot\boldsymbol{u}] \tag{30.99c}$$

$$= g\,\rho\,h\,(w_\mathrm{b} - \eta\,\nabla\cdot\boldsymbol{u}), \tag{30.99d}$$

where we used the definition $h = \eta - \eta_\mathrm{b}$. We make use of this form of the potential energy change for developing the mechanical energy budget in Section 30.98b.

30.6.2 Kinetic energy

As for the flat bottom case, the kinetic energy per horizontal area is

$$\mathscr{K} = \frac{1}{2}\int_{\eta_\mathrm{b}}^{\eta} \rho\,\boldsymbol{u}^2\,\mathrm{d}z = \frac{1}{2}\,\rho\,h\,\boldsymbol{u}^2, \tag{30.100}$$

which has a material time derivative given by

$$\frac{\mathrm{D}\mathscr{K}}{\mathrm{D}t} = -g\,\rho\,h\,\boldsymbol{u}\cdot\nabla\eta - \frac{\rho\,h\,\boldsymbol{u}^2}{2}\,\nabla\cdot\boldsymbol{u} \tag{30.101a}$$

$$= -g\,\rho\,h\,\boldsymbol{u}\cdot\nabla\eta - \mathscr{K}\,\nabla\cdot\boldsymbol{u}. \tag{30.101b}$$

$$= -g\,\rho\,h\,\boldsymbol{u}\cdot\nabla\eta + \frac{\mathscr{K}}{h}\,\frac{\mathrm{D}h}{\mathrm{D}t}. \tag{30.101c}$$

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.

30.6.3 Mechanical energy

The material time derivative of the mechanical energy per horizontal area is given by

$$\frac{\mathrm{D}(\mathscr{K}+\mathscr{P})}{\mathrm{D}t} = -g\,h\,\rho\,\boldsymbol{u}\cdot\nabla\eta - \mathscr{K}\,\nabla\cdot\boldsymbol{u} + g\,h\,\rho\,(w_\mathrm{b} - \eta\,\nabla\cdot\boldsymbol{u}). \tag{30.102}$$

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 (30.102) into its Eulerian components leads to

$$\frac{\partial}{\partial t}(\mathscr{K}+\mathscr{P}) + \boldsymbol{u}\cdot\nabla\,(\mathscr{K}+\mathscr{P}) = -g\,h\,\rho\,\boldsymbol{u}\cdot\nabla\eta - \mathscr{K}\,\nabla\cdot\boldsymbol{u} + g\,h\,\rho\,(w_\mathrm{b} - \eta\,\nabla\cdot\boldsymbol{u}), \tag{30.103}$$

with rearrangement rendering

$$\frac{\partial}{\partial t}(\mathcal{K}+\mathcal{P}) + \nabla\cdot[\boldsymbol{u}\,(\mathcal{K}+\mathcal{P})] = \mathcal{P}\,\nabla\cdot\boldsymbol{u} - g\,h\,\rho\,\boldsymbol{u}\cdot\nabla\eta + g\,h\,\rho\,(w_{\mathrm{b}} - \eta\,\nabla\cdot\boldsymbol{u}) \tag{30.104a}$$
$$= (\mathcal{P} - g\,h\,\rho\,\eta)\,\nabla\cdot\boldsymbol{u} + g\,h\,\rho\,(w_{\mathrm{b}} - \boldsymbol{u}\cdot\nabla\eta) \tag{30.104b}$$
$$= -(g\,\rho\,h^2/2)\,\nabla\cdot\boldsymbol{u} + g\,h\,\rho\,(w_{\mathrm{b}} - \boldsymbol{u}\cdot\nabla\eta) \tag{30.104c}$$
$$= -\nabla\cdot(\boldsymbol{u}\,g\,\rho\,h^2/2) + g\,\rho\,h\,\boldsymbol{u}\cdot\nabla h + g\,h\,\rho\,(w_{\mathrm{b}} - \boldsymbol{u}\cdot\nabla\eta) \tag{30.104d}$$
$$= -\nabla\cdot(\boldsymbol{u}\,g\,\rho\,h^2/2) + g\,\rho\,h\,(\boldsymbol{u}\cdot\nabla(h-\eta) + w_{\mathrm{b}}) \tag{30.104e}$$
$$= -\nabla\cdot(\boldsymbol{u}\,g\,\rho\,h^2/2) + g\,\rho\,h\,(-\boldsymbol{u}\cdot\nabla\eta_{\mathrm{b}} + w_{\mathrm{b}}) \tag{30.104f}$$
$$= -\nabla\cdot(\boldsymbol{u}\,g\,\rho\,h^2/2), \tag{30.104g}$$

where we used the identity

$$w_{\mathrm{b}} = \frac{\mathrm{D}\eta_{\mathrm{b}}}{\mathrm{D}t} = \boldsymbol{u}\cdot\nabla\eta_{\mathrm{b}}, \tag{30.105}$$

which follows since $\partial\eta_{\mathrm{b}}/\partial t = 0$. We are thus left with the conservation law of the form

$$\frac{\partial\mathcal{E}}{\partial t} + \nabla\cdot\boldsymbol{F} = 0, \tag{30.106}$$

which has the specific expression

$$\frac{\partial}{\partial t}(\mathcal{K}+\mathcal{P}) + \nabla\cdot[\boldsymbol{u}\,(\mathcal{K}+\mathcal{P}+g\,\rho\,h^2/2)] = 0, \tag{30.107}$$

where the total mechanical energy is

$$\mathcal{E} = \mathcal{K}+\mathcal{P} = \frac{1}{2}\rho\,h\,\boldsymbol{u}^2 + \frac{1}{2}\rho\,g\,(\eta^2 - \eta_{\mathrm{b}}^2), \tag{30.108}$$

and the flux of mechanical energy is

$$\boldsymbol{F} = \boldsymbol{u}\,(\mathcal{K}+\mathcal{P}+g\,\rho\,h^2/2) \tag{30.109a}$$
$$= \boldsymbol{u}\,(\mathcal{K}+\rho\,g\,\eta\,h). \tag{30.109b}$$

Recall in our discussion of total energy for a continously stratified fluid in Section 23.4.3, 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.

30.7 Available potential energy of a shallow water layer

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.

The gravitational potential energy for a single shallow water layer is given by

$$P = g\,\rho\int \mathrm{d}A\int_0^{\eta} z\,\mathrm{d}z = \frac{g\,\rho}{2}\int \eta^2\,\mathrm{d}A, \tag{30.110}$$

where $\int \mathrm{d}A$ 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 sea surface interface to a uniform value $z = H$, so that

$$P_{\mathrm{ref}} = \frac{g\,\rho}{2} \int H^2 \,\mathrm{d}A. \tag{30.111}$$

The available potential energy is the difference

$$E_{\mathrm{APE}} = P - P_{\mathrm{ref}} \tag{30.112a}$$

$$= \frac{g\,\rho}{2} \int (\eta^2 - H^2)\,\mathrm{d}A \tag{30.112b}$$

$$= \frac{g\,\rho}{2} \int \left(\eta^2 - \overline{\eta}^2\right)\,\mathrm{d}A \tag{30.112c}$$

$$= \frac{g\,\rho}{2} \int (\eta')^2 \,\mathrm{d}A, \tag{30.112d}$$

where

$$\overline{\eta} = \frac{\int \eta \,\mathrm{d}A}{A} \tag{30.113}$$

is the area averaged surface height, and

$$\eta' = \eta - \overline{\eta} \tag{30.114}$$

is the anomalous sea surface. Furthermore, we set

$$\overline{\eta} = H \tag{30.115}$$

since the total volume of the shallow water layer is constant. Equation (30.112d) 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.

30.8 Exercises

EXERCISE 30.1: FLOW IN AN OCEAN EDDY

Figure 30.7 shows a vertical-zonal slice through the upper portion of the middle latitude ocean in the northern hemisphere ($f > 0$). The central region consists of a mesoscale eddy, whose signature is a slight 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 30.11). Assume the fluid flow within the eddy region is in geostrophic balance.

(a) If the density is dominated by temperature, is the central portion of the eddy anomously cold or warm relative to the surrounding fluid?

(b) When looking down from above ("plan view"), is the flow within the eddy clockwise or anti-clockwise?

(c) Correspondingly, is the geostrophically balanced circulation within the eddy cyclonic or anti-cyclonic?

(d) If the free surface undulation is on the order of a meter, what is the order of magnitude for the undulation of the pycnocline?

Figure 30.7: *Vertical-zonal slice through a mid-latitude ocean eddy looking from the south to the north (north is into the page). The ocean free surface is depressed down whereas the pycnocline (region of enhanced vertical density gradient) is deformed up.*

(e) Discuss the assumptions used to answer this question, beyond the assumptions made in the question statement.

EXERCISE 30.2: POTENTIAL TEMPERATURE SLOPES IN ATMOSPHERE AND OCEAN

Use the two-layer thermal wind relations from Section 30.1.2, 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 30.3: CIRCULAR STEADY GEOSTROPHIC FLOW

Consider a single layer of shallow water fluid in steady geostrophic balance on a f-plane so that

$$f\,\hat{\boldsymbol{z}} \wedge \boldsymbol{u}_{\mathrm{g}} = -g\,\nabla\eta, \tag{30.116}$$

where $f > 0$ (northern hemisphere). Assume the free surface has a circular Gaussian shape

$$\eta = \eta_0\, e^{-r^2/(2\,\sigma^2)} \tag{30.117}$$

where $r^2 = x^2 + y^2$ is the squared radius and σ is the standard deviation of the Gaussian.

(a) Determine the horizontal geostrophic velocity components corresponding to this free surface undulation.

(b) Determine the streamlines for the flow. Hint: recall the discussion in Section 12.8.2.

EXERCISE 30.4: 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

$$\boldsymbol{P} = \int \rho \, \boldsymbol{u} \, \mathrm{d}V = \int \rho \, h \, \boldsymbol{u} \, \mathrm{d}A. \tag{30.118}$$

Show that

$$\frac{\mathrm{d}\boldsymbol{P}}{\mathrm{d}t} = 0 \tag{30.119}$$

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 30.5: THICKNESS WEIGHTED MOMENTUM IN TWO-LAYER SYSTEM
Following the methods from Section 30.3, derive the thickness weighted momentum equation for an inviscid two-layer stacked shallow water fluid. That is, derive the two-layer version of equation (30.42).

EXERCISE 30.6: TOPOGRAPHIC FORM STRESS FOR A RIDGE
As in Section 30.4, 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, $\eta'_{\mathrm{b}}(x)$. 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 30.8. Following the force balance (30.69), derive an expression for the sea surface height zonally averaged over the upstream side of the ridge, minus the sea surface height zonally averaged over the downstream sides of the ridge,

$$\Delta\eta' = (\overline{\eta'})_{\mathrm{up}} - (\overline{\eta'})_{\mathrm{dn}}, \tag{30.120}$$

where

$$(\overline{\eta'})_{\mathrm{up}} = \frac{\int_{x_{\mathrm{up}}}^{x_0} \eta' \, \mathrm{d}x}{L_{\mathrm{up}}} \qquad \text{and} \qquad (\overline{\eta'})_{\mathrm{dn}} = \frac{\int_{x_0}^{x_{\mathrm{dn}}} \eta' \, \mathrm{d}x}{L_{\mathrm{dn}}}. \tag{30.121}$$

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.

EXERCISE 30.7: 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 30.8: APE FOR A SINGLE SHALLOW WATER LAYER WITH BOTTOM TOPOGRAPHY
Generalize the APE discussion in Section 30.7 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.

Figure 30.8: *Zonal-vertical view of a single layer of shallow water fluid moving in a zonally reenetrant channel for use in Exercise 30.6. 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 30.9: APE FOR TWO SHALLOW WATER LAYERS

Compute the APE for two shallow water layers as in Figure 29.6 with nontrivial bottom topography, $z = \eta_b(x, y)$. Show that the APE is non-negative. Assume the domain is simply connected.

EXERCISE 30.10: 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

$$\frac{E_{\rm KE}}{E_{\rm APE}} \sim \left[\frac{L_{\rm d}^{\rm ext}}{L}\right]^2 . \tag{30.122}$$

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

$$L_{\rm d}^{\rm ext} = \frac{\sqrt{g\,H}}{f} \tag{30.123}$$

is the external deformation radius. The scaling (30.122) means that for scales larger than the external deformation radius, $L_{\rm d}$, the available potential energy is larger than the kinetic energy. The converse holds for scales smaller than $L_{\rm d}$.

EXERCISE 30.11: RATIO OF KE TO APE FOR 1.5 LAYER f-PLANE GEOSTROPHY

Consider a reduced gravity system (Section 29.4) in geostrophic balance on an f-plane. Show that the ratio of kinetic energy to available potential energy scales like

$$\frac{E_{\rm KE}}{E_{\rm APE}} \sim \left[\frac{L_{\rm d}^{\rm int}}{L}\right]^2 . \tag{30.124}$$

In this equation, L is the horizontal length scale for the fluctuation of the internal interface η_1 (i.e., $\nabla\eta_1 \sim \eta_1'/L$), and

$$L_{\rm d}^{\rm int} = \frac{\sqrt{g_1'\,\overline{h}}}{f} \tag{30.125}$$

is the internal deformation radius with $g_1' = g\,(\rho_2 - \rho_1)/\rho_1$ the reduced gravity and $h = A^{-1}\int(\eta_0 - \eta_1)\,{\rm d}A$ the area averaged layer thickness (see Figure 29.4). The scaling (30.124) means that for

scales larger than the internal deformation radius, $L_{\mathrm{d}}^{\mathrm{int}}$, the available potential energy is larger than the kinetic energy, and conversely for scales smaller than $L_{\mathrm{d}}^{\mathrm{int}}$.

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 29.5.
- The scaling (30.124) is identical to that found for the quasi-geostrophic system in Section 42.5.4. However, to solve this exercise it is <u>not</u> sufficient to merely reproduce the scaling discussed in Section 42.5.4. Instead, use here the expressions for APE and KE appropriate for the shallow water system.

31

Shallow water gravity waves

Waves are oscillatory fluctuations that result from a restoring force. 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.[1] To develop the mathematical equations for these gravity waves, we linearize the equations of motion and then develop constraints that must be satisfied for the existence of nontrivial solutions. We identify basic properties of the gravity waves, and study their generation from small amplitude undulating topography. We then examine an initial value problem to illustrate the adjustment of the fluid from an unbalanced state to a geostrophically balanced state.

READER'S GUIDE TO THIS CHAPTER

We provide two key applications of linear gravity waves in this chapter. The first is an explicit dynamical solution that manifests topographic form stress in a zonal channel, following from our discussion in Section 30.4. The second is geostrophic adjustment, which is the transient process whereby an unbalanced state emits gravity waves as it evolves toward geostrophic balance. Topographic form stress and geostrophic adjustment are fundamental aspects of geophysical fluids, with their study in a single shallow water layer allowing us to focus on their essential physical features.

31.1 The linearized shallow water equations **554**
31.1.1 Linearizing the shallow water equations 554
31.1.2 Relative vorticity of linear shallow water fluctuations 555
31.1.3 Potential vorticity of linear shallow water fluctuations 555
31.2 Non-rotating shallow water gravity waves **555**
31.2.1 Wave ansatz and dispersion relation 556
31.2.2 Vanishing relative vorticity 557
31.3 Inertia-gravity waves **558**
31.3.1 Dispersion relation 558
31.3.2 Shortwave limit 559
31.3.3 Longwave limit 560
31.3.4 Horizontally longitudinal and transverse components 560
31.4 Inertia-gravity waves generated by topography **561**
31.4.1 Linearized equations 561

[1] Rossby waves also arise when considering a non-constant Coriolis parameter.

31.4.2 Galilean transformation 562
31.4.3 Wave equation and relation between amplitudes 562
31.4.4 To Do 564
31.5 Geostrophic adjustment **564**
31.5.1 Posing the initial value problem 564
31.5.2 Adjustment in the absence of rotation 565
31.5.3 Adjustment with rotation 565
31.5.4 Comments 567
31.6 Exercises **568**

31.1 The linearized shallow water equations

We here develop the linear shallow water equations, including the velocity, surface height, and potential vorticity equations.

31.1.1 Linearizing the shallow water equations

Recall the shallow water equations of motion are given by the momentum and continuity equations, written here in their Eulerian form

$$\frac{\partial \boldsymbol{u}}{\partial t} + (\boldsymbol{u}\cdot\nabla)\,\boldsymbol{u} + f\,\hat{\boldsymbol{z}}\wedge\boldsymbol{u} = -g\,\nabla\eta \tag{31.1a}$$

$$\frac{\partial h}{\partial t} + h\,\nabla\cdot\boldsymbol{u} + \boldsymbol{u}\cdot\nabla h = 0. \tag{31.1b}$$

Since the bottom is assumed flat, the surface height equals to the thickness (see Figure 29.1)

$$\eta = h \qquad \text{flat bottom.} \tag{31.2}$$

We linearlize the momentum and thickness equation about a state of rest. For this purpose, consider small fluctuations of the thickness and velocity relative to the rest state

$$\eta(x,y,t) = H + \eta'(x,y,t) \tag{31.3a}$$

$$\boldsymbol{u}(x,y,t) = 0 + \boldsymbol{u}'(x,y,t). \tag{31.3b}$$

Substution into the thickness equation (31.1b) leads to

$$\frac{\partial \eta'}{\partial t} + (H+\eta')\,\nabla\cdot\boldsymbol{u}' + \boldsymbol{u}'\cdot\nabla\eta' = 0. \tag{31.4}$$

The products $\eta'\,\nabla\cdot\boldsymbol{u}'$ and $\boldsymbol{u}'\cdot\nabla\eta'$ are second order in fluctuating quantities. Dropping these terms leads to the linearized surface height (or thickness) equation

$$\frac{\partial \eta'}{\partial t} + H\,\nabla\cdot\boldsymbol{u}' = 0 \qquad \text{linearized surface height equation.} \tag{31.5}$$

Similarly, the linearized momentum equation takes the form (dropping the nonlinear advection term $(\boldsymbol{u}'\cdot\nabla)\,\boldsymbol{u}'$)

$$\frac{\partial \boldsymbol{u}'}{\partial t} + f\,\hat{\boldsymbol{z}}\wedge\boldsymbol{u}' = -g\,\nabla\eta' \qquad \text{linearized momentum equation.} \tag{31.6}$$

31.1.2 Relative vorticity of linear shallow water fluctuations

We here consider how the vertical component of relative vorticity

$$\zeta = \hat{\boldsymbol{z}} \cdot (\nabla \wedge \boldsymbol{u}) \tag{31.7}$$

evolves for the small amplitude shallow water fluctuations determined by equations (31.5) and (31.6). Taking the curl of the linearized momentum equation (31.6) leads to

$$\frac{\partial \zeta'}{\partial t} = -\nabla \cdot (f\, \boldsymbol{u}'). \tag{31.8}$$

On the $f-$plane and with the non-divergence condition $\nabla \cdot \boldsymbol{u}' + \partial_z w' = 0$, we have

$$\frac{\partial \zeta'}{\partial t} = f\, \frac{\partial w'}{\partial z}. \tag{31.9}$$

We can use the linearized thickness equation (31.5) to render an equivalent result

$$\frac{\partial \zeta'}{\partial t} = \frac{f}{H}\, \frac{\partial \eta'}{\partial t}. \tag{31.10}$$

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.

31.1.3 Potential vorticity of linear shallow water fluctuations

Rearranging the linearized vorticity eqution (31.10) leads to the local (i.e., no advection) conservation law

$$\frac{\partial q}{\partial t} = 0, \tag{31.11}$$

where the linearized shallow water potential vorticity is given by[2]

$$q = \zeta' - \frac{f \eta'}{H}. \tag{31.12}$$

As mentioned in Section 31.1.2, relative vorticity is locally constant for non-rotating gravity waves. For the rotating case, we see that potential vorticity is locally constant in the presence of inertia-gravity waves. As seen in Section 31.5, this property of the waves provides a constraint on the resulting steady state after the linear fluctuations (i.e., waves) pass.

31.2 Non-rotating shallow water gravity waves

For the non-rotating case ($f = 0$), the linear velocity and thickness equations are

$$\frac{\partial \eta'}{\partial t} = -H\, \nabla \cdot \boldsymbol{u}' \tag{31.13a}$$

$$\frac{\partial \boldsymbol{u}'}{\partial t} = -g\, \nabla \eta'. \tag{31.13b}$$

[2]The lineariezed potential vorticity (31.12) can be obtained by taking the limit of $|f| \gg |\zeta|$ and $H \gg \Delta\eta$ in the shallow water potential vorticity (see Section 33.3) $Q = (f + \zeta)/h$.

Equation (31.13a) 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

$$\frac{\partial^2 \eta'}{\partial t^2} = -H\,\frac{\partial(\nabla\cdot\boldsymbol{u}')}{\partial t} \tag{31.14a}$$

$$\frac{\partial(\nabla\cdot\boldsymbol{u}')}{\partial t} = -g\,\nabla^2\eta'. \tag{31.14b}$$

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

$$\frac{\partial^2 \eta'}{\partial t^2} - g\,H\,\nabla^2\eta' = 0. \tag{31.15}$$

The complement substitution leads to the slightly more general wave equation for the perturbation horizontal velocity

$$\frac{\partial^2 \boldsymbol{u}'}{\partial t^2} - g\,H\,\nabla\,(\nabla\cdot\boldsymbol{u}') = 0. \tag{31.16}$$

We discussed the linear wave equation in Section 3.6, 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.

31.2.1 Wave ansatz and dispersion relation

The linearized shallow water system (31.13a)-(31.13b) exhibits horizontal spatial symmetry. We are thus motivated to consider the linear wave ansatz[3]

$$(u', v', \eta') = (\tilde{u}, \tilde{v}, \tilde{\eta})\,e^{\mathrm{i}(\boldsymbol{k}\cdot\boldsymbol{x}-\omega\,t)}, \tag{31.17}$$

where $\mathrm{i} = \sqrt{-1}$ 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

$$\boldsymbol{k} = \hat{\boldsymbol{x}}\,k + \hat{\boldsymbol{y}}\,l \tag{31.18}$$

is the horizontal vector wave-number and ω is the radial frequency with

$$\tau_{\text{period}} = 2\pi/\omega \tag{31.19}$$

the wave period.

Making use of the linearized momentum equation (31.13b) allows us to connect the amplitude of the free surface undulation to the velocity

$$\tilde{\boldsymbol{u}} = \boldsymbol{k}\,\frac{g\,\tilde{\eta}}{\omega} \Longrightarrow \tilde{\boldsymbol{u}}\wedge\boldsymbol{k} = 0 \qquad \text{and} \qquad \tilde{\boldsymbol{u}}\cdot\boldsymbol{k} = \frac{g\,\tilde{\eta}\,|\boldsymbol{k}|^2}{\omega}. \tag{31.20}$$

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

[3] *Ansatz* is a German word meaning "educated guess". For the linear waves considered here, the wave ansatz (31.17) is a valid solution of the linearized shallow water equations (31.13a)-(31.13b) if the dispersion relation (31.22) is satisfied.

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 (29.36)

$$w(z) = w(\eta_b) - (z - \eta_b)\,\nabla \cdot \boldsymbol{u}. \tag{31.21}$$

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 horizontal motion that is longitudinal.

Substitution of the surface height fluctuation (31.17) into the wave equation (31.15) leads to the relation between frequency and wave-number

$$\omega = \pm c\,|\boldsymbol{k}|, \tag{31.22}$$

where

$$c = \sqrt{g\,H} \tag{31.23}$$

is the shallow water gravity wave speed. We emphasize that the wave ansatz (31.17) is a valid solution of the linearized shallow water equations (31.13a)-(31.13b) only if the *dispersion relation* (31.22) is satisfied.

The dispersion relation (31.22) 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* waves. Furthermore, use of the dispersion relation brings the amplitude relations (31.20) into the form

$$\tilde{\boldsymbol{u}} = \tilde{\eta}\,\frac{\boldsymbol{k}}{|\boldsymbol{k}|}\sqrt{\frac{g}{H}} \Longrightarrow \tilde{\boldsymbol{u}} \cdot \boldsymbol{k} = \tilde{\eta}\,\frac{\omega}{H}. \tag{31.24}$$

Since the wave vector is real, this relation means that the velocity and the free surface are in phase.

31.2.2 Vanishing relative vorticity

Equation (31.14b) 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 (31.13b)), the associated vorticity has a zero time tendency

$$\frac{\partial(\nabla \wedge \boldsymbol{u}')}{\partial t} = 0. \tag{31.25}$$

This result follows from the discussion of vorticity for the shallow water waves in Section 31.1.2. Hence, if the initial flow configuration has zero vorticity, the linear gravity waves retain zero vorticity as they propagate. The velocity fluctuation (31.20) indeed has zero vorticity since

$$\hat{\boldsymbol{z}} \cdot (\nabla \wedge \boldsymbol{u}') = i\,\hat{\boldsymbol{z}} \cdot (\boldsymbol{k} \wedge \tilde{\boldsymbol{u}})\,e^{\mathrm{i}(\boldsymbol{k}\cdot\boldsymbol{x} - \omega\,t)} = 0, \tag{31.26}$$

where we used equation (31.20) to set $\boldsymbol{k} \wedge \tilde{\boldsymbol{u}} = 0$.

31.3 Inertia-gravity waves

Now let the Coriolis parameter, f, be a nonzero constant, so that the linearized thickness equation (31.5) and momentum equation (31.6) take the form

$$\frac{\partial \eta'}{\partial t} = -H\,\nabla\cdot\boldsymbol{u}' \tag{31.27a}$$

$$\frac{\partial \boldsymbol{u}'}{\partial t} + f\,\hat{\boldsymbol{z}}\wedge\boldsymbol{u}' = -g\,\nabla\eta'. \tag{31.27b}$$

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 10.3, 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 31.2. 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 (31.27a), remains the same as for the non-rotating case in equation (31.13a). 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.

31.3.1 Dispersion relation

The linearized shallow water system (31.27a)-(31.27b) exhibits horizontal spatial symmetry given that the rotation is assumed constant. We are thus motivated to consider the linear wave ansatz

$$(u', v', \eta') = (\tilde{u}, \tilde{v}, \tilde{\eta})\, e^{\mathrm{i}(\boldsymbol{k}\cdot\boldsymbol{x} - \omega\, t)}, \tag{31.28}$$

which is the same as assumed for the non-rotating case in Section 31.2.1. Although the same ansatz, as we will see in the following, the dispersion relation and amplitude relations differ due to $f \neq 0$. In particular, we will here need to make use of the complex nature of the wave amplitudes $(\tilde{u}, \tilde{v}, \tilde{\eta})$ in order to realize nontrivial solutions.

Substituting the wave ansatz (31.28) into equations (31.27a)-(31.27b) renders the homogeneous matrix-vector equation

$$\begin{pmatrix} -i\,\omega & -f & i\,g\,k \\ f & -i\,\omega & i\,g\,l \\ i\,H\,k & i\,H\,l & -i\,\omega \end{pmatrix} \begin{pmatrix} \tilde{u} \\ \tilde{v} \\ \tilde{\eta} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}. \tag{31.29}$$

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

$$\omega\left[\omega^2 - f^2 - g\,H\,(k^2 + l^2)\right] = 0. \tag{31.30}$$

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 (31.27a) has $\nabla \cdot \boldsymbol{u}' = 0$. We studied shallow water geostrophic motion in Section 30.1. 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 (31.30) satisfy the quadratic dispersion relation

$$\omega^2 = f^2 \left[1 + L_{\mathrm{d}}^2 \left(k^2 + l^2\right)\right], \tag{31.31}$$

where

$$L_{\mathrm{d}} = \frac{c}{f} = \frac{\sqrt{g\,H}}{f} \tag{31.32}$$

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 (31.31) are known as *inertia-gravity waves* and Figure 31.1 depicts this relation. These waves have a frequency, ω, whose magnitude is greater than the inertial frequency

$$\omega^2 \geq f^2, \tag{31.33}$$

and are thus said to be *super-inertial waves*.

31.3.2 Shortwave limit

The short wave limit is in the regime where

$$k^2 + l^2 \gg \frac{f^2}{g\,H} = \frac{1}{L_{\mathrm{d}}^2}, \tag{31.34}$$

so that the shortwave limit occurs when the wavelength is much shorter than the deformation radius. For a wave moving in the $\hat{\boldsymbol{x}}$ direction with wavelength $\lambda = 2\pi/k$, the shortwave limit occurs when

$$\lambda \ll 2\pi L_{\mathrm{d}}. \tag{31.35}$$

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 (31.31) reduces to the non-rotating dispersion relation (31.22)

$$\omega \approx \pm|\boldsymbol{k}|\,c = \pm|\boldsymbol{k}|\,\sqrt{g\,H}. \tag{31.36}$$

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 31.31.

Figure 31.1: *Illustrating the dispersion relation (31.31) for shallow water inertia-gravity waves (hyperbolic curve). Frequency is scaled by the Coriolis frequency, f, and wavenumber by the deformation radius $L_d = \sqrt{gH}/f$. For small wave numbers ($|\boldsymbol{k}| \ll L_d^{-1}$ or $\lambda \gg 2\,\pi\,L_d$), the inertial-gravity wave frequency approaches the inertia frequency, f. 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 $\omega = k\,\sqrt{g\,H}$. 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.*

31.3.3 Longwave limit

The opposite limit occurs when

$$k^2 + l^2 \ll \frac{1}{L_{\mathrm{d}}^2}, \tag{31.37}$$

so the waves are much longer than the deformation radius. In this limit the dispersion relation is

$$\omega = \pm f. \tag{31.38}$$

This dispersion relation is satisfied by inertial oscillations. As discussed in Section 10.3, 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.

31.3.4 Horizontally longitudinal and transverse components

As discussed in Section 31.2.1, 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 (31.28) into the linearized continuity equation (31.27a) leads to

$$\boldsymbol{k} \cdot \tilde{\boldsymbol{u}} = \tilde{\eta}\,\frac{\omega}{H}, \tag{31.39}$$

which is the same as for the non-rotating case given by equation (31.24). The wave ansatz (31.28) in the velocity equation (31.27b) leads to

$$\hat{\boldsymbol{z}} \cdot (\boldsymbol{k} \wedge \tilde{\boldsymbol{u}}) = -\tilde{\eta}\, \frac{i\, f}{H}, \tag{31.40}$$

which is a result of vanishing linearized potential vorticity as per equation (31.12). In reading this equation, it is important to recall that the wave amplitudes, $(\tilde{\boldsymbol{u}}, \tilde{\eta})$, are generally complex numbers.

Making use of equations (31.39) and (31.40) leads to the following decomposition of the wave velocity amplitude

$$\tilde{\boldsymbol{u}} = \frac{\tilde{\eta}}{H\, |\boldsymbol{k}|^2}\, [\omega\, \boldsymbol{k} - i\, f\, (\boldsymbol{k} \wedge \hat{\boldsymbol{z}})]\,, \tag{31.41}$$

which in turn leads to the wave velocity

$$\boldsymbol{u}' = \frac{\tilde{\eta}}{H\, |\boldsymbol{k}|^2}\, [\omega\, \boldsymbol{k} \cos(\boldsymbol{k} \cdot \boldsymbol{x} - \omega\, t) + f\, (\boldsymbol{k} \wedge \hat{\boldsymbol{z}}) \sin(\boldsymbol{k} \cdot \boldsymbol{x} - \omega\, t)]\,. \tag{31.42}$$

The horizontally transverse component of the velocity vector is $\pi/2$ out of phase with the horizontally longitudinal component. 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.

31.4 Inertia-gravity waves generated by topography

The inertia-gravity waves discussed in Sections 31.3 and 31.3 are the natural modes of oscillation for a shallow water layer with a resting background state. In this section we consider a geophysically relevant means to force gravity waves via the flow of fluid over topography. This study also provides the opportunity to revisit the notion of topographic form stress considered in Section 30.4.

31.4.1 Linearized equations

The physical system consists of flow in a zonally reentrant channel with small amplitude topography, $\eta_{\rm b} = \eta'_{\rm b}$. We also introduce a constant zonal mean flow, $U\, \hat{\boldsymbol{x}}$. 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

$$\eta = H + \eta'_{\rm b} + \eta' \qquad \text{and} \qquad u = U + u' \qquad \text{and} \qquad v = 0 + v', \tag{31.43}$$

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 (31.43) into the shallow water equations (31.1a) and (31.1b) leads to the linearized equations

$$\frac{\partial \boldsymbol{u}'}{\partial t} + U\, \frac{\partial \boldsymbol{u}'}{\partial x} + f\, \hat{\boldsymbol{z}} \wedge \boldsymbol{u}' = -g\, \nabla \eta' \tag{31.44a}$$

$$\frac{\partial (\eta' - \eta'_{\rm b})}{\partial t} + U\, \frac{\partial (\eta' - \eta'_{\rm b})}{\partial x} = -H\, \nabla \cdot \boldsymbol{u}'. \tag{31.44b}$$

We chose to retain $\eta'_{\rm b}$ as part of the time derivative even though the topography is time-independent.

31.4.2 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

$$\bar{t} = t \qquad \text{and} \qquad \overline{x} = x - U\,t \qquad \text{and} \qquad \overline{y} = y. \tag{31.45}$$

Following our discussion of Galilean transformations in Section 12.6, we know that the derivative operators transform according to[4]

$$\frac{\partial}{\partial \bar{t}} = \frac{\partial}{\partial t} + U\,\frac{\partial}{\partial x} \qquad \text{and} \qquad \frac{\partial}{\partial \overline{x}} = \frac{\partial}{\partial x} \qquad \text{and} \qquad \frac{\partial}{\partial \overline{y}} = \frac{\partial}{\partial y}. \tag{31.46}$$

With this coordinate transformation the linear equations (31.44a)-(31.44b) take the form

$$\frac{\partial \boldsymbol{u}'}{\partial \bar{t}} + f\,\hat{\boldsymbol{z}} \wedge \boldsymbol{u}' = -g\,\nabla \eta' \tag{31.47a}$$

$$\frac{\partial(\eta' - \eta'_{\mathrm{b}})}{\partial \bar{t}} = -H\,\nabla \cdot \boldsymbol{u}'. \tag{31.47b}$$

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 $U\,\hat{\boldsymbol{x}}$, the topography moves in the opposite direction since

$$\eta'_{\mathrm{b}}(x, y) = \eta'_{\mathrm{b}}(\overline{x} + U\,\bar{t}, \overline{y}). \tag{31.48}$$

31.4.3 Wave equation and relation between amplitudes

Taking $\partial_{\bar{t}}$ of the continuity equation (31.47b) and then substituting the time derivative of the velocity equation (31.47a) leads to the forced wave equation

$$\frac{\partial^2 \eta'}{\partial \bar{t}^2} - g\,H\,\nabla^2\,\eta' + f\,H\,\zeta' = \frac{\partial^2 \eta'_{\mathrm{b}}}{\partial \bar{t}^2}, \tag{31.49}$$

where

$$\zeta' = \hat{\boldsymbol{z}} \cdot (\nabla \wedge \boldsymbol{u}') = \partial_x v' - \partial_y u' \tag{31.50}$$

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

$$\eta'_{\mathrm{b}} = \eta'_{\mathrm{b}}(x) = \eta'_{\mathrm{b}}(\overline{x} + U\,\bar{t}). \tag{31.51}$$

- Seek a steady solution whereby $\partial_t = 0$ so that $\partial_{\bar{t}} = U\,\partial_x$.

[4]Note that the Galilean transformation considered in Section 12.6 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.

- 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

$$\eta'_{\rm b}(x) = \widetilde{\eta}_{\rm b}\, e^{{\rm i} k_{\rm b}\, x} = \widetilde{\eta}_{\rm b}\, e^{{\rm i} k_{\rm b}\, (\overline{x} + U\, \overline{t})} = \widetilde{\eta}_{\rm b}\, e^{{\rm i}(k_{\rm b}\, \overline{x} - \omega_{\rm b}\, \overline{t})}, \tag{31.52}$$

  where $\widetilde{\eta}_{\rm b}$ is the specified amplitude of the topography, $k_{\rm b}$ is the specified wave number of the topography, and

$$\omega_{\rm b} = -U\, k_{\rm b} \tag{31.53}$$

  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, $k_{\rm b}$, and intrinsic frequency, $\omega_{\rm b}$, of the topography. These considerations lead us to study the wave ansatz

$$(u', v', \eta') = (\tilde{u}, \tilde{v}, \tilde{\eta})\, e^{{\rm i}(k_{\rm b}\, \overline{x} - \omega_{\rm b}\, \overline{t})}. \tag{31.54}$$

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 expression the free surface wave amplitude, $\tilde{\eta}$, in terms of the topography amplitude, $\widetilde{\eta}_{\rm b}$.

Relating the amplitudes

Use of the wave ansatz (31.54) in the wave equation (31.49) leads to

$$(-\omega_{\rm b}^2 + g\, H\, k_{\rm b}^2)\, \tilde{\eta} = -\omega_{\rm b}^2\, \widetilde{\eta}_{\rm b} - i\, f\, H\, k_{\rm b}\, \tilde{v}. \tag{31.55}$$

To simplify this expression, make use of the velocity equation (31.47a) and continuity equation (31.47b), which take on the modal form

$$\omega_{\rm b}\, \hat{\boldsymbol{z}} \cdot (\boldsymbol{k}_{\rm b} \wedge \tilde{\boldsymbol{u}}) = -i\, f\, \boldsymbol{k}_{\rm b} \cdot \tilde{\boldsymbol{u}} \tag{31.56a}$$

$$H\, \boldsymbol{k}_{\rm b} \cdot \tilde{\boldsymbol{u}} = \omega_{\rm b}\, (\tilde{\eta} - \widetilde{\eta}_{\rm b}), \tag{31.56b}$$

which, for $\boldsymbol{k}_{\rm b} = \hat{\boldsymbol{x}}\, k_{\rm b}$, leads to

$$i\, k_{\rm b}\, H\, \tilde{v} = f\, (\tilde{\eta} - \widetilde{\eta}_{\rm b}). \tag{31.57}$$

Substitution of equation (31.57) into the dispersion relation (31.55) leads to

$$\widetilde{\eta} = \left[\frac{\omega_{\rm b}^2 - f^2}{\omega_{\rm b}^2 - f^2 - g\, H\, k_{\rm b}^2}\right] \widetilde{\eta}_{\rm b} = \left[\frac{(f/k_{\rm b})^2 - U^2}{(f/k_{\rm b})^2 + (g\, H - U^2)}\right] \widetilde{\eta}_{\rm b}, \tag{31.58}$$

where the second equality follows by setting $\omega_{\rm b} = -U\, k_{\rm b}$ 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 29.3), 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

$$|\tilde{\eta}| \ll |\tilde{\eta}_{\text{b}}|. \tag{31.59}$$

From the discussion of topographic form stress in Section 30.4, 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.

31.4.4 To Do

- Schematic
- Expressions for velocity wave amplitudes.
- The single layer case corresponds to the $N^2 = 0$ case discussed in *Vallis* (2017) Section 7.7. These waves are decaying away from the topography and have no phase shift so no net form stress. To get a phase shift requires a nonzero N^2 and long waves. Should try to get this effect with a two-layer case. That might be nice to work through, and something distinct from Vallis.

31.5 Geostrophic adjustment

The geostrophic balance presented in Sections 26.3 and 30.1 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 31.1. Consequently, the adjustment consists of linear inertia-gravity waves that maintain a locally static potential vorticity (Section 31.1.3). For brevity in notation, we here drop all primes on the linear fluctuating terms.

31.5.1 Posing the initial value problem

We solve for the $t > 0$ evolution of surface height and velocity by making use of the linearized equations

$$\frac{\partial \boldsymbol{u}}{\partial t} + f\,\hat{\boldsymbol{z}} \wedge \boldsymbol{u} = -g\,\nabla\eta \tag{31.60}$$

$$\frac{\partial \eta}{\partial t} + H\,\nabla\cdot\boldsymbol{u} = 0 \tag{31.61}$$

$$\zeta - \frac{f\eta}{H} = q(x,y) \tag{31.62}$$

where we dropped the primes for brevity, and where $q(x, y)$ is the static potential vorticity determined by the initial conditions (Section 31.1.3). To illustrate the geostrophic adjustment in an analytically tractable manner, consider the following step initial conditions for the surface height

$$\eta(x, t = 0) = \begin{cases} +\eta_0 & x < 0 \\ -\eta_0 & x > 0, \end{cases} \tag{31.63}$$

which can be written

$$\eta(x, t = 0) = -\eta_0 \, \mathrm{sgn}(x). \tag{31.64}$$

The velocity is assumed to be zero initially

$$\boldsymbol{u}(x, y, t = 0) = 0. \tag{31.65}$$

Correspondingly, the initial relative vorticity vanishes so that the linearized potential vorticity is initialized as

$$q(x, y) = \frac{f \eta_0}{H} \, \mathrm{sgn}(x). \tag{31.66}$$

Since $\partial q / \partial t = 0$, 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 their evolution.

31.5.2 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 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 form

$$\eta(x, t) = -\frac{\eta_0}{2} \left[\mathrm{sgn}(x + c\,t) + \mathrm{sgn}(x - c\,t) \right], \tag{31.67}$$

where $c = \sqrt{g\,H}$ is the speed for non-rotating gravity waves. The meridional velocity remains zero, whereas the zonal velocity

$$\frac{\partial u}{\partial t} = -g \, \frac{\partial \eta}{\partial x} \tag{31.68}$$

is given by

$$u(x, t) = \frac{g \eta_0}{2c} \left[\mathrm{sgn}(x + c\,t) - \mathrm{sgn}(x - c\,t) \right]. \tag{31.69}$$

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.

31.5.3 Adjustment with rotation

With rotation, the transient solution consists of inertia-gravity waves 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

$$f\hat{\boldsymbol{z}} \wedge \boldsymbol{u} = -g\,\nabla\eta \tag{31.70}$$

$$\nabla\cdot\boldsymbol{u} = 0 \tag{31.71}$$

$$q = \frac{f\eta_0}{H}\,\mathrm{sgn}(x). \tag{31.72}$$

Conservation of potential vorticity constrains the solution so that the steady state surface height is indeed sloped according to a geostrophically balanced state. That is, an equilibrium state of no-motion is not allowed by potential vorticity conservation.

Computing the equilibrium state

As the flow is geostrophic on an f-plane, we make use of the geostrophic streamfunction

$$\psi = \frac{g\,\eta}{f}. \tag{31.73}$$

The equilibrium state is written in terms of the streamfunction according to

$$u = -\frac{\partial\psi}{\partial y} \qquad v = \frac{\partial\psi}{\partial x} \qquad \left(\nabla^2 - L_{\mathrm{d}}^{-2}\right)\psi = q(x,y), \tag{31.74}$$

where we introduced the *deformation radius*

$$L_{\mathrm{d}} = \frac{\sqrt{g\,H}}{f}. \tag{31.75}$$

We motivate the name "deformation radius" in the following.

The initial condition (31.64) 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

$$\frac{\mathrm{d}^2\psi}{\mathrm{d}x^2} - L_{\mathrm{d}}^{-2}\,\psi = \frac{f\eta_0}{H}\,\mathrm{sgn}(x). \tag{31.76}$$

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

$$\frac{\mathrm{d}^2\psi}{\mathrm{d}x^2} - L_{\mathrm{d}}^{-2}\,\psi = \frac{f\eta_0}{H}. \tag{31.77}$$

The particular solution is

$$\psi_{\mathrm{p}} = -L_{\mathrm{d}}^2\,\frac{f\eta_0}{H} = -\frac{g\eta_0}{f} \tag{31.78}$$

and the homogeneous solution is

$$\psi_{\mathrm{h}} = \frac{g\eta_0}{f}\,\mathrm{e}^{-x/L_{\mathrm{d}}} \tag{31.79}$$

so that

$$\psi = -\frac{g\eta_0}{f}\left[1 - \mathrm{e}^{-x/L_{\mathrm{d}}}\right]. \tag{31.80}$$

The $x < 0$ solution is found similarly, so that

$$\psi = \frac{g\eta_0}{f} \begin{cases} -\left(1 - \mathrm{e}^{-x/L_\mathrm{d}}\right) & x > 0 \\ \left(1 - \mathrm{e}^{x/L_\mathrm{d}}\right) & x < 0, \end{cases} \tag{31.81}$$

which means that the equlibrium surface height is

$$\eta = \eta_0 \begin{cases} -\left(1 - \mathrm{e}^{-x/L_\mathrm{d}}\right) & x > 0 \\ \left(1 - \mathrm{e}^{x/L_\mathrm{d}}\right) & x < 0, \end{cases} \tag{31.82}$$

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

$$u = 0 \qquad v = -\frac{g\eta_0}{f\,L_\mathrm{d}}\,\mathrm{e}^{-|x|/L_\mathrm{d}}. \tag{31.83}$$

The equilibrum velocity thus consists of a jet that is perpendicular to the surface height front.

Figure 31.2: *This is Figure 3.10 of Vallis (2017), illustrating solutions to the linear geostrophic adjustment of a rotating shallow water layer to an surface height perturbation. The top panel shows the initial surface height (31.64), and the second panel the equilibrium surface height (31.82). The third panel shows the equilibrium meridional velocity (31.83) comprised of a jet centered at $x = 0$. The final panel shows the static potential vorticity (31.72).*

31.5.4 Comments

As illustrated in Figure 31.2, the equlibrium 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".

The 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.

31.6 Exercises

EXERCISE 31.1: NON-DIMENSIONALIZED LINEAR SHALLOW WATER EQUATIONS
For the linear equations (31.27a)-(31.27b), introduce

$$\boldsymbol{x} = L\,\hat{\boldsymbol{x}}, \quad \boldsymbol{u}' = U\,\hat{\boldsymbol{u}}, \quad t = \frac{L\,\hat{t}}{U}, \quad f = \hat{f}\,T^{-1}, \quad \eta' = H\,\hat{\eta}, \tag{31.84}$$

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 (31.27a)-(31.27b) and identify the non-dimensional ratio of the advective velocity scale to the non-rotating gravity wave speed

$$\mathrm{Fr} \equiv \frac{U}{\sqrt{g\,H}}. \tag{31.85}$$

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 29.3.

EXERCISE 31.2: 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 boundary[5] 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 the one-dimensional wave equation for the zonal velocity fluctuation.
3. What is the wave speed?
4. Following the discussion in Section 3.6, write the linear wave equation solution in the form

$$u'(x, y, t) = F_1(x + c\,t, y) + F_2(x - c\,t, y), \tag{31.86}$$

   Verify that this form satisfies the wave equation derived above for any general functions F_1 and F_2.
5. What is the exprssion for the surface height displacement, η'?

From the above we can to determine the y-dependence to the Kelvin wave

$$L_{\mathrm{d}}\,\frac{\partial F_1}{\partial y} = F_1 \qquad \text{and} \qquad L_{\mathrm{d}}\,\frac{\partial F_2}{\partial y} = -F_2, \tag{31.87}$$

where $L_{\mathrm{d}} = c/f = \sqrt{g\,H}/f$ is the shallow water deformation radius given by equation (31.32). Solutions to equations (31.87) are given by

$$F_1 = F(x + c\,t)\,e^{(y-y_0)/L_{\mathrm{d}}} \qquad \text{and} \qquad F_2 = G(x - c\,t)\,e^{-(y-y_0)/L_{\mathrm{d}}}. \tag{31.88}$$

[5] Kelvin 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.

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

$$\boldsymbol{u}' = \hat{\boldsymbol{x}}\, e^{-(y-y_0)/L_{\mathrm{d}}}\, G(x - c\,t) \qquad \text{and} \qquad \eta' = (H/g)^{1/2}\, e^{-(y-y_0)/L_{\mathrm{d}}}\, G(x - c\,t). \tag{31.89}$$

The specific form of the wave signal, $G(x - c\,t)$, 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 $\hat{\boldsymbol{x}}$ 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 31.3: 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, $L_{\mathrm{d}}^{\mathrm{int}} = \sqrt{g'\,h}/f$ 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 $g' = g/1000$.

EXERCISE 31.4: GEOSTROPHIC ADJUSTMENT

This exercise revisits many of the points raised in this chapter with the aim to ensure a thorough understanding of geostrophic adjustment. We consider again the linearized shallow water equations on an f-plane with flat bottom topography

$$\frac{\partial \boldsymbol{u}}{\partial t} + f\,\hat{\boldsymbol{z}} \wedge \boldsymbol{u} = -g\,\nabla\eta \tag{31.90a}$$

$$\frac{\partial \eta}{\partial t} + H\,\nabla\cdot\boldsymbol{u} = 0, \tag{31.90b}$$

where

$$h(x,y,t) = H + \eta(x,y,t) \tag{31.91}$$

is the layer thickness, η is the free surface undulation relative to the resting layer, H is the resting layer thickness, f is the constant Coriolis parameter, and g is the constant gravitational acceleration. We dropped primes to reduce notational clutter.

(a) Derive the linearized shallow water potential vorticity conservation equation directly using the linearized shallow water equations.

(b) For the special case of a geostrophic flow, show how to express the horizontal velocity $\boldsymbol{u} = u\,\hat{\boldsymbol{x}} + v\,\hat{\boldsymbol{y}}$ and the thickness h in terms of a scalar streamfunction, ψ.

(c) Consider an arbitrary initial condition for u, v, h and *assume* the system evolves into a state of geostrophic balance at infinite time. Using the results from parts (a) and (b), write down the streamfunction for the geostrophically adjusted final state.

(d) Specialize the result in part (c) to the case of an initial condition of zero motion and initial height $\eta(x, y, t = 0) = A$ inside a circle centered at $(x, y) = (0, 0)$ and $\eta = 0$ outside the circle. Write an equation for the geostrophically adjusted final state in plane polar coordinates coordinates.

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 notion of angular velocity 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 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 (PV) is a strategically chosen component of vorticity whose conservation properties are much simpler than the full vorticity vector. PV has found great use for understanding and predicting geophysical fluid flows.

We start this part of the book by introducing vorticity and circulation in Chapter 32, 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. We follow in Chapter 33 with an introduction to the mechanics of vorticity and potential vorticity within a shallow water fluid. Chapter 34 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 35 we take a brief diversion to explore the formal relation, and distinction, between vorticity and angular momentum. Chapter 36 presents the foundations of potential vorticity, with Chapter 37 focusing on the remarkable impermeability property of the potential vorticity flux. We finish this part of the book in Chapter 38, which formulates the momentum and vorticity evolution equations using generalized vertical coordinates.

32

Vorticity and circulation

Vorticity is a property of the fluid that measures its local spin. 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. Helmholz 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 here are fundamental to the remaining chapters in this part of the book.

32.1 Vorticity **574**
- 32.1.1 Vorticity generates the rotation of line elements 574
- 32.1.2 Rotating reference frame 575
- 32.1.3 There are no vorticity sources 575
- 32.1.4 Vorticity is a pseudo-vector 575
- 32.1.5 Further study 575

32.2 Irrotational flows **575**

32.3 Circulation of the velocity field **576**

32.4 The free vortex **577**
- 32.4.1 Motion of a fluid particle 577
- 32.4.2 Velocity circulation 578

32.5 Translation and solid-body rotation **578**
- 32.5.1 Absolute vorticity equals planetary vorticity plus relative vorticity . 579
- 32.5.2 Solid-body rotation on a plane 580
- 32.5.3 Circulation for solid-body rotation 580
- 32.5.4 Comments 580

32.6 Exercises **581**

32.1 Vorticity

Vorticity is the curl of the velocity field

$$\boldsymbol{\omega} = \nabla \wedge \boldsymbol{v}. \tag{32.1}$$

Vorticity measures the rotation or spin of fluid motion and it does so without reference to an origin. Figure 32.1 provides an example zonal flow with a meridional strain (shear). The vertical component to the vorticity is negative for this flow. 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 32.1: *An example zonal flow with a meridional shear, $\boldsymbol{v} = u(y)\,\hat{\boldsymbol{x}}$, and a corresponding vertical vorticity component that is negative: $\zeta = \hat{\boldsymbol{z}} \cdot (\nabla \wedge \boldsymbol{v}) = \partial v/\partial x - \partial u/\partial y = -\partial u/\partial y < 0$. 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 $\hat{z}$.*

32.1.1 Vorticity generates the rotation of line elements

In Section 13.2 we considered the kinematic evolution of a line element, $\delta\boldsymbol{x}$, separating two fluid particles, with that evolution provided by equation (13.20)

$$\frac{\mathrm{D}(\delta x_m)}{\mathrm{D}t} = \delta x_n \, \frac{\partial v_m}{\partial x_n} \Longrightarrow \frac{\mathrm{D}(\delta \boldsymbol{x})}{\mathrm{D}t} = (\delta \boldsymbol{x} \cdot \nabla)\, \boldsymbol{v}. \tag{32.2}$$

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, $2\,\mathbb{S}_{mn} = \partial_n v_m + \partial_m v_n$, whose action generates changes in the distance between the fluid particles (Section 13.2.6). The anti-symmetric portion to the velocity gradient tensor is known as the rotation tensor, $2\,\mathbb{A}_{mn} = \partial_n v_m - \partial_m v_n$. The rotation tensor is related to vorticity via equation (13.32)

$$\mathbb{A}_{mn} = -\epsilon_{mnp}\,\omega_p/2 \Longleftrightarrow \mathbb{A} = \frac{1}{2}\begin{bmatrix} 0 & -\omega_3 & \omega_2 \\ \omega_3 & 0 & -\omega_1 \\ -\omega_2 & \omega_1 & 0 \end{bmatrix}, \tag{32.3}$$

so that

$$2\,\mathbb{A}_{mn}\,\delta x_n = -\epsilon_{mnp}\,\omega_p\,\delta x_n \Longrightarrow 2\,\mathbb{A}\cdot\delta\boldsymbol{x} = \boldsymbol{\omega}\wedge\delta\boldsymbol{x}. \tag{32.4}$$

From our discussion of rotation in Section 9.4, this equation means that vorticity generates a rigid rotation of a material line element (Section 13.2.7). This result accords with Figure 32.1, whereby vorticity leads to the spin of a test paddle wheel; i.e., the rotation of line elements.

32.1.2 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 (9.48), we know that the velocity observed in the non-rotating reference frame, $\boldsymbol{v}$, is related to the rotating reference frame velocity, $\boldsymbol{v}_{\text{rot}}$, via

$$\boldsymbol{v} = \boldsymbol{v}_{\text{rot}} + \boldsymbol{\Gamma} \wedge \boldsymbol{x}. \tag{32.5}$$

The vorticity in the two reference frames is thus related by

$$\boldsymbol{\omega} = \nabla \wedge \boldsymbol{v} = \nabla \wedge \boldsymbol{v}_{\text{rot}} + \nabla \wedge (\boldsymbol{\Gamma} \wedge \boldsymbol{x}) = \boldsymbol{\omega}_{\text{rot}} + 2\,\boldsymbol{\Gamma}. \tag{32.6}$$

If there is a point in the fluid whereby the rotating reference frame's angular velocity equals to one-half the local vorticity, $\boldsymbol{\Gamma} = \boldsymbol{\omega}/2$, then the rotating reference frame's vorticity vanishes at that point

$$\boldsymbol{\Gamma} = \boldsymbol{\omega}/2 \Longrightarrow \boldsymbol{\omega}_{\text{rot}} = 0. \tag{32.7}$$

Hence, we may interpret $\boldsymbol{\omega}/2$ as the angular velocity of the fluid circulating around any point in the fluid. Furthermore, if the vorticity, $\boldsymbol{\omega}$, is spatially constant, then we can move to a rotating reference frame in which the flow is everywhere irrotational, which is flow having zero vorticity.

32.1.3 There are no vorticity sources

Vorticity has zero divergence

$$\nabla \cdot \boldsymbol{\omega} = \nabla \cdot (\nabla \wedge \boldsymbol{v}) = 0. \tag{32.8}$$

This property is akin to the non-divergent nature of the velocity vector in an incompressible flow (see Chapter 16). However, $\nabla \cdot \boldsymbol{\omega} = 0$ 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 formal kinematic properties of vorticity in Chapter 34.

32.1.4 Vorticity is a pseudo-vector

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. 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 above the south pole (see Figure 9.2).

32.1.5 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.

32.2 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, ignoring the planetary vorticity component, as when focused on small scale motions, there are some geophysically relevant flows with vanishing vorticity. Linear gravity waves in the absence of planetary rotation provide a particularly relevant example (Section 31.2.2). Irrotational fluid flow is characterized by a zero vorticity

$$\boldsymbol{\omega} = 0 = \text{irrotational flow}. \tag{32.9}$$

Since the curl of a gradient vanishes, irrotational flow has a velocity field equal to the gradient of a velocity potential

$$\nabla \wedge \boldsymbol{v} = 0 \Rightarrow \boldsymbol{v} = \nabla \Psi. \tag{32.10}$$

Irrotational flow is therefore sometimes call *potential flow*. Figure 32.2 illustrates a two-dimensional flow field generated by taking the gradient of a scalar potential so that the flow has zero vorticity. 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 25.2), then the velocity potential is a harmonic function since it satisfies Laplace's equation

$$\nabla \cdot \boldsymbol{v} = 0 \Rightarrow \nabla^2 \Psi = 0. \tag{32.11}$$

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.

Figure 32.2: *An example horizontal flow based on a potential,* $\Psi = \sin(x/5)\,\sin(y/5)$. *The flow has zero vorticity,* $\boldsymbol{\omega} \cdot \hat{\boldsymbol{z}} = \zeta = \partial v/\partial x - \partial u/\partial y = 0$, *since the flow is based on a scalar potential:* $\boldsymbol{\omega} = \nabla \wedge \boldsymbol{v} = \nabla \wedge \nabla \Psi = 0$. *This example illustrates how irrotational flow may have nontrivial structure even though a test paddle wheel will not spin since there is zero vorticity.*

32.3 Circulation of the velocity field

The velocity circulation, or more briefly the *circulation*, is defined as the oriented closed loop integral of velocity as projected onto the path of the loop

$$\mathcal{C} \equiv \oint_{\partial \mathcal{S}} \boldsymbol{v} \cdot \mathrm{d}\boldsymbol{r}, \tag{32.12}$$

with Figure 32.3 offering a schematic. The line element, $\mathrm{d}\boldsymbol{r}$, is oriented in the counter-clockwise direction around the circuit $\partial \mathcal{S}$. More precisely, let $\boldsymbol{r}(\varphi)$ 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

$$\mathrm{d}\boldsymbol{r} = \boldsymbol{r}(\varphi + \delta\varphi) - \boldsymbol{r}(\varphi). \tag{32.13}$$

By construction, $\mathrm{d}\boldsymbol{r}$ is tangent to the circuit so that $\boldsymbol{v} \cdot \mathrm{d}\boldsymbol{r}$ picks out the component of the velocity that is tangent to the path.

Figure 32.3: *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, $\boldsymbol{v}\cdot\mathrm{d}\boldsymbol{r}$. 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, $\mathcal{C} = \oint_{\partial\mathcal{S}} \boldsymbol{v}\cdot\mathrm{d}\boldsymbol{r} = \int_{\mathcal{S}} \boldsymbol{\omega}\cdot\hat{\boldsymbol{n}}\,\mathrm{d}\mathcal{S}$.*

Stokes' Theorem (Section 2.6) renders the very important identity

$$\mathcal{C} = \oint_{\partial\mathcal{S}} \boldsymbol{v}\cdot\mathrm{d}\boldsymbol{r} = \int_{\mathcal{S}} (\nabla\wedge\boldsymbol{v})\cdot\hat{\boldsymbol{n}}\,\mathrm{d}\mathcal{S} = \int_{\mathcal{S}} \boldsymbol{\omega}\cdot\hat{\boldsymbol{n}}\,\mathrm{d}\mathcal{S}, \tag{32.14}$$

where $\hat{\boldsymbol{n}}$ 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 Helmholz to the circulation theories of Kelvin.

32.4 The free vortex

Consider a two-dimensional rotating fluid in the $x-y$ plane with angular velocity given by

$$\boldsymbol{\Omega} = \frac{\boldsymbol{x}\wedge\boldsymbol{v}}{r^2} = \frac{K\,\hat{\boldsymbol{z}}}{r^2}. \tag{32.15}$$

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 $\hat{\boldsymbol{z}}$ 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. Some refer to this flow as the "vr-vortex" whereas we refer to it as a *free vortex*. As shown in this section, the fluid flow associated with the free vortex has zero vorticity and zero velocity 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 32.4, and pursued in more detail in Chapter 35, nonzero angular momentum can arise in a fluid with zero circulation so long as there is a nonzero strain within the fluid.

32.4.1 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, $\boldsymbol{x} = x\,\hat{\boldsymbol{x}} + y\,\hat{\boldsymbol{y}}$, with respect to the origin

$$\boldsymbol{v}\cdot\boldsymbol{x} = 0. \tag{32.16}$$

The velocity for this pure rotational flow is given by (see Section 9.4)

$$\boldsymbol{v} = \boldsymbol{\Omega} \wedge \boldsymbol{x} = \frac{K\,(-y\,\hat{\boldsymbol{x}} + x\,\hat{\boldsymbol{y}})}{r^2} = \frac{K\hat{\boldsymbol{\theta}}}{r}, \tag{32.17}$$

where $\hat{\boldsymbol{\theta}}$ is the polar angle unit vector oriented in the counter-clockwise direction (see Section 7.3). Each component of the velocity falls off as $1/r$ when moving away from the origin. Away from the origin, the vorticity vector vanishes

$$\boldsymbol{\omega} = \nabla \wedge \boldsymbol{v} = 0, \tag{32.18}$$

whereas it is singular at the origin.

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 strain in the fluid (see Exercise 32.4), 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

$$\boldsymbol{r} \wedge \boldsymbol{v} = r\,\hat{\boldsymbol{r}} \wedge (K/r)\,\hat{\boldsymbol{\theta}} = K\,\hat{\boldsymbol{z}}. \tag{32.19}$$

This result follows since the velocity falls of 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 (32.17) in Figure 32.4 along with two 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.

32.4.2 Velocity circulation

The velocity 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

$$\mathcal{C} = \oint_{\partial\mathcal{S}} \boldsymbol{v} \cdot \mathrm{d}\boldsymbol{r} = \int_0^{2\pi} \boldsymbol{v} \cdot \hat{\boldsymbol{\theta}}\, r\, \mathrm{d}\theta = 2\pi K. \tag{32.20}$$

To reach this result, we set the line element to

$$\mathrm{d}\boldsymbol{r} = \hat{\boldsymbol{\theta}} r\, \mathrm{d}\theta \tag{32.21}$$

and inserted the velocity (32.17) represented in cylindrical polar coordinates, $\boldsymbol{v} \cdot \hat{\boldsymbol{\theta}} = K/r$. Hence, the singular point vortex at $r = 0$ induces a nonzero circulation for all circuits that enclose the vortex.

32.5 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

$$\boldsymbol{v} = \boldsymbol{U} + \boldsymbol{\Omega} \wedge \boldsymbol{x}, \tag{32.22}$$

Figure 32.4: *Irrotational counter-clockwise planar flow in the presence of a free vortex with velocity* $\boldsymbol{v} = (K/r)\,\hat{\boldsymbol{\theta}}$. *The tangential velocity decays as* $1/r$ *from the origin and the vorticity,* $\nabla \wedge \boldsymbol{v}$, *vanishes for all points except the origin. Also, test paddle wheels 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.

where $\boldsymbol{x}$ is the position vector relative to the origin, $\boldsymbol{U}$ is a translation velocity, and $\boldsymbol{\Omega}$ is an angular velocity. For rigid body motion, both $\boldsymbol{U}$ 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 13.1)

$$S_{mn} = \frac{1}{2}(\partial_m v_n + \partial_n v_m) = 0. \tag{32.23}$$

However, the vorticity is nonzero (see Exercise 32.1)

$$\boldsymbol{\omega} = \nabla \wedge (\boldsymbol{\Omega} \wedge \boldsymbol{x}) = 2\,\boldsymbol{\Omega}. \tag{32.24}$$

We encountered this vorticity in Section 32.1.2 when connecting vorticity and angular velocity. Note that the factor of two in this equation is geometric; it needs no physical explanation.

32.5.1 Absolute vorticity equals planetary vorticity plus relative vorticity

For planetary fluid dynamics, rotation of the planet imparts *planetary vorticity* to fluids. Hence, the total or *absolute* vorticity of a fliud is the vector sum of the *relative vorticity*, $\boldsymbol{\omega}$, plus the planetary vorticity

$$\boldsymbol{\omega}_{\text{absolute}} = \boldsymbol{\omega}_{\text{planet}} + \boldsymbol{\omega}. \tag{32.25}$$

In this equation,

$$\boldsymbol{\omega}_{\text{planet}} = 2\,\boldsymbol{\Omega}_{\text{planet}} \tag{32.26}$$

is the planetary vorticity associated with solid-body motion of a fluid particle stationary with respect to the planet, and

$$\boldsymbol{\omega} = \nabla \wedge \boldsymbol{v} \tag{32.27}$$

is the relative vorticity. The relative vorticity measures the vorticity of the fluid due to motion relative to the rotating sphere, with $\boldsymbol{v}$ the velocity relative to the rotating sphere.

32.5.2 Solid-body rotation on a plane

Consider the circular solid-body rotation on a plane shown in Figure 32.5, in which the velocity is purely tangential and linearly proportional to the distance from the center

$$\boldsymbol{v} = \boldsymbol{\Omega} \wedge \boldsymbol{x} = |\Omega| \left(-y\,\hat{\boldsymbol{x}} + x\,\hat{\boldsymbol{y}}\right) = |\Omega|\, r\,\hat{\boldsymbol{\theta}}. \tag{32.28}$$

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 32.5.

32.5.3 Circulation for solid-body rotation

For solid-body rotation, the velocity circulation around a circular path of radius R is given by

$$\mathcal{C} = \oint \boldsymbol{v} \cdot \mathrm{d}\boldsymbol{r} = \oint (\boldsymbol{\Omega} \wedge \boldsymbol{r}) \cdot \mathrm{d}\boldsymbol{r} = R^2\,|\Omega| \oint \mathrm{d}\theta = 2\pi R^2\,|\Omega| = 2A\,|\Omega|, \tag{32.29}$$

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

$$\mathcal{C}/A = |\boldsymbol{\omega}| = 2\,|\boldsymbol{\Omega}|. \tag{32.30}$$

Figure 32.5: *Rigid body fluid motion, whereby the fluid velocity is purely tangential and linearly proportional to the radial distance from the vortex center,* $\boldsymbol{v} = |\Omega|\, r\,\hat{\boldsymbol{\theta}}$*; fluid particles maintain a fixed relative position; and vorticity is constant and points perpendicular to the page,* $\boldsymbol{\omega} = 2\,\boldsymbol{\Omega} = 2\,|\Omega|\,\hat{\boldsymbol{z}}$*. Test paddle wheels rigidly move around the center, and they exhibit a spin about their axis thus manifesting the nonzero vorticity.*

32.5.4 Comments

As seen in Section 32.4, 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 35 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 nicely illustrated in this short video

32.6 Exercises

EXERCISE 32.1: VORTICITY FOR SOLID-BODY ROTATION
Show that a fluid in solid-body rotation with angular velocity

$$\boldsymbol{v}_{\text{solid-body}} = \boldsymbol{\Omega} \wedge \boldsymbol{x}, \tag{32.31}$$

has a vorticity given by

$$\nabla \wedge \boldsymbol{v}_{\text{solid-body}} = 2\,\boldsymbol{\Omega}. \tag{32.32}$$

EXERCISE 32.2: PLANETARY ROTATION IS NON-DIVERGENT
Show that a fluid in solid-body rotation with angular velocity

$$\boldsymbol{v}_{\text{solid-body}} = \boldsymbol{\Omega} \wedge \boldsymbol{x}, \tag{32.33}$$

has zero divergence

$$\nabla \cdot \boldsymbol{v}_{\text{solid-body}} = 0. \tag{32.34}$$

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.

EXERCISE 32.3: VELOCITY POTENTIAL FOR THE FREE VORTEX
What is the velocity potential (32.10) for the free vortex whose velocity field is given by (32.17)? Hint: The problem is two-dimensional and rotationally symmetric, so it is convenient to make use of polar coordinates $x = r\cos\theta$ and $y = r\sin\theta$ as in Appendix 7.3.

EXERCISE 32.4: STRAIN TENSOR FOR THE FREE VORTEX
Determine all components to the strain tensor

$$S_{pq} = \begin{bmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{bmatrix} = \begin{bmatrix} \frac{\partial u}{\partial x} & \frac{1}{2}\left[\frac{\partial u}{\partial y} + \frac{\partial v}{\partial x}\right] \\ \frac{1}{2}\left[\frac{\partial v}{\partial x} + \frac{\partial u}{\partial y}\right] & \frac{\partial v}{\partial y} \end{bmatrix} \tag{32.35}$$

for the free vortex as specified by the velocity field (32.17). Present the answer in the form of a 2×2 matrix.

EXERCISE 32.5: VANISHING VISCOUS FRICTION FOR SOLID-BODY MOTION
As seen in Section 47.2.5, viscous effects from molecular viscosity in an incompressible fluid appear in the momentum equation as a Laplacian weighted by a constant molecular viscosity

$$\text{viscous force per mass} = \mu\,\nabla^2 \boldsymbol{v}, \tag{32.36}$$

where μ is the molecular kinematic viscosity, assumed here to be a constant. Show that the viscous opeator 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.

33

Shallow water vorticity

We consider vorticity and potential vorticity within the shallow water system. In this chapter we derive 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 very important mechanical constraint on the fluid flow. We consider a variety of flow regimes, mostly with an ocean focus, where the study of vorticity and potential vorticity enhance our understanding of the fluid mechanics.

READER'S GUIDE FOR THIS CHAPTER

The shallow water fluid offers a useful conceptual model to introduce the dynamical equations for vorticity and potential vorticity while requiring a relatively modest level of mathematical details. Even so, we here require vector calculus identities for Cartesian coordinates as detailed in Chapter 2. We also require an understanding of the shallow water mechanics from Chapters 29 and 30, as well as the vorticity kinematics introduced in Chapter 32. The concepts and methods developed in this chapter are fundamental to the remaining chapters in this part of the book.

33.1 Shallow water vorticity dynamics **584**
33.2 PV for a rotating deformable cylinder **585**
33.2.1 Mass conservation 585
33.2.2 Angular momentum conservation 586
33.2.3 Potential vorticity conservation 587
33.2.4 Connecting angular momentum and vorticity 587
33.2.5 Comments and further reading 587
33.3 Potential vorticity for a shallow water layer **588**
33.3.1 Mass conservation plus the vorticity equation 588
33.3.2 Motivating the name 589
33.3.3 Mass conservation + Kelvin's circulation theorem 590
33.3.4 Material conservation of an arbitrary function of PV 591
33.4 Example implications of PV material invariance **591**
33.4.1 Absolute vorticity invariance when h is fixed 591
33.4.2 Planetary geostrophic PV 593
33.4.3 Spin up of converging flow 594
33.4.4 Further study 594
33.5 Circulation and a role for non-conservative processes **595**

33.5.1 Material time evolution of PV 595
33.5.2 Eulerian flux-form conservation equation 596
33.5.3 Circulation around a closed horizontal contour 597
33.5.4 Circulation from wind stress and Rayleigh drag 598
33.5.5 Further study 599
33.6 Western intensification of ocean gyres **599**
33.6.1 Steady and large-scale vorticity balance 599
33.6.2 Planetary geostrophic flow and the Sverdrup balance 600
33.6.3 Sverdrup flow in a closed domain with anti-cyclonic wind stress . . . 600
33.6.4 Western intensification and the role of beta 601
33.6.5 A role for topographic form stress 602
33.6.6 Comments and further study 604
33.7 Exercises **604**

33.1 Shallow water vorticity dynamics

In this section we formulate the vorticity equation for the shallow water fluid. For this purpose we introduce vorticity for the full velocity field as well as that just for the horizontal flow

$$\boldsymbol{\omega} = \nabla \wedge \boldsymbol{v} \qquad \text{and} \qquad \boldsymbol{\omega}^* = \nabla \wedge \boldsymbol{u} = \left[\frac{\partial v}{\partial x} - \frac{\partial u}{\partial y}\right] \hat{\boldsymbol{z}} = \zeta\, \hat{\boldsymbol{z}}, \tag{33.1}$$

where

$$\zeta = \hat{\boldsymbol{z}} \cdot \boldsymbol{\omega} = \hat{\boldsymbol{z}} \cdot \boldsymbol{\omega}^* = \frac{\partial v}{\partial x} - \frac{\partial u}{\partial y} \tag{33.2}$$

is the vertical component to the relative vorticity. The vector identity (see Section 2.3.4)

$$(\boldsymbol{u} \cdot \nabla)\, \boldsymbol{u} = (1/2)\, \nabla(\boldsymbol{u} \cdot \boldsymbol{u}) - \boldsymbol{u} \wedge (\nabla \wedge \boldsymbol{u}) \tag{33.3}$$

brings the inviscid shallow water momentum equation (29.5)

$$\frac{\partial \boldsymbol{u}}{\partial t} + (\boldsymbol{u} \cdot \nabla)\, \boldsymbol{u} + f\, \hat{\boldsymbol{z}} \wedge \boldsymbol{u} = -g\, \nabla \eta \tag{33.4}$$

into its "vector-invariant" form

$$\frac{\partial \boldsymbol{u}}{\partial t} + \boldsymbol{\omega}_{\mathrm{a}}^* \wedge \boldsymbol{u} = -\nabla \left(g\, \eta + \boldsymbol{u}^2/2\right). \tag{33.5}$$

We here introduced the absolute vorticity

$$\boldsymbol{\omega}_{\mathrm{a}}^* = (\zeta + f)\, \hat{\boldsymbol{z}}, \tag{33.6}$$

which is the sum of the relative vorticity of the horizontal flow, $\boldsymbol{\omega}^* = \zeta\, \hat{\boldsymbol{z}}$, plus the solid-body vorticity, $f\, \hat{\boldsymbol{z}}$, due to motion of the rotating reference frame (recall Section 32.5.1).

We next make use of the vector identity from Section 2.3.4

$$\nabla \wedge (\boldsymbol{\omega}_{\mathrm{a}}^* \wedge \boldsymbol{u}) = \boldsymbol{\omega}_{\mathrm{a}}^* (\nabla \cdot \boldsymbol{u}) - \boldsymbol{u}\, (\boldsymbol{\omega}_{\mathrm{a}}^* \cdot \nabla) + (\boldsymbol{u} \cdot \nabla)\, \boldsymbol{\omega}_{\mathrm{a}}^* - (\boldsymbol{\omega}_{\mathrm{a}}^* \cdot \nabla)\, \boldsymbol{u} \tag{33.7a}$$

$$= \boldsymbol{\omega}_{\mathrm{a}}^* (\nabla \cdot \boldsymbol{u}) + (\boldsymbol{u} \cdot \nabla)\, \boldsymbol{\omega}_{\mathrm{a}}^*. \tag{33.7b}$$

This result required setting

$$\nabla \cdot \boldsymbol{\omega}_{\mathrm{a}}^* = \nabla \cdot \boldsymbol{\omega}^* + \nabla \cdot (f\, \hat{\boldsymbol{z}}) = 0, \tag{33.8}$$

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

$$(\boldsymbol{\omega}_{\mathrm{a}}^{*} \cdot \nabla)\, \boldsymbol{u} = |\boldsymbol{\omega}_{\mathrm{a}}^{*}|\, \partial_z \boldsymbol{u} = 0, \tag{33.9}$$

which follows since the horizontal velocity in a shallow water fluid is depth independent (see Section 29.1). Applying the operator $\hat{\boldsymbol{z}} \cdot (\nabla \wedge\,)$ onto the vector-invariant momentum equation (33.5) yields a prognostic equation for the absolute vorticity

$$\frac{\partial \zeta_{\mathrm{a}}}{\partial t} + \nabla \cdot (\boldsymbol{u}\, \zeta_{\mathrm{a}}) = 0, \tag{33.10}$$

where

$$\zeta_{\mathrm{a}} = \hat{\boldsymbol{z}} \cdot (\boldsymbol{\omega}^{*} + \hat{\boldsymbol{z}}\, f) = \hat{\boldsymbol{z}} \cdot (\boldsymbol{\omega} + \hat{\boldsymbol{z}}\, f) = \zeta + f \tag{33.11}$$

is the vertical component of the absolute vorticity.

The Eulerian flux-form evolution equation (33.10) means that the vertical component to the absolute vorticity at a point in the inviscid shallow water fluid changes according to the horizontal convergence of vorticity advected to that point

$$\frac{\partial \zeta_{\mathrm{a}}}{\partial t} = -\nabla \cdot (\boldsymbol{u}\, \zeta_{\mathrm{a}}). \tag{33.12}$$

Alternatively, we can write the vorticity equation (33.10) in the material form

$$\frac{\mathrm{D}\zeta_{\mathrm{a}}}{\mathrm{D}t} = -\zeta_{\mathrm{a}}\, \nabla \cdot \boldsymbol{u}, \tag{33.13}$$

where the material time derivative for the shallow water fluid is determined just by the horizontal flow

$$\frac{\mathrm{D}}{\mathrm{D}t} = \frac{\partial}{\partial t} + \boldsymbol{u} \cdot \nabla. \tag{33.14}$$

The material evolution equation (33.13) 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 multipied by the absolute vorticity.

33.2 PV for a rotating deformable cylinder

To conceptually introduce potential vorticity, consider a deformable 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 example may appear contrived for a fluid. However, it offers a useful conceptual picture for a rotating material region of a constant density fluid, in which time derivatives in the following are interpreted as material derivatives. It also lends useful intuition for the motion of spinning fluids constrained by mass and angular momentum conservation.

33.2.1 Mass conservation

Mass conservation is a kinematic property of the cylinder. With a constant density, mass conservation means that the volume of the cylinder is fixed. Hence, mass conservation constrains the relative changes to the radius and height of the cylinder. A materially constant cylinder mass thus renders

$$M = \pi\, R^2\, H\, \rho \tag{33.15}$$

implies

$$\frac{2}{R}\frac{\mathrm{D}R}{\mathrm{D}t} = -\frac{1}{H}\frac{\mathrm{D}H}{\mathrm{D}t}. \tag{33.16}$$

That is, mass conservation means that as twice the relative radius increases the relative height decreases.

33.2.2 Angular momentum conservation

A second constraint on cylinder rotation 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

$$\boldsymbol{\Omega} = \Omega\,\hat{\boldsymbol{z}}. \tag{33.17}$$

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)

$$I_{mn} = \delta_{mn}\,\frac{MR^2}{2}. \tag{33.18}$$

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

$$\boldsymbol{L} = \frac{MR^2}{2}\,\Omega\,\hat{\boldsymbol{z}}. \tag{33.19}$$

Figure 33.1: *Illustrating the conservation of angular momentum for a constant mass rotating cylinder undergoing solid-body rotation around its central axis. The moment of inertia (relative to the central axis) for the left configuration is larger since more of its mass is distributed away from the central axis than in the right configuration. Assuming the two configurations have identical angular momentum means that the right configuration rotates more rapidly since its moment of inertia is smaller. This example exemplifies the familiar ice-skater experience, whereby the skater's spin 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 familiar "ice skater" example occurs when the cylinder radius changes (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 the angular velocity Ω increases (rotates faster) when the cylinder radius decreases, and vice versa. Explicitly for the cylinder we have $\mathrm{d}\boldsymbol{L}/\mathrm{d}t = 0$ and $\mathrm{d}M/\mathrm{d}t = 0$ thus rendering

$$\frac{2}{R}\frac{\mathrm{D}R}{\mathrm{D}t} = -\frac{1}{\Omega}\frac{\mathrm{D}\Omega}{\mathrm{D}t}. \tag{33.20}$$

This tradeoff between spin rate and radius holds in general and we encounter it again in Figure 33.8 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 leads (converging mass), through angular momentum conservation, to an increase in rotation speed. The opposite occurs when mass diverges from a region, thus reducing the rotation speed.

33.2.3 Potential vorticity conservation

Combining angular momentum conservation (33.20) with mass conservation (33.16) leads to the conservation law

$$\frac{\mathrm{D}}{\mathrm{D}t}\frac{\Omega}{H} = 0. \tag{33.21}$$

Equation (33.21) is a statement of potential vorticity conservation for the material fluid column, with potential vorticity given by

$$Q \equiv \frac{\Omega}{H}. \tag{33.22}$$

For example, if the column thickens then the rotational velocity increases to maintain $Q = \Omega/H$ constant.

33.2.4 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 (33.22) is generalized to the absolute vorticity. We encounter this generalization in Section 33.3. Furthermore, as shown in Section 32.5, the vorticity for solid body motion equals to twice the rotation rate, $2\,\boldsymbol{\Omega}$. Hence, the numerator for the potential vorticity of the solid-body rotating cylinder equals to one-half the vorticity. In Chapter 35, we connect angular momentum and vorticity (and strain) for arbitrary fluid motion.

33.2.5 Comments and further reading

The discussion in this section is motivated by a similar presentation given by *Salmon* (1998). The solid-body rotating cylinder succintly identifies the two mechanical properties contributing to the potential vorticity conservation law (33.21): a kinematic property (mass conservation) and a dynamical property (angular momentum conservation). Notably, 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 study this effect later in ths chapter and more thoroughly in Section 34.6.

33.3 Potential vorticity for a shallow water layer

We now consider the potential vorticity (PV) for a single layer of shallow water fluid. The result is directly analogous to that derived for the rotating cylinder in Section 33.2. However, the derivation here makes use of fluid mechanical notions rather than rigid body dynamics. 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 (we discuss Kelvin's Theorem in Chapter 34). Figure 33.2 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.

Figure 33.2: *Illustrating the conservation of PV for a layer of shallow water fluid. 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 $\zeta\,A$).*

33.3.1 Mass conservation plus the vorticity equation

To derive the potential vorticity equation, we here make use of the vorticity equation (33.13) and combine it with mass conservation.

Shallow water vorticity and vortex stretching

Mass conservation in the form of the material thickness equation (29.17) leads to the following expression for the divergence of the horizontal velocity

$$\nabla \cdot \boldsymbol{u} = -\frac{1}{h}\frac{\mathrm{D}h}{\mathrm{D}t}. \tag{33.23}$$

Making use of this result in the vorticity equation (33.13) allows us to eliminate the horizontal divergence, thus yielding

$$\frac{\mathrm{D}\zeta_{\mathrm{a}}}{\mathrm{D}t} = \frac{\zeta_{\mathrm{a}}}{h}\frac{\mathrm{D}h}{\mathrm{D}t}. \tag{33.24}$$

This equation says that material changes in shallow water absolute vorticity arise only from material changes in the layer thickness. Hence, shallow water absolute vorticity changes if the fluid column stretches or compresses. We see in Section 34.5.2 that vorticity in more general fluids is affected by vortex stretching as well as vortex tilting and torques from baroclinicity. The shallow water fluid is thus a very special case whereby absolute vorticity changes materially only through vortex stretching. This behavior is a result of the depth independence of the horizontal velocity and the associated column-like nature of shallow water fluid motion.

Material invariance of PV

Equation (33.24) can be written as an expression of the material invariance of the shallow water potential vorticity

$$\frac{\mathrm{D}Q}{\mathrm{D}t} = 0, \tag{33.25}$$

where

$$Q = \frac{\zeta_{\mathrm{a}}}{h} = \frac{\zeta + f}{h} \tag{33.26}$$

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 (33.25) says that this ratio remains constant for the shallow water layer.

33.3.2 Motivating the name

We can understand the "potential" in the name by noting that potential vorticity measures the potential 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 (see Figure 33.3). 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 22.9.

Figure 33.3: *Left panel: an arbitrary shallow water column with zero relative vorticity and potential vorticity $Q = f/h$. Right panel: the same fluid column moved to the equator (where $f = 0$) and stretched to have the standard thickness, $h_{\mathrm{s}} > h$. The relative vorticity of the column at the equator is given by $\zeta = f\,(h_{\mathrm{s}}/h)$, with $f > 0$ assumed for this figure (northern hemisphere). 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 as part of the name.*

33.3.3 Mass conservation + Kelvin's circulation theorem

Although we have yet to discuss Kelvin's Theorem (Section 34.2), we here invoke it to illustrate another way to derive shallow water PV conservation. When applied to an infinitesimal circuit in an inviscid and constant density fluid, Kelvin's theorem says that

$$\frac{\mathrm{D}\left(\boldsymbol{\omega}_{\mathrm{a}} \cdot \hat{\boldsymbol{n}}\, \delta \mathcal{S}\right)}{\mathrm{D}t} = 0, \tag{33.27}$$

where $\boldsymbol{\omega}_{\mathrm{a}}$ is the absolute vorticity

$$\boldsymbol{\omega}_{\mathrm{a}} = \boldsymbol{\omega} + f\, \hat{\boldsymbol{z}}, \tag{33.28}$$

$\hat{\boldsymbol{n}}\, \delta \mathcal{S}$ is the infinitesimal surface area enclosed by the closed circuit, with $\hat{\boldsymbol{n}}$ the outward normal to the surface. For the shallow water system, we decompose absolute vorticity into

$$\boldsymbol{\omega}_{\mathrm{a}} = \hat{\boldsymbol{z}}\, (\zeta + f) + \boldsymbol{\omega}_h, \tag{33.29}$$

where ζ is the vertical component to the relative vorticity (equation (33.2)), and

$$\boldsymbol{\omega}_h = \hat{\boldsymbol{x}}\, \frac{\partial w}{\partial y} - \hat{\boldsymbol{y}}\, \frac{\partial w}{\partial x} \tag{33.30}$$

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 (33.29) into Kelvin's theorem (33.29) leads to

$$\frac{\mathrm{D}}{\mathrm{D}t}\left[(\zeta + f)\, \delta A + \boldsymbol{\omega}_h \cdot \hat{\boldsymbol{n}}\, \delta S\right] = 0, \tag{33.31}$$

where the horizontal area element, δA, is the projection of the surface area element onto the vertical direction

$$\delta A = \hat{\boldsymbol{z}} \cdot \hat{\boldsymbol{n}}\, \mathrm{d}\mathcal{S}. \tag{33.32}$$

Shallow water fluid dynamics 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 (33.31) is much smaller than the first. Ignoring this term then leads to

$$\frac{\mathrm{D}}{\mathrm{D}t}\left[\left(\frac{\zeta + f}{h}\right) h\, \delta A\right] = 0, \tag{33.33}$$

where h is the layer thickness and $h\, \delta A$ is the volume of an infinitesimal fluid column. The volume of a column of shallow water fluid is materially constant

$$\frac{\mathrm{D}\left(h\, \delta A\right)}{\mathrm{D}t} = 0, \tag{33.34}$$

so that equation (33.33) yields the material invariance of shallow water PV

$$\frac{\mathrm{D}}{\mathrm{D}t}\left[\frac{\zeta + f}{h}\right] = \frac{\mathrm{D}Q}{\mathrm{D}t} = 0, \tag{33.35}$$

where $Q = (\zeta + f)/h$ is the same shallow water potential vorticity derived above in Section 33.3.1. The essence of this derivation, for the case of constant f, is depicted in Figure 33.4.

$$\frac{\mathrm{D}(h\,\delta A)}{\mathrm{D}t} = 0 \qquad \frac{\mathrm{D}[\delta A\,(f+\zeta)]}{\mathrm{D}t} = 0 \implies \frac{\mathrm{D}[(f+\zeta)/h]}{\mathrm{D}t} = 0$$

Figure 33.4: *Material invariance for PV in a shallow water layer results from combining material invariance of the volume of the fluid column to the 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 beta effect discussed in Section 34.6). We here depict that case with f constant, so that converging flow causes the column to spin faster by increasing its relative vorticity, just like the spinning cylinder in Figure 33.1. Conversely, as the cross-sectional area increases, as in a divergening flow, then the column spins slower by reducing its relative vorticity.*

33.3.4 Material conservation of an arbitrary function of PV

The material conservation law for PV, equation (33.25), means that any function, $F(Q)$ is also materially constant. We see this property through the chain rule

$$\frac{\mathrm{D}F}{\mathrm{D}t} = F'(Q)\left[\frac{\mathrm{D}Q}{\mathrm{D}t}\right] = 0. \tag{33.36}$$

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.

33.4 Example implications of PV material invariance

The material invariance of shallow water PV provides a nontrivial constraint on the fluid motion, saying that f, h, ζ cannot change independently of the other. Rather, the combination $Q = (f+\zeta)/h$ must remain materially invariant. 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 PV-thinking.

33.4.1 Absolute vorticity invariance when h is fixed

To illustrate the power of material invariance of PV to constrain the flow, consider a column of shallow water fluid that maintains a fixed thickness. In this case the material invariance of PV reduces to the material invariance of absolute vorticity

$$\frac{\mathrm{D}(\zeta+f)}{\mathrm{D}t} = 0 \qquad \text{if } h = \text{constant}. \tag{33.37}$$

This situation holds for the two-dimensional barotropic flows described in Chapter 39 when moving over a flat bottom domain. To use this constraint for a column of shallow water fluid, we observe

that a column of fluid curving to the left (facing downstream) is associated with a positive relative vorticity for the column, $\zeta > 0$. In contrast, the oppositely curved column has negative relative vorticity, $\zeta < 0$. If the motion is initially zonal but then turns, the meridional motion on the sphere is associated with a change in planetary vorticity since f changes. To maintain constant absolute vorticity, $\zeta + f$, when a column moves meridionally requires the relative vorticity induced by the curved motion to counteract the change in planetary vorticity. We now show that the constraint of fixed absolute vorticity means that eastward flow (westerly winds) cannot turn meridionally while maintaining fixed absolute vorticity, whereas westward flow (easterly winds) can.

Figure 33.5: *Illustrating the constraints on a homogeneous and constant thickness fluid layer imposed by material invariance of absolute vorticity: $\zeta + f$ = constant. A realization of this case is a two-dimensional barotropic fluid (Chapter 39) moving over a flat bottom domain. In each of the four cases depicted, the entering flow has zero relative vorticity, $\zeta = 0$, which means that absolute vorticity must remain constant at the initial Coriolis parameter, $\zeta_a = f_{initial}$. The red eastward flow (westerly winds) is disallowed by material invariance of absolute vorticity. In these cases 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 show motion for both the northern and southern hemispheres. This figure is taken after 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 has an associated $\zeta < 0$ and an increase in the planetary vorticity (f increases). Likewise, a westward flowing fluid column that turns equatorward (to the left) has a positive 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 hold in the southern hemisphere. These scenarios are depicted in Figure 33.5.

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 $\zeta > 0$ 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 $\zeta < 0$ 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 33.6, whereby inviscid flow in the interior of a fluid 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.

Figure 33.6: *Illustrating the constraints on a homogeneous and constant thickness fluid layer (blue region) imposed by material invariance of absolute vorticity: $\zeta + f =$ constant. As per the results from Figure 33.5, inviscid flow can enter a western boundary layer as depicted here, whereas it cannot enter an eastern boundary layer. This figure is taken after Figure 19.12 of Vallis (2017).*

33.4.2 Planetary geostrophic PV

As introduced in Section 26.4, planetary geostrophy (PG) is used to study the large-scale laminar ocean circulation where relative vorticity is ignored. Furthermore, as shown in Section 40.3, the inviscid and adiabatic PG system materially preserves the PG potential vorticity, $Q = f/h$, so that

$$\frac{D(f/h)}{Dt} = 0. \tag{33.38}$$

Consequently, fluid particles respecting the inviscid PG equations follow contours of constant f/h. 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 33.7 illustrates sample f/H

contours for a topographic bump in the southern hemisphere. Fluid columns are steered towards the equator as they encounter the bump. Those near to the bump are trapped as they circle around the bump. There are hence two regimes of contours: those that are trapped by the bump and cannot escape to *infinity*, and those that are untrapped.

Figure 33.7: *When a shallow water fluid is governed by the inviscid and adiabatic planetary geostrophic equations, fluid columns maintain fixed PG potential vorticity,* $\mathrm{D}(f/h)/\mathrm{D}t = 0$. *Ignoring free surface undulations relative to the bottom depth means that* f/H *remains fixed following the inviscid geostrophic flow, where* $z = -H(x, y) = \eta_\mathrm{b}(x, y)$ *is the bottom topography. We illustrate these geostrophic contours for a Gaussian bump (shaded feature) in the southern hemisphere. Contours of constant* f/H *that are far from the bump follow lines of constant latitude, on which* f *is constant. Contours closer to the bump are diverted according to the requirement of keeping* f/H *fixed. When* H *decreases, as per a bump, then the* f/H *contours move equatorward. Trajectories near the bump are trapped as they circle around the bump.*

33.4.3 Spin up of converging flow

Consider the flow shown in Figure 33.8, whereby mass in the shallow water layer converges into a region. Just as described in the PV derivation Figure 33.4, 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 33.2, 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 33.4).

33.4.4 Further study

Section 4.3 of *Holton* (1992) discusses the case of flow over topography where the full shallow water PV is materially invariant, $\mathrm{D}(f + \zeta)/\mathrm{D}t = 0$. 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 33.8: *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 33.1 and 33.4. Namely, converging a region of constant mass reduces its moment of inertia so that angular momentum conservation leads to an increase in positive spin.*

a useful controlled setting to observe leewaves, such as shown near the 20 minute mark in this video from Prof. Dave 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.

33.5 Circulation and a role for non-conservative processes

In this section we consider the role of a non-conservative force per mass, $\boldsymbol{F}$, with this term arising from friction and boundary stresses. Additionally, we allow for the presence of a boundary volume source, $\mathbb{p}$ (as in precipitation minus evaporation), thus changing the volume in the layer (see also Exercise 29.2). With these two non-conservative processes, the shallow water equations take on the form

$$\frac{\mathrm{D}\boldsymbol{u}}{\mathrm{D}t} + f\,\hat{\boldsymbol{z}} \wedge \boldsymbol{u} = -g\,\nabla\eta + \boldsymbol{F} \qquad \text{and} \qquad \frac{\mathrm{D}h}{\mathrm{D}t} + h\,\nabla\cdot\boldsymbol{u} = \mathbb{p}. \tag{33.39}$$

We study how these processes affect both the vorticity and potential vorticity.

33.5.1 Material time evolution of PV

In the presence of friction, the absolute vorticity equation (33.10) takes on the form

$$\frac{\partial\zeta_{\mathrm{a}}}{\partial t} + \nabla\cdot(\boldsymbol{u}\,\zeta_{\mathrm{a}}) = \hat{\boldsymbol{z}}\cdot(\nabla\wedge\boldsymbol{F}) \Longrightarrow \frac{\mathrm{D}\zeta_{\mathrm{a}}}{\mathrm{D}t} + \zeta_{\mathrm{a}}\,\nabla\cdot\boldsymbol{u} = \hat{\boldsymbol{z}}\cdot(\nabla\wedge\boldsymbol{F}), \tag{33.40}$$

so that vorticity is now affected by the curl of friction. As before, we next make use of the thickness equation to replace the horizontal divergence according to

$$\nabla\cdot\boldsymbol{u} = \frac{1}{h}\left[-\frac{\mathrm{D}h}{\mathrm{D}t} + \mathbb{p}\right]. \tag{33.41}$$

The presence of $\mathbb{p}$ thus modifies the divergence of the horizontal velocity beyond that for a conservative fluid. We are thus led to the PV equation

$$h\,\frac{\mathrm{D}Q}{\mathrm{D}t} = -Q\,\mathbb{p} + \hat{\boldsymbol{z}}\cdot(\nabla\wedge\boldsymbol{F}). \tag{33.42}$$

Hence, $\mathbb{p}\neq 0$ and/or $\hat{\boldsymbol{z}}\cdot(\nabla\wedge\boldsymbol{F})\neq 0$ render potential vorticity no longer materially invariant.

33.5.2 Eulerian flux-form conservation equation

We can convert equation (33.42) into an Eulerian flux-form conservation equation by making use of the thickness equation

$$h\,\frac{\mathrm{D}Q}{\mathrm{D}t} = h\left[\frac{\partial Q}{\partial t} + \boldsymbol{u}\cdot\nabla Q\right] + Q\left[\frac{\partial h}{\mathrm{D}t} + \nabla\cdot(h\,\boldsymbol{u}) - \mathbb{p}\right] = \frac{\partial(h\,Q)}{\partial t} + \nabla\cdot(\boldsymbol{u}\,h\,Q) - Q\,\mathbb{p}, \tag{33.43}$$

to render

$$\frac{\partial(h\,Q)}{\partial t} + \nabla\cdot(h\,\boldsymbol{u}\,Q) = \hat{\boldsymbol{z}}\cdot(\nabla\wedge\boldsymbol{F}). \tag{33.44}$$

As a final step, make use of the identity

$$\hat{\boldsymbol{z}}\cdot(\nabla\wedge\boldsymbol{F}) = -\nabla\cdot(\hat{\boldsymbol{z}}\wedge\boldsymbol{F}), \tag{33.45}$$

so that the thickness weighted PV equation (33.44) can be written

$$\frac{\partial(h\,Q)}{\partial t} = -\nabla\cdot(h\,Q\,\boldsymbol{u} + \hat{\boldsymbol{z}}\wedge\boldsymbol{F}). \tag{33.46}$$

We offer the following remarks on this equation.

- Note how the volume source term, $\mathbb{p}$, does not explicitly appear in the flux-form equation (33.46) since the effects from $\mathbb{p}$ are captured by the divergence, $\nabla\cdot\boldsymbol{u}$, as per equation (33.41).

- It is remarkable that even with non-conservative forcing, the thickness weighted potential vorticity is determined by the convergence of a flux,

$$\boldsymbol{J} = h\,Q\,\boldsymbol{u} + \hat{\boldsymbol{z}}\wedge\boldsymbol{F}. \tag{33.47}$$

  This form for the evolution of "PV-stuff" leads to special properties for potential vorticity budgets, and we focus on these properties in Chapters 37 and 46.

- For the shallow water fluid, the thickness weighted potential vorticity equals to the absolute vorticity

$$h\,Q = \zeta_{\mathrm{a}}. \tag{33.48}$$

  Consequently, the flux-form conservation form of the PV equation (33.46) is identical to equation (33.40) for the absolute vorticity

$$\frac{\partial(h\,Q)}{\partial t} = -\nabla\cdot(h\,Q\,\boldsymbol{u} + \hat{\boldsymbol{z}}\wedge\boldsymbol{F}) \Longleftrightarrow \frac{\partial\zeta_{\mathrm{a}}}{\partial t} = -\nabla\cdot(\zeta_{\mathrm{a}}\,\boldsymbol{u} + \hat{\boldsymbol{z}}\wedge\boldsymbol{F}). \tag{33.49}$$

33.5.3 Circulation around a closed horizontal contour

As an application of the vorticity equation (33.49), consider the evolution of circulation around the boundary of a horizontal area $\mathcal{A}$ (see Figure 33.9)

$$\mathcal{C} = \int_{\mathcal{A}} \zeta \, \mathrm{d}A = \oint_{\partial\mathcal{A}} \boldsymbol{u} \cdot \mathrm{d}\boldsymbol{l}. \tag{33.50}$$

In this equation, $\mathrm{d}\boldsymbol{l} = \hat{\boldsymbol{t}} \, \mathrm{d}l$ is the vector line increment around the boundary, and $\hat{\boldsymbol{t}}$ is the tangent vector orienting the contour integral in a counterclockwise direction. Since the area is horizontal, so too is the tangent vector $\hat{\boldsymbol{t}}$.

Figure 33.9: *Illustrating the calculation of circulation around a static horizontal area within a shallow water fluid, with the circulation theorem (33.54) providing the means to compute the time changes in circulation as a function of the advection of absolute vorticity crossing the boundary plus the circulation of friction. The unit outward normal is $\hat{\boldsymbol{n}}$ and the unit tangent is $\hat{\boldsymbol{t}}$.*

Formulation

To develop an evolution equation for $\mathcal{C}$ we integrate the vorticity equation (33.49) over the area $\mathcal{A}$ to yield

$$\int_{\mathcal{A}} \frac{\partial \zeta}{\partial t} \, \mathrm{d}A = \int_{\mathcal{A}} [-\nabla \cdot (\zeta_{\mathrm{a}} \, \boldsymbol{u} + \hat{\boldsymbol{z}} \wedge \boldsymbol{F})] \, \mathrm{d}A = -\oint_{\partial\mathcal{A}} [\zeta_{\mathrm{a}} \, \boldsymbol{u} + \hat{\boldsymbol{z}} \wedge \boldsymbol{F})] \cdot \hat{\boldsymbol{n}} \, \mathrm{d}l, \tag{33.51}$$

where $\oint$ symbolizes an integral around the area boundary. To reach this equation we used Gauss's divergence theorem on the right hand side, with $\hat{\boldsymbol{n}}$ the horizontal outward normal vector along the area boundary, $\partial\mathcal{A}$, and $\mathrm{d}l$ is the line increment along the boundary contour, We also set $\partial f/\partial t = 0$ as part of the time derivative of the absolute vorticity. The friction term can be written as

$$(\hat{\boldsymbol{z}} \wedge \boldsymbol{F}) \cdot \hat{\boldsymbol{n}} = (\hat{\boldsymbol{n}} \wedge \hat{\boldsymbol{z}}) \cdot \boldsymbol{F} = -\hat{\boldsymbol{t}} \cdot \boldsymbol{F}, \tag{33.52}$$

thus leading to

$$\int_{\mathcal{A}} \frac{\partial \zeta}{\partial t} \, \mathrm{d}A = -\oint_{\partial\mathcal{A}} \zeta_{\mathrm{a}} \, \boldsymbol{u} \cdot \hat{\boldsymbol{n}} \, \mathrm{d}l + \oint_{\partial\mathcal{A}} \boldsymbol{F} \cdot \hat{\boldsymbol{t}} \, \mathrm{d}l, \tag{33.53}$$

where $\oint$ is the counter-clockwise oriented closed integral around the boundary. We next assume the area $\mathcal{A}$ is constant in time, so that the Eulerian time derivative can be pulled across the integral to render

$$\frac{\mathrm{d}\mathcal{C}}{\mathrm{d}t} = -\oint_{\partial\mathcal{A}} \zeta_{\mathrm{a}}\, \boldsymbol{u}\cdot\hat{\boldsymbol{n}}\,\mathrm{d}l + \oint_{\partial\mathcal{A}} \boldsymbol{F}\cdot\hat{\boldsymbol{t}}\,\mathrm{d}l. \tag{33.54}$$

The first term on the right hand side arises from the horizontal advection of absolute vorticity across the area boundary, and the second term arises from the circulation of any non-conservative accelerations. We now consider various specialized forms of this circulation theorem.

Integration around a closed streamline

For steady circulation the thickness equation (29.16) means that $\nabla\cdot(h\,\boldsymbol{u}) = 0$. Hence, we can introduce a streamfunction, Ψ (with dimensions of $\mathrm{L}^3\ \mathrm{T}^{-1}$), so that

$$h\,\boldsymbol{u} = \hat{\boldsymbol{z}}\wedge\nabla\Psi. \tag{33.55}$$

In a bounded domain with no boundary volume sources, the contours of constant Ψ, the streamlines, are closed. Integrating the steady state form of the circulation theorem (33.54) around a closed streamline thus leads to the constraint on the non-conservative forcing

$$\oint_{\text{streamline}} \boldsymbol{F}\cdot\hat{\boldsymbol{t}}\,\mathrm{d}l = \int_{\text{streamline area}} (\nabla\wedge\boldsymbol{F})\cdot\hat{\boldsymbol{z}}\,\mathrm{d}A = 0, \tag{33.56}$$

where we set $\boldsymbol{u}\cdot\hat{\boldsymbol{n}} = 0$ along a streamline and the second equality follows from Stokes' theorem applied over the area bounded by the streamline. Equation (33.56) provides a constraint on the non-conservative forcing that must be satisfied to enable a steady flow. Notably, this constraint was arrived at by assuming a steady circulation and it only has a diagnostic value of use with knowledge of the streamlines. Even so, it is useful for our study of ocean gyres in Section 33.6.

33.5.4 Circulation from wind stress and Rayleigh drag

A particularly simple form of non-conservative acceleration is given by

$$\boldsymbol{F} = -\lambda\,\boldsymbol{u} + \boldsymbol{f}_{\text{wind}}. \tag{33.57}$$

The first term is referred to as Rayleigh drag with $\lambda > 0$ a constant inverse time scale. Rayleigh drag damps all flow to rest with λ^{-1} the e-folding time for the damping. The second term in equation (33.57), $\boldsymbol{f}_{\text{wind}}$, arises from surface boundary forcing such as wind stress. For a homogeneous shallow water layer we imagine this wind is applied throughout the layer depth as a body stress rather than as a contact stress. Correspondingly, $\boldsymbol{f}_{\text{wind}}$ is a normalized version of the surface stress, where the normalization is by the layer thickness so to not over-count the transfer of momentum across the layer boundary. In a differential form, $\boldsymbol{f}_{\text{wind}} = \rho_0^{-1}\partial_z\boldsymbol{\tau}_{\text{wind}}$, where $\boldsymbol{\tau}_{\text{wind}}$ is the wind stress applied as a contact stress to the ocean surface in a continously stratified fluid (see, for example, Section 26.4). This treatment of surface stresses as a body force is common for models comprised of homogeneous layers.

Plugging the acceleration (33.57) into the time dependent circulation theorem (33.54) leads to

$$\left[\frac{\mathrm{d}}{\mathrm{d}t} + \lambda\right]\mathcal{C} = -\int_{\partial\mathcal{A}} \zeta_{\mathrm{a}}\,\boldsymbol{u}\cdot\hat{\boldsymbol{n}}\,\mathrm{d}l + \oint_{\partial\mathcal{A}} \boldsymbol{f}_{\text{wind}}\cdot\hat{\boldsymbol{t}}\,\mathrm{d}l. \tag{33.58}$$

Specializing to a steady state and choosing the contour as a closed streamline (along which $\boldsymbol{u}\cdot\hat{\boldsymbol{n}} = 0$), renders

$$\mathcal{C} = \frac{1}{\lambda} \oint_{\text{streamline}} \boldsymbol{f}_{\text{wind}} \cdot \hat{\boldsymbol{t}}\, \mathrm{d}l. \tag{33.59}$$

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 λ^{-1}.

33.5.5 Further study

Elements of this section are motivated by Godfrey's Island Rule as presented by *Tomczak and Godfrey* (1994) and *Pedlosky et al.* (1997), although their presentations generally assume a two-dimensionally non-divergent depth-integrated flow without friction. Here we allowed for horizontal divergence by working with a single-layer of shallow water fluid. We also incorporated non-conservative terms from friction and boundary stresses.

33.6 Western intensification of ocean gyres

Ocean gyres are a prominent feature of ocean circulation, with the North Atlantic middle-latitude gyre a canonical example. 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 24.2). We discuss the beta effect more formally in Section 34.6, 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 30.4.

33.6.1 Steady and large-scale vorticity balance

The steady circulation theorem (33.59) 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 (33.49), again in the presence of wind forcing and Rayleigh drag

$$\nabla \cdot (\boldsymbol{u}\,\zeta_{\text{a}}) = -\lambda\,\zeta + \hat{\boldsymbol{z}} \cdot (\nabla \wedge \boldsymbol{f}_{\text{wind}}). \tag{33.60}$$

The steady thickness equation (29.16) means that $\nabla \cdot (h\,\boldsymbol{u}) = 0$, so that we can write

$$\nabla \cdot (\boldsymbol{u}\,\zeta_{\text{a}}) = \nabla \cdot (h\,\boldsymbol{u}\,Q), \tag{33.61}$$

where $Q = \zeta_{\text{a}}/h$ is the potential vorticity. We thus see that

$$h\,\boldsymbol{u} \cdot \nabla Q = -\lambda\,\zeta + \hat{\boldsymbol{z}} \cdot (\nabla \wedge \boldsymbol{f}_{\text{wind}}), \tag{33.62}$$

which reflects the ability of Rayleigh drag and wind stress curl to cause PV to be modified when following the flow, whereas in their absence PV is a flow invariant.

33.6.2 Planetary geostrophic flow and the Sverdrup balance

For large-scale flow away from lateral boundaries, the flow has an absolute vorticity dominated by planetary vorticity so that

$$Q \approx Q_{\rm PG} = f/h, \tag{33.63}$$

which is the PV for planetary geostrophic flow (Section 26.4 and 40.3). Correspondingly, we assume the Rayleigh drag term is negligible in this region since the relative vorticity is small, so that the PV equation (33.62) takes the form

$$h\,\boldsymbol{u}\cdot\nabla Q_{\rm PG} = \hat{\boldsymbol{z}}\cdot(\nabla\wedge\boldsymbol{f}_{\rm wind}). \tag{33.64}$$

Expanding the left hand side and introducing the planetary vorticity gradient renders the *shallow water Sverdrup balance*

$$\beta\,v = Q_{\rm PG}\,\boldsymbol{u}\cdot\nabla h + \hat{\boldsymbol{z}}\cdot(\nabla\wedge\boldsymbol{f}_{\rm wind}) \qquad \text{shallow water Sverdrup balance.} \tag{33.65}$$

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 29.1). 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

$$\beta\,v = \hat{\boldsymbol{z}}\cdot(\nabla\wedge\boldsymbol{f}_{\rm wind}) \qquad \text{traditional Sverdrup balance.} \tag{33.66}$$

33.6.3 Sverdrup flow in a closed domain with anti-cyclonic wind stress

As a particular example, consider a closed northern hemisphere middle latitude domain driven by an anti-cyclonic wind stress

$$\hat{\boldsymbol{z}}\cdot(\nabla\wedge\boldsymbol{f}_{\rm wind}) < 0 \qquad \text{northern hemisphere anti-cyclonic wind stress.} \tag{33.67}$$

This situation is depicted in Figure 33.10, where we also illustrate a commonly used wind stress profile that are purely zonal and have a co-sinusoidal meridional structure that is symmetric about the central latitude of the domain

$$\boldsymbol{f}_{\rm wind} = -\hat{\boldsymbol{x}}\,A\,\cos[\pi\,(y-y_0+L/2)/L] \tag{33.68a}$$

$$\hat{\boldsymbol{z}}\cdot(\nabla\wedge\boldsymbol{f}_{\rm wind}) = -(\pi\,A/L)\sin[\pi\,(y-y_0+L/2)/L]. \tag{33.68b}$$

In these equations, $A>0$ is the magnitude of the wind stress acceleration applied to the layer, and the domain extends meridionally from $y_0 - L/2 \le y \le y_0 + L/2$ 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 $\hat{\boldsymbol{z}}\cdot(\nabla\wedge\boldsymbol{\tau}) < 0$ throughout the domain.

The Sverdrup balance (33.66) 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 (33.68a) 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 33.10: *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 33.6.4, a linear vorticity balance between beta, winds, and friction lead to a western boundary return flow and corresponding western intensification.*

33.6.4 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 our discussion of Figure 33.6 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 33.10, 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 (33.59) 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. What we have yet to provide an argument for intensification. For that purpose, recall that return flow region is

not in traditional Sverdrup balance. To see what terms can break that balance, consider again the steady vorticity balance (33.62). Continuing to assume a flat bottom and rigid lid leads to

$$\beta\, v = -\lambda\, \zeta + \hat{\boldsymbol{z}} \cdot (\nabla \wedge \boldsymbol{f}_{\text{wind}}), \tag{33.69}$$

where we dropped the nonlinear advection term $\boldsymbol{u} \cdot \nabla \zeta$ 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 (33.69) to be maintained in the northward return flow region? To answer this question, expose the signs on the terms in equation (33.69)

$$\underbrace{\beta\, v}_{\text{positive}} + \underbrace{\lambda\, \zeta}_{\text{unspecified}} = \underbrace{\hat{\boldsymbol{z}} \cdot (\nabla \wedge \boldsymbol{f}_{\text{wind}})}_{\text{negative}}. \tag{33.70}$$

We have $\beta\, v > 0$ since we are concerned with the region of northward return flow, and $\hat{\boldsymbol{z}} \cdot (\nabla \wedge \boldsymbol{f}_{\text{wind}}) < 0$ by assumption of anti-cyclonic wind stress over the full domain. Hence, for the balance (33.70) to be realized requires $\zeta < 0$, with the value large enough to balance both the winds and the meridional advection of planetary vorticity

$$\underbrace{\lambda\, \zeta}_{\text{negative}} = \underbrace{-\beta\, v}_{\text{negative}} + \underbrace{\hat{\boldsymbol{z}} \cdot (\nabla \wedge \boldsymbol{f}_{\text{wind}})}_{\text{negative}}. \tag{33.71}$$

For anti-cyclonic gyre flow, as required by the circulation condition (33.59), 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 have no interior region in Sverdrup balance driving southward flow. The flow would thus only be subject to the circulation condition (33.59) 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.

33.6.5 A role for topographic form stress

Recall the shallow water Sverdrup balance (33.65), here with the addition of Rayleigh drag

$$\beta\, v = Q_{\text{PG}}\, \boldsymbol{u} \cdot \nabla h - \lambda\, \zeta + \hat{\boldsymbol{z}} \cdot (\nabla \wedge \boldsymbol{f}_{\text{wind}}). \tag{33.72}$$

For the right hand side, we have thus far considered a flat bottom domain and ignored free surface undulations, in which case $\nabla h = 0$. We here sketch the situation with nontrivial topographic variations with a focus on how bottom pressure gradients can render meridional flow in the presence of topography.

Topographic form stress curl

We are just concerned with that portion of $\boldsymbol{u}\cdot\nabla h$ due to pressure gradients, in which case we insert the geostrophic flow

$$\begin{aligned}
\boldsymbol{u}\cdot\nabla h &= (g/f)\,\nabla h\cdot(\hat{\boldsymbol{z}}\wedge\nabla\eta) && (33.73a)\\
&= (g/f)\,\hat{\boldsymbol{z}}\cdot(\nabla\eta\wedge\nabla h) && (33.73b)\\
&= (g/f)\,\hat{\boldsymbol{z}}\cdot(\nabla\eta_{\mathrm{b}}\wedge\nabla h) && (33.73c)\\
&= 1/(\rho\,f)\,\hat{\boldsymbol{z}}\cdot(\nabla\eta_{\mathrm{b}}\wedge\nabla p_{\mathrm{b}}) && (33.73d)\\
&= -1/(\rho\,f)\,\hat{\boldsymbol{z}}\cdot(\nabla\wedge p_{\mathrm{b}}\nabla\eta_{\mathrm{b}}), && (33.73e)
\end{aligned}$$

where $\eta = h+\eta_{\mathrm{b}}$ according to Figure 29.1 and the bottom pressure $p_{\mathrm{b}} = g\,\rho\,h$. Making use of this result in the vorticity balance (33.72) leads to

$$\beta\,\rho\,h\,v = -\hat{\boldsymbol{z}}\cdot(\nabla\wedge p_{\mathrm{b}}\nabla\eta_{\mathrm{b}}) + \rho\,h\,[-\lambda\,\zeta + \hat{\boldsymbol{z}}\cdot(\nabla\wedge\boldsymbol{f}_{\mathrm{wind}})]\,. \tag{33.74}$$

The term $p_{\mathrm{b}}\nabla\eta_{\mathrm{b}}$ is the topographic form stress discussed in Section 20.2. We offer some special cases to see how this term affects the meridional flow in a gyre circulation.

Vanishing topographic form stress curl for geostrophic f-plane motion

In the absence of friction and wind forcing, β times the depth integrated meridional mass transport is balanced by the curl of the topographic form stress

$$\beta\,\rho\,h\,v = -\hat{\boldsymbol{z}}\cdot(\nabla\wedge p_{\mathrm{b}}\nabla\eta_{\mathrm{b}}) \qquad \text{linear inviscid and unforced.} \tag{33.75}$$

We conclude from this balance that a linear inviscid and unforced geostrophic flow on an f-plane satisfies $\hat{\boldsymbol{z}}\cdot(\nabla\wedge p_{\mathrm{b}}\nabla\eta_{\mathrm{b}}) = \hat{\boldsymbol{z}}\cdot(\nabla p_{\mathrm{b}}\wedge\nabla\eta_{\mathrm{b}}) = 0$. Hence, on an f-plane the isolines of surface height align with isolines of bottom topography since $p_{\mathrm{b}} = g\,\rho\,h = g\,\rho\,(\eta-\eta_{\mathrm{b}})$. We see this property again as part of Exercise 40.4.

Vanishing integrated topographic form stress curl

Consider a closed domain comprised of a topography with isobaths closed within the domain. Use of Stokes' theorem reveals that an integral of $\hat{\boldsymbol{z}}\cdot(\nabla\wedge p_{\mathrm{b}}\nabla\eta_{\mathrm{b}})$ over the area defined by an isobath vanishes

$$\int_{\mathcal{A}}\hat{\boldsymbol{z}}\cdot(\nabla\wedge p_{\mathrm{b}}\nabla\eta_{\mathrm{b}})\,\mathrm{d}A = \oint_{\partial\mathcal{A}} p_{\mathrm{b}}\nabla\eta_{\mathrm{b}}\cdot\mathrm{d}\boldsymbol{l} = 0, \tag{33.76}$$

where $\partial\mathcal{A}$ is determined by an isobath so that $\nabla\eta_{\mathrm{b}}\cdot\mathrm{d}\boldsymbol{l} = 0$. We can employ the same argument for an area bounded by isobars of bottom pressure. Hence, if there is no circulation, then topographic form stress will not spontaneously generate circulation. Correspondingly, topographic form stress is *not* an external force akin to wind stress, whereby wind stress curl is not constrained to vanish when integrated over a closed domain. Rather, the topographic form stress arises only in the presence of flow.

Local generation of topographic form stress curl

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. We discussed an analogous misalignment is Section 30.4 when studying the force balances in

a steady zonally re-entrant channel with bottom topography. Wind stress forcing and dissipation offer another means for $(\eta, \eta_{\rm b})$-misalignment, as seen merely by rewriting the vorticity balance (33.74) as an expression for the topographic form stress curl

$$\hat{\boldsymbol{z}} \cdot (\nabla \wedge p_{\rm b} \nabla \eta_{\rm b}) = -\beta \, \rho \, h \, v + \rho \, h \, [-\lambda \, \zeta + \hat{\boldsymbol{z}} \cdot (\nabla \wedge \boldsymbol{f}_{\rm wind})] \, . \tag{33.77}$$

In our discussion of western intensification in Section 33.6.4, we ignored the role of topographic form stress curls. However, as seen by this balance, topographic form stress curls can play in a role when topography and surface height are misaligned, with that role in some locations more important than friction (see *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 30.4, whereby topographic form stress, of sufficient strength, can balance the momentum input by the winds thus making friction less important for the channel flow.

33.6.6 Comments and further study

The study of ocean gyres originates from the pioneering work of Henry Stommel in the 1950s. 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 its many subsequent variants, such as that from Walter Munk who considered a viscous closure rather than Rayleigh drag. Our treatment here is very brief by comparison. Furthermore, the Stommel model and its variants are themselves very idealized renditions of the ocean gyres occuring in Nature. Notable further factors become important in studying Nature's gyres, such as topography (briefly discussed in Section 33.6.5), 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.

Following Stommel, we see that 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 that from Fofonoff most notable. We consider elements of the Fofonoff gyre in Exercise 33.10.

33.7 Exercises

EXERCISE 33.1: 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 33.11.

(a) What is the potential vorticity for the lower layer? Hint: recall the equations for a 1.5 layer model with a stagnant upper layer were derived in Exercise 29.5.

(b) Now assume an f-plane. If the lower layer potential vorticity is (somehow) horizonally 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 33.4.2 and further pursued in Section 40.3. Also, ignore any non-steady processes; we are only interested here in the steady flow.

(c) 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 33.11: *A reduced gravity model (see Section 29.4) with a stagnant upper layer and a dynamic lower layer as in Exercise 29.5. A seamount sits fully within the lower layer.*

EXERCISE 33.2: PV FOR TWO SHALLOW WATER LAYERS

Consider the inviscid Boussinesq two-layer shallow water model as discussed in Section 29.5. Derive the potential vorticity equation for each layer, showing the mathematical steps used in the derivation.

EXERCISE 33.3: AVERAGE VORTICITY IN A SHALLOW WATER LAYER

Consider a single layer of shallow water fluid on a rotating plane with rotation rate $\boldsymbol{\Omega} = \hat{\boldsymbol{z}}\,\Omega$. Assume the fluid is contained in an arbitrary horizontal region and that it has a constant total volume given by

$$\mathcal{V} = \int \mathrm{d}A \int \mathrm{d}z = \int h\,\mathrm{d}A = \int (H + \Delta\eta - \eta_b)\,\mathrm{d}A = H\,\mathcal{A}, \tag{33.78}$$

where $\mathcal{A}$ 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 29.1). Additionally, recall that $z = 0$ is set according to

$$\int \eta_b\,\mathrm{d}A = 0, \tag{33.79}$$

and volume conservation ensures that

$$\int \Delta\eta\,\mathrm{d}A = 0. \tag{33.80}$$

(a) Determine the volume average of the vorticity $\hat{\boldsymbol{z}} \cdot \boldsymbol{\omega}_{\text{solid}}$ arising from the solid-body rotation

$$\langle \hat{\boldsymbol{z}} \cdot \boldsymbol{\omega}_{\text{solid}} \rangle = \mathcal{V}^{-1} \int \hat{\boldsymbol{z}} \cdot \boldsymbol{\omega}_{\text{solid}}\,\mathrm{d}V. \tag{33.81}$$

(b) Determine the area average of the relative vorticity,

$$\overline{\zeta} = \mathcal{A}^{-1} \int \hat{\boldsymbol{z}} \cdot \boldsymbol{\omega} \, \mathrm{d}A, \tag{33.82}$$

in terms of the circulation around the boundary of the domain.

(c) Determine the volume average of the relative vorticity

$$\langle \zeta \rangle = \mathcal{V}^{-1} \int \hat{\boldsymbol{z}} \cdot \boldsymbol{\omega} \, \mathrm{d}V. \tag{33.83}$$

Write the expression in terms of the area average vorticity, $\overline{\zeta}$, the resting layer thickness, H, and the deviation of the surface height from resting, $\Delta\eta$.

EXERCISE 33.4: 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

$$\frac{\mathrm{D}}{\mathrm{D}t} \left[\frac{\zeta + f}{\eta - \eta_b} \right] = 0, \tag{33.84}$$

where η is the height of the free surface and η_b is the height of the bottom topography (see Figure 29.1). 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 <u>then</u> 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 33.5: 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 33.6: SOME FLOW PROPERTIES OF THE STEADY STATE SHALLOW WATER

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

$$h \, \boldsymbol{u} = \nabla \wedge (\hat{\boldsymbol{z}} \, \Psi). \tag{33.85}$$

(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

$$Q = Q(\Psi). \tag{33.86}$$

(d) Show that the Bernoulli function,

$$B = g\,\eta + \boldsymbol{u}\cdot\boldsymbol{u}/2 \tag{33.87}$$

is also a constant along the same streamlines; i.e.,

$$B = B(\Psi). \tag{33.88}$$

(e) What is the functional relation between the Bernoulli function and the potential vorticity?

EXERCISE 33.7: 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 33.12.

Figure 33.12: *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 30.1).*

(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 33.8: RAYLEIGH DRAG AND GALILEAN INVARIANCE

Recall the discussion of Galilean invariance from Section 12.6. Is the Rayleigh drag used in equation (33.57) Galilean invariant? Why? If not, then should that be of concern for its use in studying flow in a closed and bounded domain?

EXERCISE 33.9: GYRES IN THE PRESENCE OF CYCLONIC WINDS

Consider the ocean gyre discussion in Section 33.6. Rather than anti-cyclonic winds, now apply a cyclonic wind stress to the domain. As per the circulation theorem (33.59), the gyre flow will have a cyclonic sense. Will the resulting gyre exhibit eastern intensification or western? Appeal to whatever arguments you wish.

EXERCISE 33.10: ELEMENTS OF THE FOFONOFF GYRE

A Fofonoff gyre is an unforced inviscid solution in a flat bottom bounded and fully domain with a rigid lid. For this fluid the absolute vorticity is materially invariant, $\mathrm{D}(\zeta + f)/\mathrm{D}t = 0$. 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 33.13. 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 $\mathrm{D}(\zeta + f)/\mathrm{D}t = 0$. Hint: recall our discussion of Figure 33.5.

Figure 33.13: *A Fofonoff gyre is an unforced inviscid flow in a bounded domain where* $\mathrm{D}(\zeta + f)/\mathrm{D}t = 0$. *We here depict elements of this double-gyre flow in the northern hemisphere as part of Exercise 33.10.*

EXERCISE 33.11: SHALLOW WATER EQUATIONS WITH DIVERGENCE-DAMPING

When breaking the continous 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 is has zero horizontal divergence (see Section 26.3 or the 2d barotropic equation in Section 39.2). 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

$$\frac{\mathrm{D}\boldsymbol{u}}{\mathrm{D}t} + f\,\hat{\boldsymbol{z}} \wedge \boldsymbol{u} = -\nabla\left(g\,\eta + \alpha\,\Gamma\right) \tag{33.89a}$$

$$\frac{\mathrm{D}h}{\mathrm{D}t} = -h\,\nabla\cdot\boldsymbol{u}. \tag{33.89b}$$

The parameter $\alpha > 0$ is a constant and the field Γ is given by the Laplacian of the horizontal flow divergence

$$\Gamma = \nabla^2\mathscr{D}, \tag{33.90}$$

where

$$\mathscr{D} = \nabla\cdot\boldsymbol{u}. \tag{33.91}$$

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

$$\tilde{\eta} = \eta + \frac{\alpha\,\Gamma}{g}. \tag{33.92}$$

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 (33.13) remains unchanged in the presence of divergence-damping.

(b) Show that the potential vorticity equation (33.25) remains unchanged in the presence of divergence-damping.

(c) Show that the horizontal divergence evolves according to

$$\frac{\partial\mathscr{D}}{\partial t} = \left[\frac{\partial\mathscr{D}}{\partial t}\right]_{\alpha=0} - \alpha\nabla^2\Gamma. \tag{33.93}$$

(d) Show that the evolution of gravitational potential energy per horizontal area

$$\mathscr{P} = g\,\rho\int_{\eta_\mathrm{b}}^{\eta} z\,\mathrm{d}z \tag{33.94}$$

remains unchanged from that determined in Section 30.6.1.

(e) Show that the kinetic energy per horizontal area evolves according to

$$\frac{\partial \mathcal{K}}{\partial t} + \nabla \cdot (\boldsymbol{u}\, \mathcal{K}) = -h \rho\, g\, \nabla \tilde{\eta}, \tag{33.95}$$

where

$$\mathcal{K} = \frac{1}{2} \int_{\eta_{\mathrm{b}}}^{\eta} \rho\, \mathbf{u}^2 \, \mathrm{d}z = \rho\, h\, \mathbf{u}^2/2, \tag{33.96}$$

is the horizontal kinetic energy per area (Section 30.6.2).

(f) Determine the evolution equation for global integrated kinetic energy

$$\frac{\partial}{\partial t} \left[\int \mathcal{K} \, \mathrm{d}A \right] = \frac{\partial}{\partial t} \left[\int \int_{\eta_{\mathrm{b}}}^{\eta} (\rho\, \boldsymbol{u}^2/2) \, \mathrm{d}z \, \mathrm{d}A \right]. \tag{33.97}$$

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 30.5. Show that the material evolution of angular momentum relative to the vertical rotational axis is given by

$$\frac{1}{\delta M} \frac{\mathrm{D}L^z}{\mathrm{D}t} = -g \frac{\partial \eta}{\partial \phi} + \mathcal{T}. \tag{33.98}$$

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

$$\frac{\partial}{\partial t} \int L^z = \alpha\, \rho \int \Gamma \frac{\partial \eta}{\partial \phi} \, \mathrm{d}A. \tag{33.99}$$

where we assume the bottom topography is flat so that $h = \eta$.

(i) The linearized thickness equation (see Section 31.3) for a flat bottom is given by

$$\frac{\partial \eta}{\partial t} + H\, \nabla \cdot \boldsymbol{u} = 0, \tag{33.100}$$

where H is the thickness of the resting fluid layer. Show that the time change for the global integrated angular momentum is given by

$$\frac{\partial}{\partial t} \int L^z = -\frac{\alpha\, \rho}{H} \int \left[\frac{\partial}{\partial t} \nabla^2 \eta \right] \frac{\partial \eta}{\partial \phi} \, \mathrm{d}A. \tag{33.101}$$

34

Vorticity mechanics

This chapter develops the basic kinematics and dynamics of vorticity, building from the shallow water discussion of Chapter 33 to here consider fully three-dimensional flows. Doing so provides us with a mathematical and physical framework to understand how physical processes affect vorticity in geophysical fluids.

READER'S GUIDE FOR THIS CHAPTER

This chapter requires notions from earlier chapters in this part of the book as well as the fluid kinematics from Part III and fluid dynamics from Part IV. As for the shallow water vorticity discussed in Chapter 33, 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, much of which is encountered in the remainder of this part of the book as well as for the balanced models considered in Part VII.

34.1 Kinematics of vortex lines and vortex tubes **612**
34.1.1 Vortex lines and vortex tubes 612
34.1.2 Kinematic properties 612
34.1.3 Helmholz's theorems 613
34.1.4 Further study 615
34.2 Kelvin's Circulation Theorem **615**
34.2.1 Formulation 615
34.2.2 Barotropic flow 616
34.2.3 Pressure gradient force for barotropic fluids 617
34.3 Vorticity dynamics **617**
34.3.1 Vector-invariant velocity equation 617
34.3.2 Basic form of the vorticity equation 618
34.3.3 Massaged form of the vorticity equation 618
34.3.4 Vorticity, angular momentum, and torques 619
34.4 Mechanics of baroclinicity **619**
34.4.1 Curl of the pressure gradient body force 619
34.4.2 Kelvin's Theorem and contact pressure forces 620
34.4.3 Bottom pressure contributions at the solid-earth boundary 621
34.5 Vorticity filaments and material line elements **622**
34.5.1 Frozen-in nature of vorticity 622
34.5.2 Stretching and tilting of vortex tubes 623
34.6 Circulation and vorticity for rotating fluids **625**

34.6.1 Material evolution of absolute circulation 626
34.6.2 The beta effect 627
34.6.3 A two-dimensional incompressible fluid 629
34.7 Vorticity in a hydrostatic-Boussinesq fluid **630**
34.7.1 Deriving the vorticity equation 630
34.7.2 Boussinesq form of the baroclinicity vector 632
34.7.3 Vertical vorticity equation 632
34.7.4 Depth integral of the vertical vorticity equation 633
34.8 Exercises **635**

34.1 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 13.2.3.

34.1.1 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)

$$\boldsymbol{x}(\varphi) = x(\varphi)\,\hat{\boldsymbol{x}} + y(\varphi)\,\hat{\boldsymbol{y}} + z(\varphi)\,\hat{\boldsymbol{z}}. \tag{34.1}$$

The three coordinates of the line are constrained so that the line is tangent to vorticity at each point, which means

$$\frac{\mathrm{d}x/\mathrm{d}\varphi}{\omega_x} = \frac{\mathrm{d}y/\mathrm{d}\varphi}{\omega_y} = \frac{\mathrm{d}z/\mathrm{d}\varphi}{\omega_z}. \tag{34.2}$$

These equations are directly analogous to those satisfied by velocity streamlines (Section 12.8.2)

$$\frac{\mathrm{d}x/\mathrm{d}\varsigma}{u} = \frac{\mathrm{d}y/\mathrm{d}\varsigma}{v} = \frac{\mathrm{d}z/\mathrm{d}\varsigma}{w}, \tag{34.3}$$

where ς 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 12.8). 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 34.1 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 12.5.

34.1.2 Kinematic properties

Vorticity has zero divergence

$$\nabla \cdot \boldsymbol{\omega} = \nabla \cdot (\nabla \wedge \boldsymbol{v}) = 0, \tag{34.4}$$

which follows since the divergence of a curl vanishes. Integrating the non-divergence relation over an arbitrary closed volume within the fluid leads to

$$\int_{\mathcal{R}} \nabla \cdot \boldsymbol{\omega}\,\mathrm{d}V = \oint_{\partial\mathcal{R}} \boldsymbol{\omega} \cdot \hat{\boldsymbol{n}}\,\mathrm{d}\mathcal{S} = 0, \tag{34.5}$$

Figure 34.1: *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* $\mathcal{C} = \oint_{\partial\mathcal{S}} \boldsymbol{v}\cdot\mathrm{d}\boldsymbol{r} = \int \boldsymbol{\omega}\cdot\hat{\boldsymbol{n}}\,\mathrm{d}\mathcal{S}$. *Since vorticity has zero divergence, the circulation is the same for any loop around the vortex tube (Helmholz's first theorem from Section 34.1.3). 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.*

where we made use of Gauss's divergence theorem to reach the surface integral expression, with $\hat{\boldsymbol{n}}\,\mathrm{d}\mathcal{S}$ the oriented area element on the boundary of the volume, $\partial\mathcal{R}$, and $\hat{\boldsymbol{n}}$ 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 (34.5) to a volume along a chosen vortex tube such as in Figure 34.1.[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

$$\int_{\mathrm{A}} \boldsymbol{\omega}\cdot\hat{\boldsymbol{n}}\,\mathrm{d}\mathcal{S}_A + \int_{\mathrm{B}} \boldsymbol{\omega}\cdot\hat{\boldsymbol{n}}\,\mathrm{d}\mathcal{S}_B = 0. \tag{34.6}$$

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

$$\oint_{\mathrm{A}} \boldsymbol{v}\cdot\mathrm{d}\boldsymbol{r} + \oint_{\mathrm{B}} \boldsymbol{v}\cdot\mathrm{d}\boldsymbol{r} = 0. \tag{34.7}$$

We thus see that the circulation around the vortex tube is the same no matter where it is computed. The circulation constraints (34.6) and (34.7) are kinematic, holding for any vorticity field. We now consider some consequences of this constraint.

34.1.3 Helmholz's theorems

There are a few basic properties of vorticity that follow from its vanishing divergence. These properties are known as Helmholz's theorems.

Helmholz's first theorem

Since the cross-sectional slices used to derive the circulation constraint (34.7) are arbitrary, the constraint holds throughout the full extent of the vortex tube. Hence, as noted following equation

[1] In Exercise 16.8, we developed a similar set of results for a streamtube in an incompressible fluid.

(34.7), 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 34.1). This result is known as Helmholz's first theorem.

Figure 34.2: *A vortex tube idealization of a tornado. Since the circulation around the tube is uniform (Helmholz'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.*

As a corollary, we refer to the vorticity constraint (34.6) 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 34.2, 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.

Helmholz's second theorem

The vorticity constraint (34.6) 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 34.3), or intersect a boundary (as for a tornado in Figure 34.2, where the ground and clouds form the boundary).

Helmholz's third theorem

Helmholz'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 34.2.

Figure 34.3: *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 as shown 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.*

34.1.4 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 demostrations of vorticity in non-rotating and rotating flows.

- Helmholz's theorems are vividly exhibited by this video from the *Physics Girl* of flow generated by a paddle in a swimming pool.
- A rotating tank experiment shown near the 15 minute mark of this video from Prof. Dave Fultz of the University of Chicago shows how vorticity is affected by vortex stretching.
- This video offers a classic tutorial on vorticity in non-rotating fluids from Prof. A. Shapiro of MIT.

34.2 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

$$\frac{\mathrm{d}\mathcal{C}}{\mathrm{d}t} = \frac{\mathrm{d}}{\mathrm{d}t}\oint_{\partial\mathcal{S}(\boldsymbol{v})} \boldsymbol{v}\cdot\mathrm{d}\boldsymbol{r} = \frac{\mathrm{d}}{\mathrm{d}t}\int_{\mathcal{S}(\boldsymbol{v})} \boldsymbol{\omega}\cdot\hat{\boldsymbol{n}}\,\mathrm{d}\mathcal{S}, \tag{34.8}$$

where $\mathcal{S}(\boldsymbol{v})$ designates a surface whose points all move with the fluid flow. We here consider non-rotating flow, with the non-rotating flow directly applicable to the absolute circulation (relative circulation plus circulation due to planetary rotation) found for the rotating case in Section 34.6.

34.2.1 Formulation

We are here concerned with circulation for material loops. Hence, the time derivative acting on the circulation integral in equation (34.8) moves inside the integral as a material time operator

$$\frac{\mathrm{d}\mathcal{C}}{\mathrm{d}t} = \frac{\mathrm{d}}{\mathrm{d}t}\oint_{\partial\mathcal{S}(\boldsymbol{v})} \boldsymbol{v}\cdot\mathrm{d}\boldsymbol{r} = \oint_{\partial\mathcal{S}(\boldsymbol{v})} \frac{\mathrm{D}(\boldsymbol{v}\cdot\mathrm{d}\boldsymbol{r})}{\mathrm{D}t}. \tag{34.9}$$

The material evolution of $\boldsymbol{v}$ is determined by Newton's Law of motion, which for a non-rotating flow is given by (see Section 23.6)

$$\frac{\mathrm{D}\boldsymbol{v}}{\mathrm{D}t} = -\frac{1}{\rho}\nabla p - \nabla\Phi + \boldsymbol{F}. \tag{34.10}$$

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 $\boldsymbol{F}$ is a force arising from non-conservative viscous stresses.

The material time derivative of the differential line element moving around the circuit equals to the differential of the velocity on the circuit

$$\frac{\mathrm{D}(\mathrm{d}\boldsymbol{r})}{\mathrm{D}t} = \mathrm{d}\boldsymbol{v}. \tag{34.11}$$

This result follows since all points along the circuit are defined by fluid particles. Consequently, evolution of circulation following a material loop becomes

$$\frac{\mathrm{d}\mathcal{C}}{\mathrm{d}t} = \frac{\mathrm{d}}{\mathrm{d}t}\oint_{\partial\mathcal{S}(\boldsymbol{v})} \boldsymbol{v}\cdot\mathrm{d}\boldsymbol{r} \tag{34.12a}$$

$$= \oint_{\partial\mathcal{S}(\boldsymbol{v})} \left[\left(-\frac{1}{\rho}\nabla p - \nabla\Phi + \boldsymbol{F}\right)\cdot\mathrm{d}\boldsymbol{r} + \frac{1}{2}\mathrm{d}\boldsymbol{v}^2\right] \tag{34.12b}$$

$$= \oint_{\partial\mathcal{S}(\boldsymbol{v})} \left[-\frac{1}{\rho}\nabla p + \boldsymbol{F}\right]\cdot\mathrm{d}\boldsymbol{r} \tag{34.12c}$$

$$= \int_{\mathcal{S}(\boldsymbol{v})} \left[-\nabla\wedge\left(\frac{1}{\rho}\nabla p\right) + \nabla\wedge\boldsymbol{F}\right]\cdot\hat{\boldsymbol{n}}\,\mathrm{d}\mathcal{S} \tag{34.12d}$$

$$= \int_{\mathcal{S}(\boldsymbol{v})} (\boldsymbol{B} + \nabla\wedge\boldsymbol{F})\cdot\hat{\boldsymbol{n}}\,\mathrm{d}\mathcal{S}. \tag{34.12e}$$

The third equality follows since $\nabla\Phi\cdot\mathrm{d}\boldsymbol{r} = \mathrm{d}\Phi$ has zero integral around a closed circuit, as does $\mathrm{d}\boldsymbol{v}^2$. The fourth equality made use of Stokes' theorem. The final equality introduced the solenoidal vector

$$\boldsymbol{B} = \frac{\nabla\rho\wedge\nabla p}{\rho^2}, \tag{34.13}$$

which we also refer to as the *baroclinicity vector*. The baroclinicity vector has physical dimensions of inverse squared time, T^{-2}. Equation (34.12e) 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.

34.2.2 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

$$p = p(\rho) \Rightarrow \text{barotropic flow}. \tag{34.14}$$

Kelvin's theorem then follows, which states that for inviscid barotropic flow the circulation around any closed material circuit remains constant

$$\frac{\mathrm{d}\mathcal{C}}{\mathrm{d}t} = \frac{\mathrm{d}}{\mathrm{d}t}\oint_{\partial\mathcal{S}(\boldsymbol{v})} \boldsymbol{v}\cdot\mathrm{d}\boldsymbol{r} = \frac{\mathrm{d}}{\mathrm{d}t}\int_{\mathcal{S}(\boldsymbol{v})} \boldsymbol{\omega}\cdot\hat{\boldsymbol{n}}\,\mathrm{d}\mathcal{S} = 0 \qquad \Longleftarrow \text{inviscid barotropic flow}. \tag{34.15}$$

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.

34.2.3 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

$$\nabla \wedge (\rho^{-1} \nabla p) = 0 \tag{34.16}$$

and there is a corresponding scalar potential for the pressure gradient force

$$\nabla \Phi_{\mathrm{p}} = \rho^{-1} \nabla p. \tag{34.17}$$

Integration renders (see page 79 of *Müller* (2006))

$$\Phi_{\mathrm{p}} = \int_{p_0}^{p} \frac{\mathrm{d}p'}{\rho(p')}, \tag{34.18}$$

where p_0 is an arbitrary reference pressure. We make use of the scalar potential Φ_{p} in Exercise 34.5.

34.3 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 23.6)

$$\rho \left[\frac{\mathrm{D}}{\mathrm{D}t} + 2\, \boldsymbol{\Omega} \wedge \right] \boldsymbol{v} = -\nabla p - \rho \nabla \Phi + \rho \boldsymbol{F}, \tag{34.19}$$

where $\boldsymbol{\Omega}$ is the angular velocity of the rotating reference frame.

34.3.1 Vector-invariant velocity equation

As for the shallow water fluid in Section 33.1, 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)

$$\boldsymbol{\omega} \wedge \boldsymbol{v} = -(1/2)\, \nabla(\boldsymbol{v} \cdot \boldsymbol{v}) + (\boldsymbol{v} \cdot \nabla)\, \boldsymbol{v} \tag{34.20}$$

to eliminate velocity self-advection in favor of vorticity

$$\frac{\partial \boldsymbol{v}}{\partial t} + \boldsymbol{\omega}_{\mathrm{a}} \wedge \boldsymbol{v} = -\frac{1}{\rho} \nabla p - \nabla \left[\frac{1}{2}\, \boldsymbol{v}^2 + \Phi \right] + \boldsymbol{F}. \tag{34.21}$$

We here introduced the absolute vorticity

$$\boldsymbol{\omega}_{\mathrm{a}} = \boldsymbol{\omega} + 2\, \boldsymbol{\Omega}, \tag{34.22}$$

which is the sum of the relative vorticity and the planetary vorticity (see Section 32.5.1).

34.3.2 Basic form of the vorticity equation

Taking the curl of the vector-invariant momentum equation (34.21) removes the mechanical energy per mass, $\boldsymbol{v}^2/2 + \Phi$, thus leaving

$$\frac{\partial \boldsymbol{\omega}}{\partial t} + \nabla \wedge (\boldsymbol{\omega}_{\mathrm{a}} \wedge \boldsymbol{v}) = \frac{1}{\rho^2}(\nabla\rho \wedge \nabla p) + \nabla \wedge \boldsymbol{F}. \tag{34.23}$$

For geophysical fluids we generally assume that $\boldsymbol{\Omega}$ has zero time tendency, so that

$$\frac{\partial \boldsymbol{\omega}_{\mathrm{a}}}{\partial t} = \frac{\partial\,(\boldsymbol{\omega} + 2\,\boldsymbol{\Omega})}{\partial t} = \frac{\partial \boldsymbol{\omega}}{\partial t}, \tag{34.24}$$

in which case equation (34.23) can be written

$$\frac{\partial \boldsymbol{\omega}_{\mathrm{a}}}{\partial t} + \nabla \wedge (\boldsymbol{\omega}_{\mathrm{a}} \wedge \boldsymbol{v}) = \boldsymbol{B} + \nabla \wedge \boldsymbol{F}, \tag{34.25}$$

where $\boldsymbol{B}$ is the baroclinicity vector introduced by equation (34.13).

34.3.3 Massaged form of the vorticity equation

Physical interpretation of the term $\nabla \wedge (\boldsymbol{\omega}_{\mathrm{a}} \wedge \boldsymbol{v})$ appearing in the prognostic equation (34.25) can be made more transparent by using yet another vector identity

$$\nabla \wedge (\boldsymbol{\omega}_{\mathrm{a}} \wedge \boldsymbol{v}) = (\boldsymbol{v}\cdot\nabla)\boldsymbol{\omega}_{\mathrm{a}} - (\boldsymbol{\omega}_{\mathrm{a}}\cdot\nabla)\boldsymbol{v} + \boldsymbol{\omega}_{\mathrm{a}}\,\nabla\cdot\boldsymbol{v} - \boldsymbol{v}\,\nabla\cdot\boldsymbol{\omega}_{\mathrm{a}} \tag{34.26a}$$

$$= (\boldsymbol{v}\cdot\nabla)\boldsymbol{\omega}_{\mathrm{a}} - (\boldsymbol{\omega}_{\mathrm{a}}\cdot\nabla)\boldsymbol{v} - \frac{\boldsymbol{\omega}_{\mathrm{a}}}{\rho}\frac{\mathrm{D}\rho}{\mathrm{D}t}. \tag{34.26b}$$

The second equality required the continuity equation

$$\frac{\mathrm{D}\rho}{\mathrm{D}t} = -\rho\,\nabla\cdot\boldsymbol{v}, \tag{34.27}$$

and the non-divergent nature of the absolute vorticity

$$\nabla\cdot\boldsymbol{\omega}_{\mathrm{a}} = \nabla\cdot(\nabla \wedge \boldsymbol{v} + 2\,\boldsymbol{\Omega}) = 0. \tag{34.28}$$

Equation (34.25) thus takes the form

$$\frac{\mathrm{D}\boldsymbol{\omega}_{\mathrm{a}}}{\mathrm{D}t} - \frac{\boldsymbol{\omega}_{\mathrm{a}}}{\rho}\frac{\mathrm{D}\rho}{\mathrm{D}t} = (\boldsymbol{\omega}_{\mathrm{a}}\cdot\nabla)\boldsymbol{v} + \frac{1}{\rho^2}(\nabla\rho \wedge \nabla p) + \nabla \wedge \boldsymbol{F}, \tag{34.29}$$

which can be written in vector form as

$$\rho\,\frac{\mathrm{D}(\boldsymbol{\omega}_{\mathrm{a}}/\rho)}{\mathrm{D}t} = (\boldsymbol{\omega}_{\mathrm{a}}\cdot\nabla)\,\boldsymbol{v} + \boldsymbol{B} + \nabla \wedge \boldsymbol{F}. \tag{34.30}$$

Equation (34.30) 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, $(\boldsymbol{\omega}_{\mathrm{a}}\cdot\nabla)\,\boldsymbol{v}$, will be explored in Section 34.5 in the simplified context of a barotropic fluid. The second term arises from baroclinicity as introduced in equation (34.13) and given a mechanical interpretation in Section 34.4. The third term arises from the curl of the friction vector, which contributes especially in boundary layer regions where friction is relatively large.

34.3.4 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 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 crosss 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 derivative operations. In Chapter 35, 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 34.4 we explore baroclinicity, which is the key mechanism for how inviscid torques from pressure modify vorticity. 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. Hence, when there is baroclinicity, the pressure force spins the fluid element. Analogous inviscid and viscous forces 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.

34.4 Mechanics of baroclinicity

Baroclinicity is present in all realistic geophysical flows, thus affecting the material evolution of circulation and vorticity. Flow with a nonzero baroclinicity vector is generally referred to as *baroclinic flow*. We here further explore the mechanical features of baroclinicity to help further its understanding.

34.4.1 Curl of the pressure gradient body force

Baroclinicity is the curl of the pressure gradient body force

$$\boldsymbol{B} = \nabla \wedge \boldsymbol{F}_{\text{press}} = \nabla \wedge (-\nabla p/\rho) = -\nabla \rho^{-1} \wedge \nabla p = \frac{\nabla \rho \wedge \nabla p}{\rho^2}. \tag{34.31}$$

As discussed in Section 34.3.4, the curl of a force provides a torque that spins the fluid, thus rendering a vorticity source. Geometrically we find it useful to introduce the notion of a solenoid, which is a tube region in the fluid that is perpendicular to both $\nabla\rho$ and ∇p. Solenoids vanish for barotropic flows, whereby $p = p(\rho)$ (see equation (34.14)). 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

$$\rho\,\boldsymbol{B} = (-\rho^{-1}\,\nabla p)\wedge\nabla\rho. \tag{34.32}$$

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 34.4. 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 shallow water layer) does the pressure force pass through the center of mass, thus creating no torque and not affecting vorticity.

Figure 34.4: *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).*

34.4.2 Kelvin's 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 34.2.1. Focusing just on the baroclinicity contribution in equation (34.12e) we have

$$\left[\frac{\mathrm{d}\mathcal{C}}{\mathrm{d}t}\right]_{\text{baroclinicity}} = \oint_{\partial\mathcal{S}(\boldsymbol{v})} -\rho^{-1}\,\nabla p\cdot\mathrm{d}\boldsymbol{r} = \oint_{\partial\mathcal{S}(\boldsymbol{v})} \rho^{-2}\,(-p\,\nabla\rho)\cdot\mathrm{d}\boldsymbol{r}. \tag{34.33}$$

The term $-p\,\nabla\rho$ 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 34.5 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 34.5).

Figure 34.5: *The material circuit on the left is assumed to be within a constant density surface. In this case,* $\nabla\rho\,\mathrm{d}\boldsymbol{r} = 0$ *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.*

34.4.3 Bottom pressure contributions at the solid-earth boundary

Now consider a fluid region that intersects the solid-earth boundary. This boundary is material so that we can apply Kelvin's Circulation Theorem to a circuit within this boundary. Consider the situation in Figure 34.6, which shows a vertical slice next to a sloping bottom with constant density surface intersecting the bottom. As in our considerations in Section 34.4.2, any material circuit that sits within the bottom boundary will cross 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 withing 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

$$\mathrm{d}\boldsymbol{r} = \hat{\boldsymbol{x}}\,\mathrm{d}x + \hat{\boldsymbol{y}}\,\mathrm{d}y + \hat{\boldsymbol{z}}\,\mathrm{d}z = (\hat{\boldsymbol{x}} - \hat{\boldsymbol{z}}\,\partial_x H)\mathrm{d}x + (\hat{\boldsymbol{y}} - \hat{\boldsymbol{z}}\,\partial_y H)\mathrm{d}y. \tag{34.34}$$

To reach this result we set $z = -H(x, y)$ since the circuit is along the bottom boundary, which in turn means that

$$\mathrm{d}z = -\mathrm{d}H = -\nabla_z H \cdot (\hat{\boldsymbol{x}}\,\mathrm{d}x + \hat{\boldsymbol{y}}\,\mathrm{d}y). \tag{34.35}$$

Consequently, the projection of the density gradient onto the circuit is given by

$$\nabla\rho \cdot \mathrm{d}\boldsymbol{r} = (\nabla_z\rho - \partial_z\rho\,\nabla_z H) \cdot \mathrm{d}\boldsymbol{x}. \tag{34.36}$$

Making use of this result in Kelvin's Circulation Theorem and focusing on the pressure contribution, as in equation (34.33), leads to

$$\left[\frac{\mathrm{d}\mathcal{C}}{\mathrm{d}t}\right]_{\text{bottom}} = -\oint_{\partial\mathcal{S}_{\text{bottom}}} \rho^{-1}\,\nabla p \cdot \mathrm{d}\boldsymbol{r} = \oint_{\partial\mathcal{S}_{\text{bottom}}} \frac{p_{\text{b}}}{\rho^2}(-\nabla_z\rho + \partial_z\rho\,\nabla_z H) \cdot \mathrm{d}\boldsymbol{x}. \tag{34.37}$$

There are two contributions to the circulation changes revealed by equation (34.37). 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 34.6: *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.*

34.5 Vorticity filaments and material line elements

To help develop an understanding for the term $(\boldsymbol{\omega}_{\mathrm{a}} \cdot \nabla)\, \boldsymbol{v}$ appearing in the vorticity equation (34.30), consider the special case of an incompressible, inviscid, non-rotating barotropic fluid, in which case the vorticity equation (34.30) reduces to

$$\frac{\mathrm{D}\boldsymbol{\omega}}{\mathrm{D}t} = (\boldsymbol{\omega} \cdot \nabla)\, \boldsymbol{v}. \tag{34.38}$$

Evolution of vorticity due to the source term $(\boldsymbol{\omega} \cdot \nabla)\, \boldsymbol{v}$ is kinematically identical to evolution of a material line element detailed in Section 13.2. We pursue this connection in the following.

34.5.1 Frozen-in nature of vorticity

Recall from Section 34.1.1 that a vortex line is a line drawn through the fluid that is everywhere parallel to the vorticity. Such a line connects material fluid elements, 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

$$\delta \boldsymbol{x}(0) = \Gamma\, \boldsymbol{\omega}(\boldsymbol{x}, 0), \tag{34.39}$$

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 (34.38) has precisely the same mathematical form as the material line element equation (13.19)

$$\frac{\mathrm{D}(\delta \boldsymbol{x})}{\mathrm{D}t} = (\delta \boldsymbol{x} \cdot \nabla)\, \boldsymbol{v}. \tag{34.40}$$

Consequently, the relation (34.39) 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 $\delta \boldsymbol{x} = \Gamma \boldsymbol{\omega}$. In this way, the vorticity is attached to the particular material line element, motivating one to say that vorticity is a "frozen-in" property. This property is illustrated in Figure 34.7. Again, this propety holds only for the case of an incompressible, inviscid, barotropic fluid. Nonetheless, it offers great insight into the more general situation occuring in real fluids.

Figure 34.7: *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 34.5.2. 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.*

34.5.2 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 barotropic incompressible vorticity equation (34.38) and focus just the vertical vorticity component

$$\frac{\mathrm{D}\omega^z}{\mathrm{D}t} = \omega^x \frac{\partial w}{\partial x} + \omega^y \left(\frac{\partial w}{\partial y}\right) + \omega^z \left(\frac{\partial w}{\partial z}\right). \tag{34.41}$$

Note that since $\nabla \cdot \boldsymbol{\omega} = 0$, we can write this equation as

$$\frac{\mathrm{D}\omega^z}{\mathrm{D}t} = \nabla \cdot (w\, \boldsymbol{\omega}), \tag{34.42}$$

though in the following we find it more useful to focus on the form given by equation (34.41). The following discussion closely emulates that given for a material line element in Section 13.2.5.

Stretching

Consider the vortex tube to be initially aligned with the (vertical) z-axis, so that $\omega^x = \omega^y = 0$, in which case there is only a single term impacting vertical vorticity[2]

$$\frac{\mathrm{D}\omega^z}{\mathrm{D}t} = \omega^z \frac{\partial w}{\partial z}. \tag{34.43}$$

Since the fluid is incompressible, the volume of an infinitesimal portion of the vortex tube is materially constant

$$\frac{\mathrm{D}(\delta V)}{\mathrm{D}t} = 0, \tag{34.44}$$

[2]Be mindful to distinguish the symbols for the vertical component of vorticity, ω^z, and the vertical component of velocity, w.

which means that the vertical extent, δz, and cross-sectional area, δA, are constrained

$$\frac{1}{\delta z}\frac{\mathrm{D}(\delta z)}{\mathrm{D}t} + \frac{1}{\delta A}\frac{\mathrm{D}(\delta A)}{\mathrm{D}t} = 0. \tag{34.45}$$

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 (34.40)) allows us to write

$$\frac{1}{\delta z}\frac{\mathrm{D}(\delta z)}{\mathrm{D}t} = \frac{\partial w}{\partial z}, \tag{34.46}$$

so that the vorticity equation (34.43) becomes

$$\frac{\mathrm{D}\omega^z}{\mathrm{D}t} = \omega^z\,\frac{\partial w}{\partial z} \tag{34.47a}$$

$$= \omega^z\left[\frac{1}{\delta z}\frac{\mathrm{D}(\delta z)}{\mathrm{D}t}\right] \tag{34.47b}$$

$$= -\omega^z\left[\frac{1}{\delta A}\frac{\mathrm{D}(\delta A)}{\mathrm{D}t}\right]. \tag{34.47c}$$

Rearrangement leads to

$$\frac{\mathrm{D}(\omega^z\,\delta A)}{\mathrm{D}t} = 0, \tag{34.48}$$

which is an expression of Kelvin's circulation theorem (equation (34.15)) 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 (34.41) and illustrated in Figure 34.8. 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 Helmholz's first theorem discussed in Section 34.1.3). 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 33.2.2 and depicted by Figure 33.1.

Tilting

Now consider an initially horizontal vortex tube as in the lower left panel of Figure 34.8 so that $\omega^z = 0$. Focus on just one of the two horizontal directions, so that equation (34.41) for the vertical vorticity becomes

$$\frac{\mathrm{D}\omega^z}{\mathrm{D}t} = \omega^x\,\frac{\partial w}{\partial x}. \tag{34.49}$$

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 aligned parallel to the x-axis, the evolution of a material line segment (equation (34.40)) is given by

$$\frac{\mathrm{D}(\delta\boldsymbol{x})}{\mathrm{D}t} = \delta x\,\frac{\partial \boldsymbol{v}}{\partial x}. \tag{34.50}$$

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.

Figure 34.8: *Ilustrating 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 33.2.2 and depicted by Figure 33.1, as well as with Helmholz's first theorem in Section 34.1.3 and Figure 34.2. 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 (i.e., $\partial w/\partial x \neq 0$) 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.*

34.6 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 9.9.1 for the inertial or absolute velocity (i.e., velocity measured in an inertial frame)

$$\boldsymbol{v}_{\mathrm{a}} = \boldsymbol{v} + \boldsymbol{\Omega} \wedge \boldsymbol{x}, \tag{34.51}$$

where $\boldsymbol{v}$ is the velocity measured in the rotating frame (relative velocity), and $\boldsymbol{x}$ 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

$$\mathcal{C}_{\mathrm{a}} = \oint_{\partial\mathcal{S}} (\boldsymbol{v} + \boldsymbol{\Omega} \wedge \boldsymbol{x}) \cdot \mathrm{d}\boldsymbol{r} = \mathcal{C} + \mathcal{C}_{\mathrm{planet}}, \tag{34.52}$$

where the circulation measured in the rotating frame is

$$\mathcal{C} = \oint_{\partial\mathcal{S}} \boldsymbol{v} \cdot \mathrm{d}\boldsymbol{r} \tag{34.53}$$

and the circulation associated with the rotating planet is

$$\mathcal{C}_{\mathrm{planet}} = \oint_{\partial\mathcal{S}} (\boldsymbol{\Omega} \wedge \boldsymbol{x}) \cdot \mathrm{d}\boldsymbol{r}. \tag{34.54}$$

Again, $\mathrm{d}\boldsymbol{r}$ is the differential line element moving around the circuit as mentioned in Section 32.3.

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

$$\mathcal{C} = \oint_{\partial\mathcal{S}} \boldsymbol{v}\cdot \mathrm{d}\boldsymbol{r} = \int_{\mathcal{S}} \boldsymbol{\omega}\cdot\hat{\boldsymbol{n}}\,\mathrm{d}\mathcal{S} \qquad \text{relative circulation} \tag{34.55a}$$

$$\mathcal{C}_{\text{planet}} = \oint_{\partial\mathcal{S}} (\boldsymbol{\Omega}\wedge\boldsymbol{x})\cdot \mathrm{d}\boldsymbol{r} = \int_{\mathcal{S}} \boldsymbol{\omega}_{\text{planet}}\cdot\hat{\boldsymbol{n}}\,\mathrm{d}\mathcal{S} \qquad \text{planetary circulation} \tag{34.55b}$$

$$\mathcal{C}_{\text{a}} = \oint_{\partial\mathcal{S}} \boldsymbol{v}_{\text{a}}\cdot \mathrm{d}\boldsymbol{r} = \int_{\mathcal{S}} \boldsymbol{\omega}_{\text{a}}\cdot\hat{\boldsymbol{n}}\,\mathrm{d}\mathcal{S} \qquad \text{absolute circulation,} \tag{34.55c}$$

where

$$\boldsymbol{\omega} = \nabla\wedge\boldsymbol{v} \qquad \text{relative vorticity} \tag{34.56a}$$

$$\boldsymbol{\omega}_{\text{planet}} = \nabla\wedge(\boldsymbol{\Omega}\wedge\boldsymbol{x}) = 2\,\boldsymbol{\Omega} \qquad \text{planetary vorticity} \tag{34.56b}$$

$$\boldsymbol{\omega}_{\text{a}} = \boldsymbol{\omega} + \boldsymbol{\omega}_{\text{planet}} \qquad \text{absolute vorticity.} \tag{34.56c}$$

Thus far we have merely substituted in the expression (34.51) 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.

34.6.1 Material evolution of absolute circulation

We now consider how the absolute circulation evolves for a material circuit that moves with the fluid

$$\frac{\mathrm{d}\mathcal{C}_{\text{a}}}{\mathrm{d}t} = \frac{\mathrm{d}}{\mathrm{d}t}\oint_{\partial\mathcal{S}(\boldsymbol{v})} (\boldsymbol{v}+\boldsymbol{\Omega}\wedge\boldsymbol{r})\cdot\mathrm{d}\boldsymbol{r}, \tag{34.57}$$

where

$$\boldsymbol{r} = \boldsymbol{x} \tag{34.58}$$

is now the position of a fluid particle on the circuit, and $\mathrm{d}\boldsymbol{r}$ 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 $\boldsymbol{v}$ rather than the absolute velocity $\boldsymbol{v}_{\text{a}}$. Following the discussion for non-rotating Kelvin's Circulation Theorem in Section 34.2 leads to

$$\frac{\mathrm{d}\mathcal{C}_{\text{a}}}{\mathrm{d}t} = \frac{\mathrm{d}}{\mathrm{d}t}\oint_{\partial\mathcal{S}(\boldsymbol{v})} (\boldsymbol{v}+\boldsymbol{\Omega}\wedge\boldsymbol{r})\cdot\mathrm{d}\boldsymbol{r} \tag{34.59a}$$

$$= \oint_{\partial\mathcal{S}(\boldsymbol{v})} \left[\frac{\mathrm{D}\boldsymbol{v}}{\mathrm{D}t} + \boldsymbol{\Omega}\wedge\frac{\mathrm{D}\boldsymbol{r}}{\mathrm{D}t}\right]\cdot\mathrm{d}\boldsymbol{r} + \oint_{\partial\mathcal{S}(\boldsymbol{v})} (\boldsymbol{v}+\boldsymbol{\Omega}\wedge\boldsymbol{r})\cdot\mathrm{d}\boldsymbol{v} \tag{34.59b}$$

$$= \oint_{\partial\mathcal{S}(\boldsymbol{v})} \left[\frac{\mathrm{D}\boldsymbol{v}}{\mathrm{D}t} + \boldsymbol{\Omega}\wedge\boldsymbol{v}\right]\cdot\mathrm{d}\boldsymbol{r} + \oint_{\partial\mathcal{S}(\boldsymbol{v})} (\boldsymbol{\Omega}\wedge\boldsymbol{r})\cdot\mathrm{d}\boldsymbol{v} \tag{34.59c}$$

$$= \oint_{\partial\mathcal{S}(\boldsymbol{v})} \left[\frac{\mathrm{D}\boldsymbol{v}}{\mathrm{D}t} + 2\,\boldsymbol{\Omega}\wedge\boldsymbol{v}\right]\cdot\mathrm{d}\boldsymbol{r}. \tag{34.59d}$$

To reach this result we set

$$\boldsymbol{v} = \frac{\mathrm{D}\boldsymbol{r}}{\mathrm{D}t}, \tag{34.60}$$

for the velocity of a fluid particle on the circuit. We also used the identity

$$\oint_{\partial\mathcal{S}(\boldsymbol{v})} \boldsymbol{v}\cdot \mathrm{d}\boldsymbol{v} = \frac{1}{2}\oint_{\partial\mathcal{S}(\boldsymbol{v})} \mathrm{d}\boldsymbol{v}^2 = 0 \tag{34.61}$$

as well as

$$\oint_{\partial\mathcal{S}(\boldsymbol{v})} (\boldsymbol{\Omega}\wedge\boldsymbol{r})\cdot \mathrm{d}\boldsymbol{v} = \oint_{\partial\mathcal{S}(\boldsymbol{v})} \mathrm{d}\left[(\boldsymbol{\Omega}\wedge\boldsymbol{r})\cdot\boldsymbol{v}\right] - \oint_{\partial\mathcal{S}(\boldsymbol{v})} (\boldsymbol{\Omega}\wedge\mathrm{d}\boldsymbol{r})\cdot\boldsymbol{v} = \oint_{\partial\mathcal{S}(\boldsymbol{v})} (\boldsymbol{\Omega}\wedge\boldsymbol{v})\cdot\mathrm{d}\boldsymbol{r}. \tag{34.62}$$

Now insert the momentum equation (34.19) into equation (34.59d) to yield

$$\frac{\mathrm{d}\mathcal{C}_{\mathrm{a}}}{\mathrm{d}t} = \oint_{\partial\mathcal{S}(\boldsymbol{v})} \left[\frac{\mathrm{D}\boldsymbol{v}}{\mathrm{D}t} + 2\,\boldsymbol{\Omega}\wedge\boldsymbol{v}\right]\cdot\mathrm{d}\boldsymbol{r}. \tag{34.63a}$$

$$= \oint_{\partial\mathcal{S}(\boldsymbol{v})} \left[-\frac{1}{\rho}\nabla p - \nabla\Phi + \boldsymbol{F}\right]\cdot\mathrm{d}\boldsymbol{r}. \tag{34.63b}$$

$$= \oint_{\partial\mathcal{S}(\boldsymbol{v})} \left[-\frac{\mathrm{d}p}{\rho} + \boldsymbol{F}\cdot\mathrm{d}\boldsymbol{r}\right]. \tag{34.63c}$$

Making use of Stokes' Theorem leads to the evolution of absolute circulation around a material loop

$$\frac{\mathrm{d}\mathcal{C}_{\mathrm{a}}}{\mathrm{d}t} = \oint_{\partial\mathcal{S}(\boldsymbol{v})} \left[-\frac{\mathrm{d}p}{\rho} + \boldsymbol{F}\cdot\mathrm{d}\boldsymbol{r}\right] = \int_{S(\boldsymbol{v})} (\boldsymbol{B} + \nabla\wedge\boldsymbol{F})\cdot\hat{\boldsymbol{n}}\,\mathrm{d}\mathcal{S}, \tag{34.64}$$

where $\boldsymbol{B} = \rho^{-2}\,\nabla\rho\wedge\nabla p$ is the baroclinicity vector from equation (34.31).

The circulation theorem (34.64) is the same as obtained for the non-rotating Kelvin's Circulation Theorem discussed in Section 34.2 (see equation (34.12e)). This equivalence proves that circulation is an objective (frame invariant) property of the fluid, in which its evolution is unchanged when moving to a non-inertial rotating frame.

34.6.2 The beta effect

As given by equation (34.52), 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)

$$\mathcal{C}_{\mathrm{a}} = \mathcal{C} + \mathcal{C}_{\mathrm{planet}} = \mathcal{C} + 2\int_{\mathcal{S}} \boldsymbol{\Omega}\cdot\hat{\boldsymbol{n}}\,\mathrm{d}\mathcal{S} \Longleftrightarrow \frac{\mathrm{d}\mathcal{C}_{\mathrm{a}}}{\mathrm{d}t} = \frac{\mathrm{d}\mathcal{C}}{\mathrm{d}t} + \frac{\mathrm{d}\mathcal{C}_{\mathrm{planet}}}{\mathrm{d}t}. \tag{34.65}$$

We can determine the processes that affect the absolute circulation around a material loop by using the circulation equation (34.64)

$$\frac{\mathrm{d}\mathcal{C}}{\mathrm{d}t} = -\frac{\mathrm{d}\mathcal{C}_{\mathrm{planet}}}{\mathrm{d}t} + \frac{\mathrm{d}\mathcal{C}_{\mathrm{a}}}{\mathrm{d}t} \tag{34.66a}$$

$$= -2\,\frac{\mathrm{d}}{\mathrm{d}t}\left[\int_{\mathcal{S}(\boldsymbol{v})} \boldsymbol{\Omega}\cdot\hat{\boldsymbol{n}}\,\mathrm{d}\mathcal{S}\right] + \int_{\mathcal{S}(\boldsymbol{v})} (\boldsymbol{B} + \nabla\wedge\boldsymbol{F})\cdot\hat{\boldsymbol{n}}\,\mathrm{d}\mathcal{S}. \tag{34.66b}$$

We generally assume that the planetary rotation is a constant in time and points through the north pole of the sphere[3] $\boldsymbol{\Omega} = \Omega\,\hat{\boldsymbol{Z}}$, so that

$$\int_{\mathcal{S}(\boldsymbol{v})} \boldsymbol{\Omega}\cdot\hat{\boldsymbol{n}}\,\mathrm{d}\mathcal{S} = \Omega\int_{\mathcal{S}(\boldsymbol{v})} \hat{\boldsymbol{Z}}\cdot\hat{\boldsymbol{n}}\,\mathrm{d}\mathcal{S} = \Omega\,A_{\perp}. \tag{34.67}$$

[3]We follow the notational conventions of Figure 7.1 with one exception. Here, the vertical Cartesian direction through the north pole is written $\hat{\boldsymbol{Z}}$ to avoid confusion with the local vertical direction $\hat{\boldsymbol{z}}$ determined by the geopotential.

The area $A_\perp$ is the projection of the spherical area enclosed by the circuit onto the horizontal equatorial plane, with Figure 34.9 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

$$\frac{\mathrm{d}\mathcal{C}}{\mathrm{d}t} = \underbrace{-2\,\Omega\,\frac{\mathrm{d}A_\perp}{\mathrm{d}t}}_{\text{beta effect}} + \underbrace{\int_{\mathcal{S}} (\boldsymbol{B} + \nabla \wedge \boldsymbol{F}) \cdot \hat{\boldsymbol{n}}\, \mathrm{d}\mathcal{S}(\boldsymbol{v})}_{\text{solenoids plus friction}} . \tag{34.68}$$

Holton (1992) calls equation (34.68) the *Bjerknes Circulation Theorem* (see his 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 34.2 and 34.4.

Figure 34.9: *Geometry of the beta effect. According to the Bjerknes circulation theorem (34.68), 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, $2\,\Omega$, 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 ($2\,\Omega$) and curvature of the sphere ($\partial_y f = \beta$); it is therefore absent on the f-plane.*

The first term in the circulation theorem (34.68) 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 by the β-effect. Planetary geostrophic flow is the canonical example of such flow, as studied in Section 26.4 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.

34.6.3 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 39. For now, we use it as a toy example to help expose the essence of the β-effect (see also Section 39.3.3).

In the rotating frame, circulation around an infinitesimal closed material loop is

$$\mathcal{C} = A\,\zeta, \tag{34.69}$$

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 16.6). The material evolution of circulation is therefore given by

$$\frac{\mathrm{D}\mathcal{C}}{\mathrm{D}t} = A\,\frac{\mathrm{D}\zeta}{\mathrm{D}t} = -2\,\Omega\,\frac{\mathrm{D}A_\perp}{\mathrm{D}t}, \tag{34.70}$$

where the second equality follows from Bjerknes' circulation theorem (34.68). Let the material circuit be at a latitude ϕ so that the projection of the loop area onto the equatorial plane is (see Figure 34.9)

$$A_\perp = A\,\sin\phi. \tag{34.71}$$

Hence, material evolution of the circulation is

$$\begin{aligned}
\frac{\mathrm{D}\mathcal{C}}{\mathrm{D}t} &= A\,\frac{\mathrm{D}\zeta}{\mathrm{D}t} && \text{(34.72a)}\\
&= -2\,\Omega\,\frac{\mathrm{D}A_\perp}{\mathrm{D}t} && \text{(34.72b)}\\
&= -2\,A\,\Omega\,\frac{\mathrm{D}\sin\phi}{\mathrm{D}t} && \text{(34.72c)}\\
&= -2\,A\,\Omega\,\cos\phi\,\frac{\mathrm{D}\phi}{\mathrm{D}t} && \text{(34.72d)}\\
&= -A\left[\frac{2\,\Omega\,\cos\phi}{R}\right]\left[R\frac{\mathrm{D}\phi}{\mathrm{D}t}\right] && \text{(34.72e)}\\
&= -A\,\beta\,v, && \text{(34.72f)}
\end{aligned}$$

where we introduced the meridional velocity component

$$v = R\,\frac{\mathrm{D}\phi}{\mathrm{D}t} \tag{34.73}$$

and the meridional derivative of the planetary vorticity

$$\beta = \frac{\mathrm{d}f}{\mathrm{d}y} = \frac{1}{R}\frac{\mathrm{d}}{\mathrm{d}\phi}(2\,\Omega\,\sin\phi) = \frac{2\,\Omega\,\cos\phi}{R}. \tag{34.74}$$

The result (34.72f) shows how meridional motion on a rotating sphere induces relative circulation. It furthermore motivates the name *beta effect* for planetary induction of relative vorticity.

34.7 Vorticity in a hydrostatic-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

$$\frac{\mathrm{D}\boldsymbol{u}}{\mathrm{D}t} + \boldsymbol{f} \wedge \boldsymbol{v} = -\nabla_z \varphi + \boldsymbol{F} \tag{34.75a}$$

$$\frac{\partial \varphi}{\partial z} = b \tag{34.75b}$$

$$\nabla \cdot \boldsymbol{v} = 0 \tag{34.75c}$$

$$\frac{\mathrm{D}b}{\mathrm{D}t} = \dot{b}. \tag{34.75d}$$

In this equation, the non-divergent velocity field is written

$$\boldsymbol{v} = (\boldsymbol{u}, w) = \boldsymbol{u} + w\,\hat{\boldsymbol{z}}. \tag{34.76}$$

The perturbation pressure is

$$\varphi = \frac{\delta p}{\rho_0} = \frac{p - p_0}{\rho_0}, \tag{34.77}$$

with $p_0 = p_0(z)$ in hydrostatic balance with the constant reference density

$$\frac{\mathrm{d}p_0}{\mathrm{d}z} = -g\,\rho_0 \tag{34.78}$$

and p the full hydrostatic pressure satisfying the hydrostatic balance

$$\frac{\partial p}{\partial z} = -g\,\rho. \tag{34.79}$$

The buoyancy is given by

$$b = -g\left[\frac{\rho - \rho_0}{\rho_0}\right], \tag{34.80}$$

with negative buoyancy for fluid more dense than the reference density, ρ_0, and vice versa. We assume the Coriolis parameter of the form $\boldsymbol{f} = f\,\hat{\boldsymbol{z}}$, in which case

$$\boldsymbol{f} \wedge \boldsymbol{v} = \boldsymbol{f} \wedge \boldsymbol{u}. \tag{34.81}$$

Finally, the horizontal friction acceleration vector is given by

$$\boldsymbol{F} = (F^x, F^y, 0) \tag{34.82}$$

and the gradient operator is

$$\nabla = \nabla_z + \hat{\boldsymbol{z}}\,\partial_z. \tag{34.83}$$

34.7.1 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

$$\frac{\mathrm{D}\boldsymbol{u}}{\mathrm{D}t} + \boldsymbol{f} \wedge \boldsymbol{v} = -\nabla\varphi + b\,\hat{\boldsymbol{z}} + \boldsymbol{F}. \tag{34.84}$$

As for the non-hydrostatic case (Section 34.3.1), we rewrite the self-advection operator, $(\boldsymbol{v}\cdot\nabla)\,\boldsymbol{u}$, before taking the curl. In turn, we introduce the hydrostatic relative vorticity given by the curl of the horizontal velocity

$$\boldsymbol{\omega}^{\mathrm{hy}} = \nabla\wedge\boldsymbol{u} = \hat{\boldsymbol{z}}\wedge\partial_z\boldsymbol{u} + \hat{\boldsymbol{z}}\,\zeta = -\hat{\boldsymbol{x}}\,\partial_z v + \hat{\boldsymbol{y}}\,\partial_z u + \hat{\boldsymbol{z}}\,\zeta, \tag{34.85}$$

where

$$\zeta = \frac{\partial v}{\partial x} - \frac{\partial u}{\partial y} \tag{34.86}$$

is the vertical component to the relative vorticity, and

$$\nabla\cdot\boldsymbol{\omega}^{\mathrm{hy}} = 0. \tag{34.87}$$

It is then straightforward to show that

$$\boldsymbol{\omega}^{\mathrm{hy}}\wedge\boldsymbol{v} = \hat{\boldsymbol{x}}\,(w\,\partial_z u - v\,\partial_x v + v\,\partial_y u) + \hat{\boldsymbol{y}}\,(w\,\partial_z v - u\,\partial_y u + u\,\partial_x v) - \hat{\boldsymbol{z}}\,\partial_z(u^2+v^2)/2 \tag{34.88a}$$

$$= w\,\partial_z\boldsymbol{u} + \zeta\,(\hat{\boldsymbol{x}}\,v - \hat{\boldsymbol{u}}\,u) - \hat{\boldsymbol{z}}\,\partial_z(u^2+v^2)/2, \tag{34.88b}$$

in which case

$$\nabla(\boldsymbol{u}^2/2) + \boldsymbol{\omega}^{\mathrm{hy}}\wedge\boldsymbol{v} = \nabla\,(u^2+v^2)/2 - \hat{\boldsymbol{z}}\,\partial_z\,(u^2+v^2)/2 + w\,\partial_z\boldsymbol{u} + \zeta\,(u\hat{\boldsymbol{y}} - v\hat{\boldsymbol{x}}) \tag{34.89a}$$

$$= (u\,\partial_x + v\,\partial_y + w\,\partial_z)\,\boldsymbol{u} \tag{34.89b}$$

$$= (\boldsymbol{v}\cdot\nabla)\,\boldsymbol{u}. \tag{34.89c}$$

The material time derivative of the horizontal velocity can thus be written as

$$\frac{\mathrm{D}\boldsymbol{u}}{\mathrm{D}t} = \frac{\partial\boldsymbol{u}}{\partial t} + (\boldsymbol{v}\cdot\nabla)\,\boldsymbol{u} = \frac{\partial\boldsymbol{u}}{\partial t} + \boldsymbol{\omega}^{\mathrm{hy}}\wedge\boldsymbol{v} + \nabla(\boldsymbol{u}^2/2), \tag{34.90}$$

which then leads to the vector-invariant horizontal velocity equation

$$\frac{\partial\boldsymbol{u}}{\partial t} + (f\,\hat{\boldsymbol{z}} + \boldsymbol{\omega}^{\mathrm{hy}})\wedge\boldsymbol{v} = -\nabla\,(\varphi + \boldsymbol{u}^2/2) + \hat{\boldsymbol{z}}\,b + \boldsymbol{F}, \tag{34.91}$$

which can be written in the equivalent form

$$\left[\frac{\partial}{\partial t} + w\,\frac{\partial}{\partial z} + (f+\zeta)\,\hat{\boldsymbol{z}}\wedge\right]\boldsymbol{u} = -\nabla\,(\varphi + \boldsymbol{u}^2/2) + \hat{\boldsymbol{z}}\,b + \boldsymbol{F}. \tag{34.92}$$

Now take the curl of equation (34.91), and make use of the identity

$$\nabla\wedge(\boldsymbol{\omega}_{\mathrm{a}}^{\mathrm{hy}}\wedge\boldsymbol{v}) = (\boldsymbol{v}\cdot\nabla)\,\boldsymbol{\omega}_{\mathrm{a}}^{\mathrm{hy}} - (\boldsymbol{\omega}_{\mathrm{a}}^{\mathrm{hy}}\cdot\nabla)\,\boldsymbol{v}, \tag{34.93}$$

where we introduced the absolute vorticity for the hydrostatic fluid

$$\boldsymbol{\omega}_{\mathrm{a}}^{\mathrm{hy}} = f\,\hat{\boldsymbol{z}} + \boldsymbol{\omega}^{\mathrm{hy}} \tag{34.94}$$

thus resulting in the vorticity equation

$$\frac{\partial\boldsymbol{\omega}^{\mathrm{hy}}}{\partial t} + (\boldsymbol{v}\cdot\nabla)\,\boldsymbol{\omega}_{\mathrm{a}}^{\mathrm{hy}} = (\boldsymbol{\omega}_{\mathrm{a}}^{\mathrm{hy}}\cdot\nabla)\,\boldsymbol{v} + \nabla\wedge\hat{\boldsymbol{z}}\,b + \nabla\wedge\boldsymbol{F}. \tag{34.95}$$

Since the Coriolis parameter is time independent, we can add it to the time derivative to yield

$$\frac{\mathrm{D}\boldsymbol{\omega}_{\mathrm{a}}^{\mathrm{hy}}}{\mathrm{D}t} = \underbrace{(\boldsymbol{\omega}_{\mathrm{a}}^{\mathrm{hy}}\cdot\nabla)\,\boldsymbol{v}}_{\text{stretching + tilting}} + \underbrace{\nabla\wedge\hat{\boldsymbol{z}}\,b}_{\text{baroclinicity}} + \underbrace{\nabla\wedge\boldsymbol{F}}_{\text{friction curl}}. \tag{34.96}$$

34.7.2 Boussinesq form of the baroclinicity vector

Equation (34.96) 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 (34.30)). 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

$$\boldsymbol{B}_{\text{bouss}} = \nabla \wedge \hat{\boldsymbol{z}}\, b = \nabla b \wedge \hat{\boldsymbol{z}}, \tag{34.97}$$

which is simpler than baroclinicity in a compressible fluid

$$\boldsymbol{B} = (\nabla \rho \wedge \nabla p)/\rho^2 \tag{34.98}$$

(equation (34.31)). We see in Exercise 36.2 that the same baroclinicity vector (34.97) 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 34.10. That is, a sloping buoyancy surface acts as a vorticity source for the Boussinesq fluid by providing a spin to fluid elements (see Section 34.4). Correspondingly, baroclinicity in a Boussinesq fluid has no impact on the vertical component to absolute vorticity, since

$$\hat{\boldsymbol{z}} \cdot \boldsymbol{B}_{\text{bouss}} = \hat{\boldsymbol{z}} \cdot (\nabla \wedge \hat{\boldsymbol{z}}\, b) = 0. \tag{34.99}$$

Hence, baroclinicity in a Boussinesq fluid only acts as a source to horizontally oriented vorticity.

Figure 34.10: *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. Furthermore, as per equation (34.99), baroclinicity in a Boussinesq fluid only acts as a source to horizontally oriented vorticity.*

34.7.3 Vertical vorticity equation

The vertical component of the vorticity equation (34.96) takes the form

$$\frac{\mathrm{D}\zeta_{\mathrm{a}}}{\mathrm{D}t} = (\boldsymbol{\omega}_{\mathrm{a}}^{\mathrm{hy}} \cdot \nabla)\, w + \hat{\boldsymbol{z}} \cdot (\nabla \wedge \boldsymbol{F}), \tag{34.100}$$

with the absence of baroclinicity noted above in Section 34.7.2. 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

$$\frac{\partial \zeta_{\mathrm{a}}}{\partial t} + (\boldsymbol{v} \cdot \nabla)\, \zeta + \beta\, v = \hat{\boldsymbol{z}} \cdot (\partial_z \boldsymbol{u} \wedge \nabla_z w) + (\zeta + f)\, \partial_z w + \hat{\boldsymbol{z}} \cdot (\nabla \wedge \boldsymbol{F}). \tag{34.101}$$

Planetary geostrophic limit

The linearized, inviscid, and steady version of the vorticity equation (34.101) leads to the linear vorticity balance studied in Section 26.4.4

$$\beta\, v = f\, \partial_z w \tag{34.102}$$

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 (34.101) in an alternative form by making use of $\nabla \cdot \boldsymbol{v} = \nabla \cdot \boldsymbol{\omega}_{\mathrm{a}}^{\mathrm{hy}} = 0$ to yield

$$\frac{\partial \zeta_{\mathrm{a}}}{\partial t} = -\nabla \cdot (\boldsymbol{v}\, \zeta_{\mathrm{a}} - w\, \boldsymbol{\omega}_{\mathrm{a}}^{\mathrm{hy}}) + \hat{\boldsymbol{z}} \cdot (\nabla \wedge \boldsymbol{F}). \tag{34.103}$$

Furthermore, we can write the friction curl in the form

$$\begin{aligned}
\hat{\boldsymbol{z}} \cdot (\nabla \wedge \boldsymbol{F}) &= \nabla z \cdot (\nabla \wedge \boldsymbol{F}) && (34.104\mathrm{a})\\
&= \nabla \cdot [z\, \nabla \wedge \boldsymbol{F}] && (34.104\mathrm{b})\\
&= \nabla \cdot [\nabla \wedge (z\, \boldsymbol{F}) - \nabla z \wedge \boldsymbol{F}] && (34.104\mathrm{c})\\
&= -\nabla \cdot (\hat{\boldsymbol{z}} \wedge \boldsymbol{F}). && (34.104\mathrm{d})
\end{aligned}$$

Hence, the vertical component of the Boussinesq vorticity evolves according to the convergence of the vorticity flux

$$\frac{\partial \zeta_{\mathrm{a}}}{\partial t} = -\nabla \cdot \boldsymbol{J}_{\zeta} \qquad \text{with} \qquad \boldsymbol{J}_{\zeta} = \boldsymbol{v}\, \zeta_{\mathrm{a}} - w\, \boldsymbol{\omega}_{\mathrm{a}}^{\mathrm{hy}} + \hat{\boldsymbol{z}} \wedge \boldsymbol{F}. \tag{34.105}$$

There is no vertical contribution to the vorticity flux vector, as seen by noting that $\hat{\boldsymbol{z}} \wedge \boldsymbol{F}$ is a horizontal vector. Furthermore, $\boldsymbol{\omega}_{\mathrm{a}}^{\mathrm{hy}} = \hat{\boldsymbol{z}} \wedge \partial_z \boldsymbol{u} + \hat{\boldsymbol{z}}\, \zeta_{\mathrm{a}}$ allows us to write

$$\boldsymbol{v}\, \zeta_{\mathrm{a}} - w\, \boldsymbol{\omega}_{\mathrm{a}}^{\mathrm{hy}} = \boldsymbol{u}\, \zeta_{\mathrm{a}} - w\, \hat{\boldsymbol{z}} \wedge \partial_z \boldsymbol{u}, \tag{34.106}$$

which also is a horizontal vector.

34.7.4 Depth integral of the vertical vorticity equation

We now consider the depth integrated the vorticity equation (34.105) over the full depth of an ocean, thus integrating from the ocean bottom at $z = \eta_{\mathrm{b}}(x, y)$ to the ocean surface at $z = \eta(x, y, t)$ (see Figure 29.1). Doing so provides a useful example of how to perform such integrals and in particular how to isolate the boundary contributions. It also 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.

The necessary manipulations are typical for the analysis of depth integrated budgets, with care especially needed to account for the boundary conditions. For that purpose we make use of Leibniz's Rule (Section 15.3.4) to move the time and space derivatives in the equation

$$\int_{\eta_\mathrm{b}}^{\eta} \frac{\partial \zeta_\mathrm{a}}{\partial t}\, \mathrm{d}z = -\int_{\eta_\mathrm{b}}^{\eta} \nabla \cdot \boldsymbol{J}_\zeta \, \mathrm{d}z \tag{34.107}$$

from inside the integrals to outside. We thus make use of the following Leibniz's Rule results

$$\int_{\eta_\mathrm{b}}^{\eta} \frac{\partial \zeta_\mathrm{a}}{\partial t}\, \mathrm{d}z = -[\zeta_\mathrm{a}\, \partial_t \eta]_{z=\eta} + \frac{\partial}{\partial t} \int_{\eta_\mathrm{b}}^{\eta} \zeta_\mathrm{a}\, \mathrm{d}z \tag{34.108}$$

$$-\int_{\eta_\mathrm{b}}^{\eta} \nabla_z \cdot \boldsymbol{J}_\zeta \, \mathrm{d}z = [\nabla_z \eta \cdot \boldsymbol{J}_\zeta]_{z=\eta} - [\nabla_z \eta_\mathrm{b} \cdot \boldsymbol{J}_\zeta]_{z=\eta_\mathrm{b}} - \nabla_z \cdot \int_{\eta_\mathrm{b}}^{\eta} \boldsymbol{J}_\zeta \, \mathrm{d}z \tag{34.109}$$

$$-\int_{\eta_\mathrm{b}}^{\eta} \frac{\partial (\hat{\boldsymbol{z}} \cdot \boldsymbol{J}_\zeta)}{\partial z}\, \mathrm{d}z = -[\hat{\boldsymbol{z}} \cdot \boldsymbol{J}_\zeta]_{z=\eta} + [\hat{\boldsymbol{z}} \cdot \boldsymbol{J}_\zeta]_{z=\eta_\mathrm{b}}. \tag{34.110}$$

These results then lead to

$$\frac{\partial}{\partial t} \int_{\eta_\mathrm{b}}^{\eta} \zeta_\mathrm{a}\, \mathrm{d}z = -\nabla_z \cdot \int_{\eta_\mathrm{b}}^{\eta} \boldsymbol{J}_\zeta \, \mathrm{d}z + [\zeta_\mathrm{a}\, \partial_t \eta - \nabla(z-\eta) \cdot \boldsymbol{J}_\zeta]_{z=\eta} + [\nabla(z-\eta_\mathrm{b}) \cdot \boldsymbol{J}_\zeta]_{z=\eta_\mathrm{b}} \,. \tag{34.111}$$

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. We next massage the boundary contributions to help understand their physical content.

Bottom boundary contribution

The bottom boundary contribution to the vorticity equation (34.111) takes on the form

$$\nabla(z-\eta_\mathrm{b}) \cdot \boldsymbol{J}_\zeta = \nabla(z-\eta_\mathrm{b}) \cdot [\boldsymbol{v}\, \zeta_\mathrm{a} - w\, \boldsymbol{\omega}_\mathrm{a}^{\mathrm{hy}} + \hat{\boldsymbol{z}} \wedge \boldsymbol{F}]_{z=\eta_\mathrm{b}} \tag{34.112a}$$

$$= -|\nabla(z-\eta_\mathrm{b})|\, \hat{\boldsymbol{n}} \cdot [-w\, \boldsymbol{\omega}_\mathrm{a}^{\mathrm{hy}} + \hat{\boldsymbol{z}} \wedge \boldsymbol{F}]_{z=\eta_\mathrm{b}} \,, \tag{34.112b}$$

where we made use of the no-normal flow bottom kinematic boundary condition $\hat{\boldsymbol{n}} \cdot \boldsymbol{v} = 0$ (Section 14.4.1) and where

$$\hat{\boldsymbol{n}} = -\left[\frac{\nabla(z-\eta_\mathrm{b})}{|\nabla(z-\eta_\mathrm{b})|}\right] = -\left[\frac{\hat{\boldsymbol{z}} - \nabla_z \eta_\mathrm{b}}{\sqrt{1 + \nabla_z \eta_\mathrm{b} \cdot \nabla_z \eta_\mathrm{b}}}\right] \tag{34.113}$$

is the outward normal at the bottom. The first term in the bottom boundary flux (34.112b) 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 vorticity perpendicular to the bottom. This term vanishes for a flat bottom, in which case $w(\eta_\mathrm{b}) = 0$. The second term provides the contribution from friction along the bottom.

Surface boundary contribution

The surface boundary contribution to the vorticity equation (34.111) takes on a similar form to the bottom, with the new feature that the free surface is both moving and permeable (Section 14.4.3).

This boundary term takes on the form

$$\zeta_{\mathrm{a}}\,\partial_t\eta - \nabla(z-\eta)\cdot\boldsymbol{J}_\zeta = \zeta_{\mathrm{a}}\,\partial_t\eta - \nabla(z-\eta)\cdot\left[\boldsymbol{v}\,\zeta_{\mathrm{a}} - w\,\boldsymbol{\omega}_{\mathrm{a}}^{\mathrm{hy}} + \hat{\boldsymbol{z}}\wedge\boldsymbol{F}\right]_{z=\eta} \tag{34.114a}$$

$$= \zeta_{\mathrm{a}}\left[\partial_t\eta + \boldsymbol{u}\cdot\nabla_z\eta - w\right]_{z=\eta} + \nabla(z-\eta)\cdot\left[w\,\boldsymbol{\omega}_{\mathrm{a}}^{\mathrm{hy}} - \hat{\boldsymbol{z}}\wedge\boldsymbol{F}\right]_{z=\eta} \tag{34.114b}$$

$$= \zeta_{\mathrm{a}}\,Q_{\mathrm{m}}/\rho_0 + |\nabla(z-\eta)|\,\hat{\boldsymbol{n}}\cdot\left[w\,\boldsymbol{\omega}_{\mathrm{a}}^{\mathrm{hy}} - \hat{\boldsymbol{z}}\wedge\boldsymbol{F}\right]_{z=\eta}, \tag{34.114c}$$

where we made use of the surface kinematic boundary condition (14.73) to introduce the surface mass flux Q_{m}, and where

$$\hat{\boldsymbol{n}} = \frac{\nabla(z-\eta)}{|\nabla(z-\eta)|} = \frac{\nabla_z\eta + \hat{\boldsymbol{z}}}{\sqrt{1+\nabla_z\eta\cdot\nabla_z\eta}} \tag{34.115}$$

is the outward normal at the surface. The first term in the surface boundary flux (34.114c) 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 passing through the boundary. The third term provides the corresponding contribution from the friction along the surface. For a rigid lid surface, $w(0) = 0$ and $Q_{\mathrm{m}} = 0$ so that the only surface boundary contribution arises from friction.

34.8 Exercises

EXERCISE 34.1: FRICTION IN THE VORTICITY EQUATION
Add a viscous term of the form

$$\boldsymbol{F} = \mu\nabla^2\boldsymbol{v}, \tag{34.116}$$

with μ a constant molecular viscosity. How is the vorticity equation modified?

EXERCISE 34.2: 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

$$\rho(x) = \rho_0\,(1-\gamma\,x) \tag{34.117}$$

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 34.4 will help with this exercise.

(a) Compute the density gradient $\nabla\rho$ and draw a schematic.

(b) Compute the pressure gradient ∇p. Draw a schematic at $x = 0$.

(c) Compute the baroclinicity/solenoidal vector $\boldsymbol{B} = \rho^{-2}\,(\nabla\rho\wedge\nabla p)$. Draw a schematic.

(d) Describe the vorticity induced by the baroclinicity vector.

EXERCISE 34.3: CIRCULATION WITH ISLANDS
Our discussion of Stokes' Theorem has been thus far restricted to a simply connected domain, in which

$$\oint_{\partial\mathcal{S}} \boldsymbol{v}\cdot\mathrm{d}\boldsymbol{r} = \int_{\mathcal{S}} \boldsymbol{\omega}\cdot\hat{\boldsymbol{n}}\,\mathrm{d}\mathcal{S}. \tag{34.118}$$

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 34.11 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 an AquaPlanet or the global atmosphere.

Derive the following expression for the circulation in multiply-connected regions

$$\oint_{\partial\mathcal{S}} \boldsymbol{v}\cdot \mathrm{d}\boldsymbol{r} = \sum_{n=1}^{N}\left(\oint_{\partial\mathcal{S}_n} \boldsymbol{v}\cdot \mathrm{d}\boldsymbol{r}\right) + \int_{\mathcal{S}} \boldsymbol{\omega}\cdot\hat{\boldsymbol{n}}\,\mathrm{d}\mathcal{S}, \tag{34.119}$$

where N is the number of islands and $\mathcal{S}_n$ is the contour surrounding each island. 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 region. Removing the islands allows the island contours to be shrunk to zero size, in which case we recover the simply connected result (34.118).

Figure 34.11: *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 34.3 is concerned with deriving an expression for the circulation of the depth-integrated flow as defined along the closed contour, $\partial\mathcal{S}$.*

EXERCISE 34.4: EVOLUTION OF CIRCULATION AROUND ISLANDS

The momentum equation for a homogeneous layer of inviscid shallow water on a tangent plane is given by

$$\frac{\partial \boldsymbol{u}}{\partial t} + (\boldsymbol{u}\cdot\nabla)\,\boldsymbol{u} + f\,\hat{\boldsymbol{z}}\wedge\boldsymbol{u} = -g\,\nabla\eta. \tag{34.120}$$

In this equation, $\boldsymbol{u} = (u, v)$ 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 position. All spatial derivatives are horizontal, so that

$$\boldsymbol{u}\cdot\nabla = u\,\partial_x + v\,\partial_y. \tag{34.121}$$

Use of a vector identity allows us to write

$$\frac{\partial \boldsymbol{u}}{\partial t} + (f + \zeta)\, \hat{\boldsymbol{z}} \wedge \boldsymbol{u} = -\nabla\, (\boldsymbol{u}^2/2 + g\, \eta), \tag{34.122}$$

where

$$\zeta = \hat{\boldsymbol{z}} \cdot (\nabla \wedge \boldsymbol{u}) \tag{34.123}$$

is the vorticity of the shallow water fluid.

Consider an island, such as one shown in Figure 34.11. Each island is static and impenetrable to fluid flow, which means that

$$\boldsymbol{u} \cdot \hat{\boldsymbol{n}} = 0 \tag{34.124}$$

where $\hat{\boldsymbol{n}}$ 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 island[4]

$$\boldsymbol{u} \wedge \mathrm{d}\boldsymbol{r} = 0. \tag{34.125}$$

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

$$\frac{\mathrm{d}}{\mathrm{d}t} \oint_I \boldsymbol{u} \cdot \mathrm{d}\boldsymbol{r} = 0. \tag{34.126}$$

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.

EXERCISE 34.5: 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, $\boldsymbol{\omega}$. Hence, the outward normal to the region boundary is orthogonal to the vorticity,

$$\hat{\boldsymbol{n}} \cdot \boldsymbol{\omega} = 0. \tag{34.127}$$

Such volumes define <u>closed</u> 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, $\boldsymbol{v} \cdot \boldsymbol{\omega}$, over the closed volume

$$\mathbb{H} = \int_{\mathcal{R}(v)} \boldsymbol{v} \cdot \boldsymbol{\omega}\, \mathrm{d}V, \tag{34.128}$$

where the volume $\mathcal{R}(v)$ is material. In Cartesian coordinates, the helicity density takes the form

$$\boldsymbol{v} \cdot \boldsymbol{\omega} = u\, (\partial_y w - \partial_z v) + v\, (\partial_z u - \partial_x w) + w\, (\partial_x v - \partial_y u). \tag{34.129}$$

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

$$\frac{\mathrm{d}\mathbb{H}}{\mathrm{d}t} = 0. \tag{34.130}$$

[4]This 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.

(b) Discuss why helicity is not defined for a shallow water fluid.

Use the following hints.

- Make use of Φ_{p} that satisfies equation (34.17).
- 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.

35

Angular momentum, vorticity, and strain

As noted in Section 32.5.4, fluid flow in the presence of a free vortex (Section 32.4) 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 32.5) 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 34, 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.

35.1 A resume of point particle mechanics **639**
35.1.1 Angular velocity and moment of inertia 640
35.1.2 Relating angular velocity to velocity 641
35.2 Linear momentum for material fluid regions **641**
35.3 Angular momentum for material fluid regions **642**
35.3.1 Taylor expanding the velocity 643
35.3.2 Strain and vorticity 643
35.3.3 Relating angular momentum to strain and vorticity 644
35.3.4 Comments 645

35.1 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 written[1]

$$\boldsymbol{P} = M\,\boldsymbol{V}, \tag{35.1}$$

[1] We use capital letters to accord with our usage of Lagrangian fluid particle trajectories.

where M is the particle's mass, which is a measure of the particle's inertia. The velocity, $\boldsymbol{V}$, is the time change of the particle position,

$$\boldsymbol{V} = \frac{\mathrm{d}\boldsymbol{X}}{\mathrm{d}t}. \tag{35.2}$$

The corresponding angular momentum is given by

$$\boldsymbol{L} = \boldsymbol{X} \wedge \boldsymbol{P} = M\,(\boldsymbol{X} \wedge \boldsymbol{V}). \tag{35.3}$$

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 10.6 that the component of angular momentum about the rotation axis is a constant of the motion.

35.1.1 Angular velocity and moment of inertia

Whereas linear momentum has physical dimensions of

$$[\boldsymbol{P}] \equiv \text{mass} \times \text{length} \times \text{time}^{-1}, \tag{35.4}$$

angular momentum has dimensions of

$$[\boldsymbol{L}] \equiv \text{mass} \times \text{length}^2 \times \text{time}^{-1}. \tag{35.5}$$

We can pursue the analog by introducing the angular velocity

$$\boldsymbol{\Omega} = \frac{\boldsymbol{X} \wedge \boldsymbol{V}}{|\boldsymbol{X}|^2}. \tag{35.6}$$

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

$$\boldsymbol{\Omega} \cdot \boldsymbol{X} = \boldsymbol{\Omega} \cdot \boldsymbol{V} = 0. \tag{35.7}$$

The angular velocity is not defined at the origin since $|\boldsymbol{X}| = 0$.

Inserting the definition of the angular velocity (35.6) into the angular momentum (35.3) renders

$$\boldsymbol{L} = M\,(\boldsymbol{X} \wedge \boldsymbol{V}) \tag{35.8a}$$

$$= M\,|\boldsymbol{X}|^2\,\boldsymbol{\Omega} \tag{35.8b}$$

$$\equiv I\,\boldsymbol{\Omega}. \tag{35.8c}$$

In the final equality we introduced the moment of inertia for a point particle

$$I = M\,|\boldsymbol{X}|^2. \tag{35.9}$$

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 35.3.3).

35.1.2 Relating angular velocity to velocity

The cross product of the position vector with the angular velocity (35.6) is given by

$$\boldsymbol{\Omega} \wedge \boldsymbol{X} = \boldsymbol{V} - \boldsymbol{X}\,\frac{\boldsymbol{V}\cdot\boldsymbol{X}}{|\boldsymbol{X}|^2}. \tag{35.10}$$

For the special case of velocity $\boldsymbol{V}$ orthogonal to the position vector, $\boldsymbol{X}$, we have

$$\boldsymbol{V} = \boldsymbol{\Omega} \wedge \boldsymbol{X} \qquad \text{when } \boldsymbol{V}\cdot\boldsymbol{X} = 0. \tag{35.11}$$

In particular, for circular motion the velocity is orthogonal to the position.

35.2 Linear momentum for material fluid regions

We now consider the velocity and linear momentum of a connected material fluid region denoted by $\mathcal{R}(\boldsymbol{v})$, with each point of the region moving with the local fluid velocity $\boldsymbol{v}(\boldsymbol{x},t)$. Let an arbitray fluid parcel within this region be marked with the material label $\boldsymbol{a}$ so that its position vector is $\boldsymbol{X}(\boldsymbol{a},t)$ and its velocity is

$$\boldsymbol{V}(\boldsymbol{a},t) = \frac{\partial \boldsymbol{X}(\boldsymbol{a},t)}{\partial t}, \tag{35.12}$$

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

$$\boldsymbol{V}(\boldsymbol{a},t) = \frac{\partial \boldsymbol{X}(\boldsymbol{a},t)}{\partial t} \tag{35.13a}$$

$$= \frac{\partial(\overline{\boldsymbol{X}} + \boldsymbol{X}')}{\partial t} \tag{35.13b}$$

$$= \overline{\boldsymbol{V}}(t) + \boldsymbol{V}'(\boldsymbol{a},t). \tag{35.13c}$$

In this equation, we introduced the velocity $\boldsymbol{V}'$ defined relative to the center of mass of the region. Furthermore, the center of mass velocity is given by

$$\overline{\boldsymbol{V}} = \frac{\mathrm{d}\overline{\boldsymbol{X}}}{\mathrm{d}t} \tag{35.14a}$$

$$= \frac{\mathrm{d}}{\mathrm{d}t}\left[\frac{\int_{\mathcal{R}(\boldsymbol{v})} \boldsymbol{x}\,\rho\,\mathrm{d}V}{\int_{\mathcal{R}(\boldsymbol{v})} \rho\,\mathrm{d}V}\right] \tag{35.14b}$$

$$= \frac{1}{M}\int_{\mathcal{R}(\boldsymbol{v})} \frac{\mathrm{D}\boldsymbol{x}}{\mathrm{D}t}\,\rho\,\mathrm{d}V \tag{35.14c}$$

$$= \frac{1}{M}\int_{\mathcal{R}(\boldsymbol{v})} \boldsymbol{v}\,\rho\,\mathrm{d}V. \tag{35.14d}$$

The identity (35.14c) follows since the material region maintains a constant mass,

$$M = \int_{\mathcal{R}(\boldsymbol{v})} \rho\,\mathrm{d}V \Longrightarrow \frac{\mathrm{d}M}{\mathrm{d}t} = 0, \tag{35.15}$$

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 14.3.4), the time

derivative moves across the integral to act materially on the position vector. The final equality, (35.14d), follows since the material time derivative of a parcel trajectory when evaluated at a point, $\boldsymbol{x}$, equals to the velocity field at that point

$$\boldsymbol{v}(\boldsymbol{x}, t) = \frac{\mathrm{D}\boldsymbol{x}}{\mathrm{D}t}. \tag{35.16}$$

It follows that the linear momentum for the material fluid region is given by

$$\boldsymbol{P} = \int_{\mathcal{R}(\boldsymbol{v})} \boldsymbol{v}\, \rho\, \mathrm{d}V = M\, \overline{\boldsymbol{V}}. \tag{35.17}$$

We conclude that the total linear momentum of an extended body equals to that of a point particle of mass $M = \int_{\mathcal{R}(\boldsymbol{v})} \rho\, \mathrm{d}V$ moving with the center of mass velocity, $\overline{\boldsymbol{V}}$.

35.3 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

$$\boldsymbol{L} = \int_{\mathcal{R}(\boldsymbol{v})} (\boldsymbol{x} \wedge \boldsymbol{v})\, \rho\, \mathrm{d}V. \tag{35.18}$$

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, $\boldsymbol{x} = \overline{\boldsymbol{x}} + \boldsymbol{x}'$, where $\overline{\boldsymbol{x}} = \overline{X}$ is the instantaneous position of the moving center of mass. The angular momentum thus takes the form

$$\boldsymbol{L} = \int_{\mathcal{R}(\boldsymbol{v})} (\boldsymbol{x} \wedge \boldsymbol{v})\, \rho\, \mathrm{d}V \tag{35.19a}$$

$$= \int_{\mathcal{R}(\boldsymbol{v})} [(\overline{\boldsymbol{x}} + \boldsymbol{x}') \wedge \boldsymbol{v}]\, \rho\, \mathrm{d}V \tag{35.19b}$$

$$= \overline{\boldsymbol{X}} \wedge \left[\int_{\mathcal{R}(\boldsymbol{v})} \boldsymbol{v}\, \rho\, \mathrm{d}V \right] + \int_{\mathcal{R}(\boldsymbol{v})} (\boldsymbol{x}' \wedge \boldsymbol{v})\, \rho\, \mathrm{d}V \tag{35.19c}$$

$$= (\overline{\boldsymbol{X}} \wedge \boldsymbol{P}) + \int_{\mathcal{R}(\boldsymbol{v})} (\boldsymbol{x}' \wedge \boldsymbol{v})\, \rho\, \mathrm{d}V. \tag{35.19d}$$

The final equality introduced the linear momentum, (35.17), for the fluid region. The first term in equation (35.19d) 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, $\int_{\mathcal{R}(\boldsymbol{v})} (\boldsymbol{x}' \wedge \boldsymbol{v})\, \rho\, \mathrm{d}V$, contributes to the angular momentum (35.18). 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)

$$(\boldsymbol{A} \wedge \boldsymbol{B})_m = \epsilon_{mnp}\, A_n\, B_p. \tag{35.20}$$

35.3.1 Taylor expanding the velocity

We now perform a Taylor expansion of the velocity $\boldsymbol{v}(\boldsymbol{x})$ around the instantaneous center of mass position, $\overline{\boldsymbol{x}} = \overline{X}$, and truncate the expansion to the leading order term[2]

$$\boldsymbol{v}(\boldsymbol{x}) = \boldsymbol{v}(\overline{\boldsymbol{x}} + \boldsymbol{x}') \approx \boldsymbol{v}(\overline{\boldsymbol{x}}) + (\boldsymbol{x}' \cdot \nabla)\, \boldsymbol{v}|_{\boldsymbol{x}=\overline{\boldsymbol{x}}}. \tag{35.21}$$

We are thus left with

$$\boldsymbol{L} = (\overline{\boldsymbol{X}} \wedge \boldsymbol{P}) + \int_{\mathcal{R}(\boldsymbol{v})} (\boldsymbol{x}' \wedge \boldsymbol{v})\, \rho\, \mathrm{d}V \tag{35.22a}$$

$$= (\overline{\boldsymbol{X}} \wedge \boldsymbol{P}) + \int_{\mathcal{R}(\boldsymbol{v})} [\boldsymbol{x}' \wedge \boldsymbol{v}(\overline{\boldsymbol{x}})]\, \rho\, \mathrm{d}V + \int_{\mathcal{R}(\boldsymbol{v})} [\boldsymbol{x}' \wedge (\boldsymbol{x}' \cdot \nabla)\, \boldsymbol{v}(\overline{\boldsymbol{x}})]\, \rho\, \mathrm{d}V. \tag{35.22b}$$

The velocity $\boldsymbol{v}(\overline{\boldsymbol{x}})$ can be removed from the integration since it is evaluated at the center of mass point. Hence, the second term in equation (35.22b) vanishes

$$\int_{\mathcal{R}(\boldsymbol{v})} [\boldsymbol{x}' \wedge \boldsymbol{v}(\overline{\boldsymbol{x}})]\, \rho\, \mathrm{d}V = \left[\int_{\mathcal{R}(\boldsymbol{v})} \boldsymbol{x}'\, \rho\, \mathrm{d}V\right] \wedge \boldsymbol{v}(\overline{\boldsymbol{x}}) = 0, \tag{35.23}$$

where $\int_{\mathcal{R}(\boldsymbol{v})} \boldsymbol{x}'\, \rho\, \mathrm{d}V = 0$ by definition of the center of mass. The angular momentum is thus given by the two terms

$$\boldsymbol{L} = (\overline{\boldsymbol{X}} \wedge \boldsymbol{P}) + \int_{\mathcal{R}(\boldsymbol{v})} [\boldsymbol{x}' \wedge (\boldsymbol{x}' \cdot \nabla)\, \boldsymbol{v}(\overline{\boldsymbol{x}})]\, \rho\, \mathrm{d}V. \tag{35.24}$$

The $m'th$ component of the second term can be written

$$\int_{\mathcal{R}(\boldsymbol{v})} [\boldsymbol{x}' \wedge (\boldsymbol{x}' \cdot \nabla)\, \boldsymbol{v}(\overline{\boldsymbol{x}})]_m\, \rho\, \mathrm{d}V = \epsilon_{mnp} \int_{\mathcal{R}(\boldsymbol{v})} x'_n\, [(\boldsymbol{x}' \cdot \nabla)\, \boldsymbol{v}(\overline{\boldsymbol{x}})]_p\, \rho\, \mathrm{d}V \tag{35.25a}$$

$$= \epsilon_{mnp} \int_{\mathcal{R}(\boldsymbol{v})} x'_n\, x'_q\, \partial_q v(\overline{\boldsymbol{x}})_p\, \rho\, \mathrm{d}V \tag{35.25b}$$

$$= \epsilon_{mnp} \left[\int_{\mathcal{R}(\boldsymbol{v})} x'_n\, x'_q\, \rho\, \mathrm{d}V\right] \partial_q v(\overline{\boldsymbol{x}})_p. \tag{35.25c}$$

We removed the velocity derivatives

$$\partial_q v(\overline{\boldsymbol{x}})_p = \left[\frac{\partial v_p}{\partial x_q}\right]_{\boldsymbol{x}=\overline{\boldsymbol{x}}} \tag{35.26}$$

from the integral, since they are evaluated at the center of mass point and so do not participate in the integration.

35.3.2 Strain and vorticity

Following from the discussion in Section 13.2.4, we know that the velocity derivatives $\partial_q v_p$ appearing in equation (35.25c) form the components to a second order tensor. To expose the kinematics of this tensor, decompose it into its symmetric and anti-symmetric components

$$\partial_q\, v_p = \frac{1}{2}(\partial_q\, v_p + \partial_p\, v_q) + \frac{1}{2}(\partial_q\, v_p - \partial_p\, v_q) \tag{35.27a}$$

$$\equiv S_{qp} + A_{qp}. \tag{35.27b}$$

[2]The velocity field evaluated at the center of mass position, $\boldsymbol{v}(\overline{\boldsymbol{x}})$, is not equal to the center of mass velocity: $\boldsymbol{v}(\overline{\boldsymbol{x}}) \neq \overline{\boldsymbol{v}}$.

The symmetric tensor

$$S_{qp} = \frac{1}{2}(\partial_q\, v_p + \partial_p\, v_q) \tag{35.28}$$

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

$$2A_{qp} = \partial_q v_p - \partial_p v_q \tag{35.29a}$$
$$= (\delta_{qm}\,\delta_{pn} - \delta_{qn}\,\delta_{pm})\,\partial_m v_n \tag{35.29b}$$
$$= \epsilon_{sqp}\,\epsilon_{smn}\,\partial_m v_n \tag{35.29c}$$
$$= \epsilon_{sqp}\,\omega_s, \tag{35.29d}$$

where $\omega_s = \epsilon_{smn}\,\partial_m v_n$ are components to the vorticity pseudo-vector

$$\boldsymbol{\omega} = \nabla \wedge \boldsymbol{v}. \tag{35.30}$$

35.3.3 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

$$L_m = (\overline{\boldsymbol{X}} \wedge \boldsymbol{P})_m + \epsilon_{mnp}\left[\int_{\mathcal{R}(\boldsymbol{v})} x'_n x'_q\,\rho\,\mathrm{d}V\right] S_{qp} + \epsilon_{mnp}\left[\int_{\mathcal{R}(\boldsymbol{v})} x'_n x'_q\,\rho\,\mathrm{d}V\right] A_{qp} \tag{35.31a}$$
$$= (\overline{\boldsymbol{X}} \wedge \boldsymbol{P})_m + \epsilon_{mnp}\left[\int_{\mathcal{R}(\boldsymbol{v})} x'_n x'_q\,\rho\,\mathrm{d}V\right] S_{qp} + \frac{1}{2}\,\epsilon_{mnp}\,\epsilon_{sqp}\left[\int_{\mathcal{R}(\boldsymbol{v})} x'_n x'_q\,\rho\,\mathrm{d}V\right]\omega_s \tag{35.31b}$$
$$= \underbrace{(\overline{\boldsymbol{X}} \wedge \boldsymbol{P})_m}_{\text{center of mass}} + \underbrace{\epsilon_{mnp}\left[\int_{\mathcal{R}(\boldsymbol{v})} x'_n x'_q\,\rho\,\mathrm{d}V\right] S_{qp}}_{\text{strain contribution}} + \underbrace{\frac{1}{2}\left[\int_{\mathcal{R}(\boldsymbol{v})} (\boldsymbol{x}'\cdot\boldsymbol{x}'\,\delta_{ms} - x'_m\,x'_s)\,\rho\,\mathrm{d}V\right]\omega_s}_{\text{vorticity contribution}}\,. \tag{35.31c}$$

As each point in the fluid can be considered the center of mass for an arbitrary material region, the decomposition (35.31c) is general.

- Center of mass angular momentum: The first term on the right hand side of equation (35.31c) 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 (35.8a)). 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 13.2.4). 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 32.5). 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 (35.31c) contains the vorticity as weighted by the moment of inertia tensor

$$I_{ms} \equiv \int_{\mathcal{R}(\boldsymbol{v})} (\boldsymbol{x}' \cdot \boldsymbol{x}' \, \delta_{ms} - x'_m \, x'_s) \, \rho \, \mathrm{d}V. \tag{35.32}$$

  Since the material region is evolving and is not rigid, the moment of inertia tensor is time dependent. The contribution

$$L_m^{\text{vorticity}} \equiv \frac{1}{2} \, I_{ms} \, \omega_s \tag{35.33}$$

  has the same form as angular momentum for a rigid body, with one-half the vorticity playing the role of angular velocity (see equation (35.8c) 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.

35.3.4 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.

36

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 ("trick") exploited for the construction of PV is to choose a scalar field to strategically orient the vorticity. If the scalar is a material invariant, and it annihilates the baroclinicity vector, then PV is a material invariant. Notably, 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 constrained, 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 allows for insights into the interpretation of that motion.

READER'S GUIDE FOR THIS CHAPTER

This chapter requires a firm understanding of vorticity from Chapter 34 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.

36.1 Material invariance of PV in perfect fluids **648**
36.1.1 Perfect barotropic fluid 648
36.1.2 Region between two scalar isosurfaces 648
36.1.3 Material invariance 649
36.1.4 Perfect baroclinic fluid 650
36.1.5 Some general remarks 651
36.2 PV and the seawater equation of state **652**
36.2.1 Baroclinicity vector 652
36.2.2 PV based on potential density 652
36.2.3 An example EOS admitting a materially invariant PV 653

36.2.4 Further reading 654
36.3 PV evolution with non-conservative processes **654**
36.4 Eulerian flux-form PV budget **655**
36.4.1 Deriving the Eulerian flux-form PV budget 656
36.4.2 PV-substance and the PV flux 656
36.5 PV budget for a hydrostatic Boussinesq fluid **657**
36.5.1 Potential vorticity 657
36.5.2 Potential vorticity flux vector 658
36.5.3 Kinematic derivation of the PV-substance flux 658
36.5.4 Potential vorticity flux vector 659
36.6 Exercises **661**

36.1 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 34.2 and 34.4), and then we generalize to a baroclinic fluid.

36.1.1 Perfect barotropic fluid

Consider a perfect barotropic fluid. As for the shallow water discussion in Section 33.3.3, we apply Kelvin's circulation theorem (Section 34.2.2) to an infinitesimal material circuit within the fluid

$$\frac{\mathrm{D}}{\mathrm{D}t}(\boldsymbol{\omega}_{\mathrm{a}} \cdot \hat{\boldsymbol{n}}\, \delta\mathcal{S}) = 0, \tag{36.1}$$

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

$$\frac{\mathrm{D}C}{\mathrm{D}t} = 0. \tag{36.2}$$

36.1.2 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 36.1, 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

$$\delta V = \delta\mathcal{S}\, \delta h, \tag{36.3}$$

where δh is the distance between the isosurfaces. The unit normal direction orienting the area $\delta\mathcal{S}$ is given by

$$\hat{\boldsymbol{n}} = \frac{\nabla C}{|\nabla C|}. \tag{36.4}$$

The separation between the two isosurfaces is related to the increment δC through

$$\delta C = \nabla C \cdot \delta\boldsymbol{x} = |\nabla C|\, \hat{\boldsymbol{n}} \cdot \delta\boldsymbol{x} = |\nabla C|\, \delta h. \tag{36.5}$$

Note that we can write this result in the equivalent manner

$$\delta C = |\nabla C|\,\delta h = (\hat{\boldsymbol{n}}\cdot\nabla C)\,\delta h = \frac{\delta C}{\delta n}\,\delta h. \tag{36.6}$$

Hence, the distance (or thickness) between the two isosurfaces is

$$\delta h = \frac{\delta C}{|\nabla C|}. \tag{36.7}$$

This equation has a straightforward geometric interpretation indicated in Figure 36.1. Namely, the geometric separation between the two isosurfaces is reduced in regions of strong scalar gradients and increased in regions of weak gradients.

Figure 36.1: *Illustrating the geometry of a cylindrical region of fluid between two iso-surfaces of a scalar field C. The volume of the region is $\delta V = \delta\mathcal{S}\,\delta h$, with δh the thickness and $\delta\mathcal{S}$ the area. By convention, the normal vector, $\hat{\boldsymbol{n}}$, points towards larger values of C. Notably, if C is a material invariant so that $\mathrm{D}C/\mathrm{D}t = 0$, then so is its infinitesimal increment, $\mathrm{D}(\delta C)/\mathrm{D}t = 0$. As per equation (36.7), the geometric thickness between the isosurfaces is related to the scalar field increment by $\delta h = \delta C/|\nabla C|$, so that the stronger 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 $\hat{\boldsymbol{n}}\cdot\boldsymbol{B} = 0$, with $\boldsymbol{B} = (\nabla\rho\wedge\nabla p)/\rho^2$ the baroclinicity.*

36.1.3 Material invariance

We now have the necessary pieces in place to write

$$\boldsymbol{\omega}_{\mathrm{a}}\cdot\hat{\boldsymbol{n}}\,\delta\mathcal{S} = \frac{\boldsymbol{\omega}_{\mathrm{a}}\cdot\nabla C}{|\nabla C|}\,\delta\mathcal{S} \tag{36.8a}$$

$$= \frac{\boldsymbol{\omega}_{\mathrm{a}}\cdot\nabla C}{|\nabla C|}\,\frac{\delta V}{\delta h} \tag{36.8b}$$

$$= (\boldsymbol{\omega}_{\mathrm{a}}\cdot\nabla C)\,\frac{\delta V}{\delta C} \tag{36.8c}$$

$$= \frac{\boldsymbol{\omega}_{\mathrm{a}}\cdot\nabla C}{\rho}\,\frac{\rho\,\delta V}{\delta C}. \tag{36.8d}$$

Mass is materially invariant so that

$$\frac{\mathrm{D}\,(\rho\,\delta V)}{\mathrm{D}t} = 0. \tag{36.9}$$

Likewise, by assumption C is materially invariant so that the increment between two C isosurfaces is materially invariant

$$\frac{\mathrm{D}(\delta C)}{\mathrm{D}t} = 0. \tag{36.10}$$

Bringing these elements into Kelvin's circulation theorem (36.1) leads us to conclude that the potential vorticity, Q, is also materially invariant

$$Q = \frac{\boldsymbol{\omega}_{\mathrm{a}} \cdot \nabla C}{\rho} = \frac{\nabla \cdot (\boldsymbol{\omega}_{\mathrm{a}} C)}{\rho} \Longrightarrow \frac{\mathrm{D}Q}{\mathrm{D}t} = 0. \tag{36.11}$$

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 36.4 and layer integrated PV Section 37.3.

36.1.4 Perfect baroclinic fluid

Consider the case of a perfect baroclinic fluid, in which Kelvin's circulation theorem for an infinitesimal circuit takes the form

$$\frac{\mathrm{D}}{\mathrm{D}t}(\boldsymbol{\omega}_{\mathrm{a}} \cdot \hat{\boldsymbol{n}}\, \delta\mathcal{S}) = \boldsymbol{B} \cdot \hat{\boldsymbol{n}}\, \delta\mathcal{S}, \tag{36.12}$$

where the source on the right hand side arises from the baroclinicity vector discussed in Sections 34.2 and 34.4.

$$\boldsymbol{B} = \frac{\nabla\rho \wedge \nabla p}{\rho^2}. \tag{36.13}$$

Now assume there exists a materially invariant scalar, θ, that annihilates the baroclinicity vector as in Figure 36.2, so that

$$\boldsymbol{B} \cdot \hat{\boldsymbol{n}} = \frac{\boldsymbol{B} \cdot \nabla\theta}{|\nabla\theta|} = 0. \tag{36.14}$$

In that case, the barotropic derivation detailed earlier in this section follows for the baroclinic case so that PV remains materially invariant

$$\frac{\mathrm{D}Q}{\mathrm{D}t} = 0 \qquad \text{where} \qquad Q = \frac{\boldsymbol{\omega}_{\mathrm{a}} \cdot \nabla\theta}{\rho}. \tag{36.15}$$

Figure 36.2: *Material invariance of PV is ensured for perfect fluids that allow for a materially invariant scalar to annihilate baroclinicity, meaning that $\hat{\boldsymbol{n}} \cdot \boldsymbol{B} = 0$ where $\hat{\boldsymbol{n}} = \nabla\theta/|\nabla\theta|$. 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 37.2.

36.1.5 Some general remarks

Perfect fluid PV material invariance ↔ Kelvin's circulation theorem

Kelvin's circulation theorem from Section 34.2.2 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 (36.11) is, on first glance, quite distinct from the shallow water expression $Q = (\zeta + f)/h$ explored in Chapter 33 (see equation (33.26)). 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 from physical oceanography.

Motivating the adjective "potential"

Write the PV in the form

$$Q = \frac{\boldsymbol{\omega}_{\mathrm{a}} \cdot \hat{\boldsymbol{n}}}{\rho} \, |\nabla\theta| \qquad \text{with} \qquad \hat{\boldsymbol{n}} = \frac{\nabla\theta}{|\nabla\theta|}. \tag{36.16}$$

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 (as well as the thermodynamic tracer, θ), In contrast, material tracers (e.g., salinity) and thermodynamic tracers (e.g., θ), are properties of the fluid and they do not directly depend on velocity, even though they are affected by velocity through advection.

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.

36.2 PV and the seawater equation of state

As seen in Section 36.1, 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 sewater 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.

36.2.1 Baroclinicity vector

Recall the baroclinicty vector given by (Sections 34.2 and 34.4)

$$\boldsymbol{B} = \frac{\nabla\rho \wedge \nabla p}{\rho^2}. \tag{36.17}$$

If we take the *in situ* density as the scalar field to define PV, then $\boldsymbol{B} \cdot \nabla\rho = 0$. 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

$$\frac{\mathrm{D}\rho}{\mathrm{D}t} = \frac{\partial\rho}{\partial S}\frac{\mathrm{D}S}{\mathrm{D}t} + \frac{\partial\rho}{\partial\theta}\frac{\mathrm{D}\theta}{\mathrm{D}t} + \frac{\partial\rho}{\partial p}\frac{\mathrm{D}p}{\mathrm{D}t}. \tag{36.18}$$

Even when salinity and potential temperature are materially constant, $\mathrm{D}S/\mathrm{D}t = 0$ and $\mathrm{D}\theta/\mathrm{D}t = 0$, the *in situ* density has a nonzero material time evolution due to mechanical effects leading to pressure changes, $\mathrm{D}p/\mathrm{D}t \neq 0$. Hence, *in situ* density is not an appropriate scalar for use in developing a materially invariant PV.

36.2.2 PV based on potential density

Potential density is commonly used in oceanography (see Section 21.3.5), with potential density the *in situ* density referenced to a chosen pressure.[1] We write potential density as

$$\sigma(S, \theta) = \rho(S, \theta, p = p_{\mathrm{R}}). \tag{36.19}$$

Potential density provides a global measure of buoyancy, with accuracy of the measure compromised in the presence of a nonlinear equation of state (Section 21.6).

The material time derivative of potential density is

$$\frac{\mathrm{D}\sigma}{\mathrm{D}t} = \frac{\partial\sigma}{\partial S}\frac{\mathrm{D}S}{\mathrm{D}t} + \frac{\partial\sigma}{\partial\theta}\frac{\mathrm{D}\theta}{\mathrm{D}t}, \tag{36.20}$$

[1] Oceanographers 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.

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

$$\rho^2 \boldsymbol{B} \cdot \nabla\sigma = (\nabla\rho \wedge \nabla p) \cdot \nabla\sigma \tag{36.21a}$$
$$= (\nabla\sigma \wedge \nabla\rho) \cdot \nabla p \tag{36.21b}$$
$$= [(\sigma_S \, \nabla S + \sigma_\theta \, \nabla\theta) \wedge (\rho_S \, \nabla S + \rho_\theta \, \nabla\theta + \rho_p \, \nabla p)] \cdot \nabla p \tag{36.21c}$$
$$= [(\sigma_S \, \nabla S + \sigma_\theta \, \nabla\theta) \wedge (\rho_S \, \nabla S + \rho_\theta \, \nabla\theta)] \cdot \nabla p \tag{36.21d}$$
$$= [\sigma_S \, \nabla S \wedge \rho_\theta \, \nabla\theta + \sigma_\theta \, \nabla\theta \wedge \rho_S \, \nabla S] \cdot \nabla p \tag{36.21e}$$
$$= (\sigma_S \, \rho_\theta - \sigma_\theta \, \rho_S) \, (\nabla S \wedge \nabla\theta) \cdot \nabla p, \tag{36.21f}$$

where we used the shorthand notation for partial derivatives

$$\rho_S = \frac{\partial\rho}{\partial S} \qquad \sigma_S = \frac{\partial\sigma}{\partial S} \tag{36.22a}$$
$$\rho_\theta = \frac{\partial\rho}{\partial\theta} \qquad \sigma_\theta = \frac{\partial\sigma}{\partial\theta}. \tag{36.22b}$$

Note that the triple product, $(\nabla S \wedge \nabla\theta) \cdot \nabla p$, also appears in the discussion of neutral helicity in Section 21.7 (see equation (21.57)). Equation (36.21f) allows us to identify cases where the baroclinicity vector is annihilated, $\boldsymbol{B} \cdot \nabla\sigma = 0$, 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 $\boldsymbol{B} \cdot \nabla\sigma = 0$.

- 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 $\boldsymbol{B} \cdot \nabla\sigma \neq 0$. Nonetheless, the baroclinicity vector is annihilated if the *in situ* density has a pressure dependence that is additive, so that we can write

$$\rho(S, \theta, p) = \sigma(S, \theta) + F(p) - F(p_{\mathrm{R}}) \Rightarrow \sigma_S \, \rho_\theta - \sigma_\theta \, \rho_S = 0, \tag{36.23}$$

  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 (36.23). 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.

36.2.3 An example EOS admitting a materially invariant PV

An explicit realization of the equation of state (36.23) 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

$$\rho(S, \theta, p) \approx \sigma(S, \theta) + \underbrace{(p - p_{\mathrm{R}}) \left[\frac{\partial\rho}{\partial p}\right]_{S=S_{\mathrm{R}}, \theta=\theta_{\mathrm{R}}, p=p_{\mathrm{R}}} + H.O.T.}_{F(p)-F(p_{\mathrm{R}})} \tag{36.24}$$

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

$$c_{\mathrm{s}}^{-2} = \frac{\partial \rho}{\partial p} \approx \text{constant}. \tag{36.25}$$

In this case, the equation of state takes the form

$$\rho(S, \theta, p) \approx \sigma(S, \theta) + \frac{p - p_{\mathrm{R}}}{c_{\mathrm{s}}^2}, \tag{36.26}$$

where again $\sigma(S, \theta) = \rho(S, \theta, p_{\mathrm{R}})$ is the potential density referenced to $p = p_{\mathrm{R}}$.

36.2.4 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 49.4.4).

36.3 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 37, 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 find it more convenient to pursue an algebraic approach that starts from the vorticity equation (34.30)

$$\rho \frac{\mathrm{D}(\boldsymbol{\omega}_{\mathrm{a}}/\rho)}{\mathrm{D}t} = (\boldsymbol{\omega}_{\mathrm{a}} \cdot \nabla)\, \boldsymbol{v} + \boldsymbol{B} + \nabla \wedge \boldsymbol{F}, \tag{36.27}$$

where $\boldsymbol{F}$ is the frictional acceleration vector and $\boldsymbol{B}$ the baroclinicity vector. Furthermore, we introduce a non-conservative scalar field

$$\frac{\mathrm{D}C}{\mathrm{D}t} = \dot{C}, \tag{36.28}$$

with $\dot{C}$ arising from diffusion or other irreversible processes.

As part of the manipulations we make use of the identity

$$(\boldsymbol{\omega}_{\mathrm{a}} \cdot \nabla) \frac{\mathrm{D}C}{\mathrm{D}t} = \boldsymbol{\omega}_{\mathrm{a}} \cdot \frac{\mathrm{D}\,(\nabla C)}{\mathrm{D}t} + [(\boldsymbol{\omega}_{\mathrm{a}} \cdot \nabla)\, \boldsymbol{v}] \cdot \nabla C, \tag{36.29}$$

which is readily proven by expanding terms and assuming Cartesian coordinates. Rearrangement, and use of the scalar equation (36.28), leads to

$$\boldsymbol{\omega}_{\mathrm{a}} \cdot \frac{\mathrm{D}\,(\nabla C)}{\mathrm{D}t} = (\boldsymbol{\omega}_{\mathrm{a}} \cdot \nabla)\, \dot{C} - [(\boldsymbol{\omega}_{\mathrm{a}} \cdot \nabla)\, \boldsymbol{v}] \cdot \nabla C. \tag{36.30}$$

Now project the vorticity equation (36.27) onto the direction normal to the C isosurfaces

$$\rho\, \nabla C \cdot \frac{\mathrm{D}(\boldsymbol{\omega}_{\mathrm{a}}/\rho)}{\mathrm{D}t} = \nabla C \cdot [(\boldsymbol{\omega}_{\mathrm{a}} \cdot \nabla)\, \boldsymbol{v}] + \nabla C \cdot (\boldsymbol{B} + \nabla \wedge \boldsymbol{F}). \tag{36.31}$$

The sum of equations (36.30) and (36.31) leads to

$$\rho \frac{\mathrm{D}(\nabla C \cdot \boldsymbol{\omega}_{\mathrm{a}}/\rho)}{\mathrm{D}t} = (\boldsymbol{\omega}_{\mathrm{a}} \cdot \nabla)\, \dot{C} + \nabla C \cdot (\boldsymbol{B} + \nabla \wedge \boldsymbol{F}). \tag{36.32}$$

This equation is general, and thus is applies to any scalar field.

To simplify the source terms on the right hand side of equation (36.32), follow the discussion from Section 36.1.4 by assuming we can find 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, or potential density. We thus have

$$\nabla\theta \cdot \boldsymbol{B} = 0, \tag{36.33}$$

which in turn leads to the potential vorticity equation in the presence of irreversible processes

$$\rho \frac{\mathrm{D}Q}{\mathrm{D}t} = (\boldsymbol{\omega}_{\mathrm{a}} \cdot \nabla)\, \dot{\theta} + \nabla\theta \cdot (\nabla \wedge \boldsymbol{F}), \tag{36.34}$$

where the entropic-PV is again given by

$$Q = \frac{\boldsymbol{\omega}_{\mathrm{a}} \cdot \nabla\theta}{\rho}. \tag{36.35}$$

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 (36.15) derived for the perfect fluid using Kelvin's circulation theorem. In their presence, the PV of a fluid particle 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.

36.4 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 $\rho\, \mathrm{D}Q/\mathrm{D}t = 0$ having its Eulerian flux-form expression

$$\frac{\partial(\rho\, Q)}{\partial t} + \nabla \cdot (\rho\, \boldsymbol{v}\, Q) = 0 \qquad \text{perfect fluid.} \tag{36.36}$$

Following the formalism established for material tracers in Section 15.3, the flux form local conservation law (36.36) 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.

[2] In Section 37.2 we study what happens when no such scalar exists.

36.4.1 Deriving the Eulerian flux-form PV budget

To transform the material evolution equation (36.34) into a flux-form Eulerian equation we make use of the following identities

$$\frac{\mathrm{D}}{\mathrm{D}t} = \frac{\partial}{\partial t} + \boldsymbol{v}\cdot\nabla \qquad \text{relating material and Eulerian time changes} \tag{36.37a}$$

$$\frac{\mathrm{D}\rho}{\mathrm{D}t} = -\rho\,\nabla\cdot\boldsymbol{v} \qquad \text{mass conservation} \tag{36.37b}$$

$$(\boldsymbol{\omega}_{\mathrm{a}}\cdot\nabla)\,\dot{\theta} = \nabla\cdot(\boldsymbol{\omega}_{\mathrm{a}}\,\dot{\theta}) \qquad \nabla\cdot\boldsymbol{\omega}_{\mathrm{a}} = 0 \tag{36.37c}$$

$$\nabla\theta\cdot(\nabla\wedge\boldsymbol{F}) = \nabla\cdot(\boldsymbol{F}\wedge\nabla\theta) \qquad \text{divergence of curl vanishes.} \tag{36.37d}$$

The identity (36.37d) follows from

$$\nabla\theta\cdot(\nabla\wedge\boldsymbol{F}) = \nabla\cdot(\theta\,\nabla\wedge\boldsymbol{F}) = \nabla\cdot[\nabla\wedge(\theta\,\boldsymbol{F}) - \nabla\theta\wedge\boldsymbol{F}] = \nabla\cdot(\boldsymbol{F}\wedge\nabla\theta), \tag{36.38}$$

where we set the divergence of a curl to zero to reach the first and third equalities. These identities then lead to the material evolution equation

$$\rho\,\frac{\mathrm{D}Q}{\mathrm{D}t} = \nabla\cdot[\boldsymbol{\omega}_{\mathrm{a}}\,\dot{\theta} + \boldsymbol{F}\wedge\nabla\theta]. \tag{36.39}$$

Now converting the material time derivative into its Eulerian expression, and making use of mass conservation, renders the flux-form Eulerian PV budget equation

$$\frac{\partial(\rho\,Q)}{\partial t} + \nabla\cdot[\rho\,Q\,\boldsymbol{v} - \boldsymbol{\omega}_{\mathrm{a}}\,\dot{\theta} - \boldsymbol{F}\wedge\nabla\theta] = 0. \tag{36.40}$$

36.4.2 PV-substance and the PV flux

The budget equation (36.40) says that the density-weighted potential vorticity

$$\rho\,Q = \boldsymbol{\omega}_{\mathrm{a}}\cdot\nabla\theta \tag{36.41}$$

has a local time tendency determined by the convergence of the PV flux vector

$$\frac{\partial(\rho\,Q)}{\partial t} = -\nabla\cdot\boldsymbol{J}_{\mathrm{Q}} \qquad \text{with} \qquad \boldsymbol{J}_{\mathrm{Q}} = \rho\,Q\,\boldsymbol{v} - \boldsymbol{\omega}_{\mathrm{a}}\,\dot{\theta} - \boldsymbol{F}\wedge\nabla\theta. \tag{36.42}$$

The budget (36.40) follows a form similar to material tracers detailed in Chapter 15, though with some specific terms in the flux vector, $\boldsymbol{J}_{\mathrm{Q}}$. The correspondence suggests that one consider equation (36.40) as the local budget for *PV-substance*, with Q the concentration of PV-substance and $\boldsymbol{J}_{\mathrm{Q}}$ its flux. This interpretation is pursued further in Chapter 37 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 (36.42) 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 36.3).

The time tendency for PV-substance, $\partial_t(\rho\,Q)$, is unchanged by adding the curl of a vector to $\boldsymbol{J}_{\mathrm{Q}}$. This ambiguity manifests a *gauge freedom* afforded the PV-substance flux vector. We offer a general discussion of gauge freedom in Section 16.5.2 and provide more specifics for PV in Chapter 37.

Figure 36.3: *The contribution from friction to the PV flux is given by* $\boldsymbol{J}_{\text{friction}} = \nabla\theta \wedge \boldsymbol{F}$*, 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* $\nabla\theta \wedge \boldsymbol{F}$ *(i.e., zero divergence), then friction does not contribute to PV evolution.*

36.5 PV budget for a hydrostatic Boussinesq fluid

Building on the vorticity budget developed in Section 34.7, we here develop the potential vorticity substance budgets for a hydrostatic Boussinesq fluid in the presence of diabatic and frictional forcing.

36.5.1 Potential vorticity

Potential vorticity evolves in the absence of baroclinicity, which is eliminated from the vorticity equation by projecting the absolute vorticity onto the direction normal to buoyancy surfaces

$$\nabla b \cdot \frac{\mathrm{D}\boldsymbol{\omega}_{\mathrm{a}}^{\mathrm{hy}}}{\mathrm{D}t} = \nabla b \cdot [(\boldsymbol{\omega}_{\mathrm{a}}^{\mathrm{hy}} \cdot \nabla)\, \boldsymbol{v}] + \nabla b \cdot (\nabla \wedge \boldsymbol{F}), \tag{36.43}$$

where we used

$$\nabla b \cdot (\nabla \wedge \hat{\boldsymbol{z}}\, b) = 0. \tag{36.44}$$

We next make use of the identity

$$\frac{\mathrm{D}(\partial b/\partial x^i)}{\mathrm{D}t} = \frac{\partial}{\partial x^i}\left[\frac{\mathrm{D}b}{\mathrm{D}t}\right] - \nabla b \cdot \frac{\partial \boldsymbol{v}}{\partial x^i} \tag{36.45a}$$

$$= \frac{\partial \dot{b}}{\partial x^i} - \nabla b \cdot \frac{\partial \boldsymbol{v}}{\partial x^i}, \tag{36.45b}$$

so that

$$\boldsymbol{\omega}_{\mathrm{a}}^{\mathrm{hy}} \cdot \left[\frac{\mathrm{D}\,\nabla b}{\mathrm{D}t}\right] = \boldsymbol{\omega}_{\mathrm{a}}^{\mathrm{hy}} \cdot \nabla \dot{b} - \nabla b \cdot [(\boldsymbol{\omega}_{\mathrm{a}}^{\mathrm{hy}} \cdot \nabla)\, \boldsymbol{v}]. \tag{36.46}$$

Making use of this result in equation (36.43) renders

$$\nabla b \cdot \frac{\mathrm{D}\boldsymbol{\omega}_{\mathrm{a}}^{\mathrm{hy}}}{\mathrm{D}t} + \boldsymbol{\omega}_{\mathrm{a}}^{\mathrm{hy}} \cdot \frac{\mathrm{D}\nabla b}{\mathrm{D}t} = \boldsymbol{\omega}_{\mathrm{a}}^{\mathrm{hy}} \cdot \nabla \dot{b} + \nabla b \cdot (\nabla \wedge \boldsymbol{F}), \tag{36.47}$$

which leads to

$$\frac{\mathrm{D}Q}{\mathrm{D}t} = \boldsymbol{\omega}_{\mathrm{a}}^{\mathrm{hy}} \cdot \nabla \dot{b} + \nabla b \cdot (\nabla \wedge \boldsymbol{F}) \tag{36.48}$$

where

$$Q = \boldsymbol{\omega}_{\mathrm{a}}^{\mathrm{hy}} \cdot \nabla b = \boldsymbol{\omega}^{\mathrm{hy}} \cdot \nabla b + f\,\frac{\partial b}{\partial z} \tag{36.49}$$

is the potential vorticity for a rotating hydrostatic Boussinesq fluid. Potential vorticity is materially invariant for the inviscid and adiabatic case, in which $\boldsymbol{F} = 0$ and $\dot{b} = 0$.

It is sometimes useful to split the hydrostatic vorticity into its vertical and horizontal terms as per equation (34.85). In this way, PV takes on the form

$$Q = \frac{\partial u}{\partial z}\frac{\partial b}{\partial y} - \frac{\partial v}{\partial z}\frac{\partial b}{\partial x} + (\zeta + f)\,\frac{\partial b}{\partial z} = \hat{\boldsymbol{z}} \cdot \left[\frac{\partial \boldsymbol{u}}{\partial z} \wedge \nabla b\right] + (\zeta + f)\,\frac{\partial b}{\partial z}. \tag{36.50}$$

36.5.2 Potential vorticity flux vector

The material form of the PV equation (36.48) is converted into its Eulerian flux-form via

$$\frac{\partial Q}{\partial t} + \nabla \cdot (\boldsymbol{v}\, Q) = \boldsymbol{\omega}_{\mathrm{a}}^{\mathrm{hy}} \cdot \nabla \dot{b} + \nabla b \cdot (\nabla \wedge \boldsymbol{F}) \tag{36.51a}$$

$$= \nabla \cdot [\dot{b}\, \boldsymbol{\omega}_{\mathrm{a}}^{\mathrm{hy}} + b\, (\nabla \wedge \boldsymbol{F})] \tag{36.51b}$$

$$= \nabla \cdot [\dot{b}\, \boldsymbol{\omega}_{\mathrm{a}}^{\mathrm{hy}} + \nabla \wedge (b\, \boldsymbol{F}) - \nabla b \wedge \boldsymbol{F}] \tag{36.51c}$$

$$= \nabla \cdot [\dot{b}\, \boldsymbol{\omega}_{\mathrm{a}}^{\mathrm{hy}} - \nabla b \wedge \boldsymbol{F}], \tag{36.51d}$$

where we used

$$\nabla \cdot \boldsymbol{v} = 0 \tag{36.52a}$$

$$\nabla \cdot \boldsymbol{\omega}_{\mathrm{a}}^{\mathrm{hy}} = 0 \tag{36.52b}$$

$$\nabla \cdot (\nabla \wedge \boldsymbol{F}) = 0 \tag{36.52c}$$

$$\nabla \cdot [\nabla \wedge (b\, \boldsymbol{F})] = 0. \tag{36.52d}$$

The conservation equation (36.51d) allows us to identify a potential vorticity flux vector

$$\boldsymbol{J}_{\mathrm{Q}} = \boldsymbol{v}\, Q - \dot{b}\, \boldsymbol{\omega}_{\mathrm{a}}^{\mathrm{hy}} + \nabla b \wedge \boldsymbol{F} + \nabla \wedge \boldsymbol{A}, \tag{36.53}$$

so that the PV equation takes the form

$$\frac{\partial Q}{\partial t} + \nabla \cdot \boldsymbol{J}_{\mathrm{Q}} = 0. \tag{36.54}$$

The potential vorticity flux (36.53) is comprised of an advective term

$$\boldsymbol{J}_{\mathrm{advective}} = \boldsymbol{v}\, Q, \tag{36.55}$$

and non-advective terms arising from diabatic and frictional forcing,

$$\boldsymbol{J}_{\text{non-advective}} = -\dot{b}\, \boldsymbol{\omega}_{\mathrm{a}}^{\mathrm{hy}} + \nabla b \wedge \boldsymbol{F}, \tag{36.56}$$

as well as a gauge term,

$$\boldsymbol{J}_{\mathrm{gauge}} = \nabla \wedge \boldsymbol{A}. \tag{36.57}$$

36.5.3 Kinematic derivation of the PV-substance flux

Following the discussion in Section 37.1.2, we consider a purely kinematic derivation of the PV flux vector. For that purpose write the hydrostatic Boussinesq PV in the form

$$Q = \boldsymbol{\omega}_{\mathrm{a}}^{\mathrm{hy}} \cdot \nabla b = \nabla \cdot (b\, \boldsymbol{\omega}_{\mathrm{a}}^{\mathrm{hy}}), \tag{36.58}$$

which follows since $\nabla \cdot \boldsymbol{\omega}_{\mathrm{a}}^{\mathrm{hy}} = 0$. Taking the Eulerian time derivative then leads to

$$\frac{\partial Q}{\partial t} = \nabla \cdot \left[\frac{\partial b}{\partial t} \boldsymbol{\omega}_{\mathrm{a}}^{\mathrm{hy}} + b \, \frac{\partial \boldsymbol{\omega}_{\mathrm{a}}^{\mathrm{hy}}}{\partial t} \right] \tag{36.59a}$$

$$= \nabla \cdot \left[\frac{\partial b}{\partial t} \boldsymbol{\omega}_{\mathrm{a}}^{\mathrm{hy}} + \frac{\partial \boldsymbol{u}}{\partial t} \wedge \nabla b \right], \tag{36.59b}$$

which made use of the identities

$$\frac{\partial \boldsymbol{\omega}_{\mathrm{a}}^{\mathrm{hy}}}{\partial t} = \frac{\partial \boldsymbol{\omega}^{\mathrm{hy}}}{\partial t} = \frac{\partial (\nabla \wedge \boldsymbol{u})}{\partial t} \qquad \text{and} \qquad \nabla \cdot [\nabla \wedge (b \, \boldsymbol{u})] = 0. \tag{36.60}$$

The PV thus evolves according to

$$\frac{\partial Q}{\partial t} = -\nabla \cdot \widetilde{\boldsymbol{J}}_{\mathrm{Q}} \tag{36.61}$$

where the PV flux is given by

$$\widetilde{\boldsymbol{J}}_{\mathrm{Q}} = -\frac{\partial b}{\partial t} \boldsymbol{\omega}_{\mathrm{a}}^{\mathrm{hy}} - \frac{\partial \boldsymbol{u}}{\partial t} \wedge \nabla b. \tag{36.62}$$

This form of the PV flux manifests the impermeability property since

$$\hat{\boldsymbol{n}} \cdot (\widetilde{\boldsymbol{J}}_{\mathrm{Q}}/Q) = -\frac{1}{|\nabla b|} \frac{\partial b}{\partial t} = \hat{\boldsymbol{n}} \cdot \boldsymbol{v}_{b\perp}, \tag{36.63}$$

where $\hat{\boldsymbol{n}} = \nabla b \, |\nabla b|^{-1}$ is the unit normal for a buoyancy surface, and with $\boldsymbol{v}_{b\perp}$ the velocity of a point on the isopycnal that satisfies

$$\frac{\partial b}{\partial t} + \boldsymbol{v}_{b\perp} \cdot \nabla b = 0. \tag{36.64}$$

36.5.4 Potential vorticity flux vector

We introduced the Bernoulli potential and Bernoulli theorem in Section 23.4.4. *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 37.2.1.

Momentum equation

We 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 (34.91), and

expand the expressions for the perturbation pressure and the buoyancy

$$\frac{\partial \boldsymbol{u}}{\partial t} + (\boldsymbol{f} + \boldsymbol{\omega}^{\mathrm{hy}}) \wedge \boldsymbol{v} = -\nabla\,(\varphi + \boldsymbol{u}^2/2) + \hat{\boldsymbol{z}}\, b + \boldsymbol{F} \tag{36.65a}$$

$$= -\nabla\,(\boldsymbol{u}^2/2) - \frac{1}{\rho_0}\nabla(p - p_0) - \hat{\boldsymbol{z}}\left[\frac{g\,(\rho - \rho_0)}{\rho_0}\right] + \boldsymbol{F} \tag{36.65b}$$

$$= -\nabla\,(\boldsymbol{u}^2/2) - \frac{1}{\rho_0}\nabla p - \hat{\boldsymbol{z}}\,\frac{g\,\rho}{\rho_0} + \boldsymbol{F} \tag{36.65c}$$

$$= -\nabla\left[\frac{\boldsymbol{u}^2}{2} + \frac{p}{\rho_0}\right] - \hat{\boldsymbol{z}}\left[\frac{g\,\rho - g\,\rho_o + g\,\rho_0}{\rho_0}\right] + \boldsymbol{F} \tag{36.65d}$$

$$= -\nabla\left[\frac{\boldsymbol{u}^2}{2} + \frac{p}{\rho_0} + g\,z\right] - \hat{\boldsymbol{z}}\left[\frac{g\,(\rho - \rho_o)}{\rho_0}\right] + \boldsymbol{F} \tag{36.65e}$$

$$= -\nabla\,\mathcal{B} + \hat{\boldsymbol{z}}\, b + \boldsymbol{F}, \tag{36.65f}$$

where we introduced the Bernoulli potential for a hydrostatic and Boussinesq fluid

$$\mathcal{B} = \frac{\boldsymbol{u}^2}{2} + \frac{p}{\rho_0} + g\,z. \tag{36.66}$$

Note that for a non-Boussinesq fluid the internal energy appears in the Bernoulli function (Section 23.4.3).

Potential vorticity flux

The flux-form potential vorticity conservation statement remains as given by equation (36.51d), and the PV-substance flux is given by equation (36.53). 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 (36.65f) to have

$$\nabla b \wedge \frac{\partial \boldsymbol{u}}{\partial t} + \nabla b \wedge (\boldsymbol{\omega}_{\mathrm{a}}^{\mathrm{hy}} \wedge \boldsymbol{v}) = -\nabla b \wedge \nabla\mathcal{B} + \nabla b \wedge \hat{\boldsymbol{z}}\, b + \nabla b \wedge \boldsymbol{F}. \tag{36.67}$$

Make use of the identity

$$\nabla b \wedge (\boldsymbol{\omega}_{\mathrm{a}}^{\mathrm{hy}} \wedge \boldsymbol{v}) = (\nabla b \cdot \boldsymbol{v})\,\boldsymbol{\omega}_{\mathrm{a}}^{\mathrm{hy}} - (\boldsymbol{\omega}_{\mathrm{a}}^{\mathrm{hy}} \cdot \nabla b)\,\boldsymbol{v} \tag{36.68}$$

in equation (36.67) to render

$$(\boldsymbol{v} \cdot \nabla b)\,\boldsymbol{\omega}_{\mathrm{a}}^{\mathrm{hy}} - (\boldsymbol{\omega}_{\mathrm{a}}^{\mathrm{hy}} \cdot \nabla b)\,\boldsymbol{v} = -\nabla b \wedge \frac{\partial \boldsymbol{u}}{\partial t} - \nabla b \wedge \nabla\mathcal{B} + \nabla b \wedge \hat{\boldsymbol{z}}\, b + \nabla b \wedge \boldsymbol{F}. \tag{36.69}$$

Now write the flux given by equation (36.53) in the form

$$\boldsymbol{J}_{\mathrm{Q}} = \boldsymbol{v}\,Q - \dot{b}\,\boldsymbol{\omega}_{\mathrm{a}}^{\mathrm{hy}} + \nabla b \wedge \boldsymbol{F} + \nabla \wedge \boldsymbol{A} \tag{36.70a}$$

$$= \boldsymbol{v}\,(\boldsymbol{\omega}_{\mathrm{a}}^{\mathrm{hy}} \cdot \nabla b) - \left[\frac{\partial b}{\partial t} + \boldsymbol{v} \cdot \nabla b\right]\boldsymbol{\omega}_{\mathrm{a}}^{\mathrm{hy}} + \nabla b \wedge \boldsymbol{F} + \nabla \wedge \boldsymbol{A} \tag{36.70b}$$

$$= [\boldsymbol{v}\,(\boldsymbol{\omega}_{\mathrm{a}}^{\mathrm{hy}} \cdot \nabla b) - (\boldsymbol{v} \cdot \nabla b)\,\boldsymbol{\omega}_{\mathrm{a}}^{\mathrm{hy}}] - \frac{\partial b}{\partial t}\,\boldsymbol{\omega}_{\mathrm{a}}^{\mathrm{hy}} + \nabla b \wedge \boldsymbol{F} + \nabla \wedge \boldsymbol{A} \tag{36.70c}$$

$$= \left[\nabla b \wedge \frac{\partial \boldsymbol{u}}{\partial t} + \nabla b \wedge \nabla\mathcal{B} - \nabla b \wedge \hat{\boldsymbol{z}}\, b - \nabla b \wedge \boldsymbol{F}\right] - \frac{\partial b}{\partial t}\,\boldsymbol{\omega}_{\mathrm{a}}^{\mathrm{hy}} + \nabla b \wedge \boldsymbol{F} + \nabla \wedge \boldsymbol{A} \tag{36.70d}$$

$$= \nabla b \wedge \frac{\partial \boldsymbol{u}}{\partial t} - \frac{\partial b}{\partial t}\,\boldsymbol{\omega}_{\mathrm{a}}^{\mathrm{hy}} + \nabla b \wedge \nabla\mathcal{B} - \nabla b \wedge \hat{\boldsymbol{z}}\, b + \nabla \wedge \boldsymbol{A}. \tag{36.70e}$$

$$= \nabla b \wedge \left[\frac{\partial \boldsymbol{u}}{\partial t} + \nabla\mathcal{B}\right] - \frac{\partial b}{\partial t}\,\boldsymbol{\omega}_{\mathrm{a}}^{\mathrm{hy}} + \nabla \wedge \left(\boldsymbol{A} - \hat{\boldsymbol{z}}\, b^2/2\right). \tag{36.70f}$$

Choosing the gauge function according to

$$\boldsymbol{A} = \hat{\boldsymbol{z}}\,(b^2/2) \tag{36.71}$$

renders the flux vector

$$\boldsymbol{J}_{\mathrm{Q}} = \nabla b \wedge \left(\frac{\partial \boldsymbol{u}}{\partial t} + \nabla \mathcal{B}\right) - \frac{\partial b}{\partial t}\,\boldsymbol{\omega}_{\mathrm{a}}^{\mathrm{hy}} \tag{36.72a}$$

$$= \boldsymbol{v}\,Q - \dot{b}\,\boldsymbol{\omega}_{\mathrm{a}}^{\mathrm{hy}} + \nabla b \wedge \boldsymbol{F}. \tag{36.72b}$$

The second equality reintroduced the first form of the PV flux given by equation (36.53); made use of the gauge choice (36.71); and furthermore dropped the non-divergent term

$$\nabla b \wedge b\,\hat{\boldsymbol{z}} = (1/2)\,\nabla b^2 \wedge \hat{\boldsymbol{z}} \qquad \text{with} \qquad \nabla \cdot (\nabla b \wedge b\,\hat{\boldsymbol{z}}) = 0. \tag{36.73}$$

In the steady state, the PV flux (36.72a) reduces to

$$\boldsymbol{J}_{\mathrm{Q}}^{\mathrm{ss}} = \nabla b \wedge \nabla \mathcal{B}. \tag{36.74}$$

Hence, the steady state potential vorticity flux is aligned with the intersection of surfaces of constant buoyancy and Bernoulli potential

$$\nabla b \cdot \boldsymbol{J}_{\mathrm{Q}}^{\mathrm{ss}} = 0 \qquad \text{and} \qquad \nabla \mathcal{B} \cdot \boldsymbol{J}_{\mathrm{Q}}^{\mathrm{ss}} = 0. \tag{36.75}$$

This result is the Boussinesq/hydrostatic form of the more general result derived in Section 37.2.3.

36.6 Exercises

EXERCISE 36.1: PV FOR A PERFECT NON-HYDROSTATIC BOUSSINESQ FLUID
Consider a perfect rotating Boussinesq fluid whose governing equations are given by

$$\frac{\mathrm{D}\boldsymbol{v}}{\mathrm{D}t} + f\,(\hat{\boldsymbol{z}} \wedge \boldsymbol{v}) = -\nabla\varphi + b\,\hat{\boldsymbol{z}} \tag{36.76}$$

$$\nabla \cdot \boldsymbol{v} = 0 \tag{36.77}$$

$$\frac{\mathrm{D}b}{\mathrm{D}t} = 0 \tag{36.78}$$

$$b = -\frac{g\,(\rho - \rho_0)}{\rho_0} = g\,\alpha\,\theta, \tag{36.79}$$

where θ is the potential temperature and $\alpha > 0$ is a constant thermal expansion coefficient. Further details are provided in Section 25.2.6. Some of this exercise follows the hydrostatic Boussinesq discussion in Section 36.5, though they differ in important places so be careful!

(a) Derive the equation for the material time evolution of potential vorticity in this fluid system. Show all steps in the derivation.

(b) The baroclinicity vector appearing in the Boussinesq vorticity equation is $\boldsymbol{B} = \nabla \wedge \hat{\boldsymbol{z}}\,b = \nabla b \wedge \hat{\boldsymbol{z}}$ (e.g., see Section 34.7.1 if you did not derive this in part (a) above). Show that this vector results upon making the Boussinesq approximation to the compressible fluid expression $\boldsymbol{B} = (\nabla\rho \wedge \nabla p)/\rho^2$.

(c) Show that the vertical portion of Q^{bouss} can be written

$$Q^{\text{bouss}}_{\text{vert}} = (\zeta + f)\, N^2 \tag{36.80}$$

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 21.4.4). Hint: this is a rather trivial question.

(d) If flow maintains geostrophic and thermal wind balance, show that the horizontal portion of Q^{bouss} can be written

$$Q^{\text{bouss}}_{\text{horz}} = \boldsymbol{\omega} \cdot \nabla_z b \approx -f^{-1} |\nabla_z b|^2. \tag{36.81}$$

Hint: recall that for geostrophic flow, the vertical velocity is much smaller than horizontal.

EXERCISE 36.2: PV FOR DIABATIC AND FRICTIONAL NON-HYDROSTATIC BOUSSINESQ FLOW

Reconsider Exercise 36.1 in the presence of irreversible friction and buoyancy sources so that the governing equations are

$$\frac{\mathrm{D}\boldsymbol{v}}{\mathrm{D}t} + f\,(\hat{\boldsymbol{z}} \wedge \boldsymbol{v}) = -\nabla\varphi + b\,\hat{\boldsymbol{z}} + \boldsymbol{F} \tag{36.82}$$

$$\nabla \cdot \boldsymbol{v} = 0 \tag{36.83}$$

$$\frac{\mathrm{D}b}{\mathrm{D}t} = \dot{b} \tag{36.84}$$

$$b = -\frac{g\,(\rho - \rho_0)}{\rho_0} = g\,\alpha\,\theta, \tag{36.85}$$

where $\dot{\theta}$ is a diabatic heating source/sink, α is a constant thermal expansion coefficient, and $\boldsymbol{F}$ is a non-conservative acceleration such as from friction or boundary stresses.

(a) Derive the equation for the material time evolution of potential vorticity in this fluid system, including the irreversible contributions from friction and heating.

(b) Derive an equation for the potential vorticity time tendency (i.e., Eulerian time derivative), written in the form

$$\frac{\partial Q}{\partial t} = -\nabla \cdot \boldsymbol{J}_{\mathrm{Q}}. \tag{36.86}$$

What is the PV flux $\boldsymbol{J}_{\mathrm{Q}}$? 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 $\boldsymbol{J}_{\mathrm{Q}}$.

(c) A common diabatic process is written in the form of a damping source

$$\dot{b} = -\mu\,(b - b^*), \tag{36.87}$$

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 $Q^* = \boldsymbol{\omega}_{\mathrm{a}} \cdot \nabla b^*$.

(d) A form for the friction operator is given by Rayleigh drag

$$\boldsymbol{F} = -\lambda\,\boldsymbol{v}, \tag{36.88}$$

with λ 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, $Q^{\text{pg}} = f\,N^2$, where $N^2 = \partial b/\partial z$ is the squared buoyancy frequency.

(e) Discuss the balance needed between forcing terms in $\boldsymbol{J}_{\mathrm{Q}}$ 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.

37

Potential vorticity budgets and impermeability

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 37.3, ρQ remains constant within an isentrope unless the isentrope intersects a boundary.

READER'S GUIDE FOR THIS CHAPTER

This chapter builds from the potential vorticity mechanics introduced in Chapter 36. It is an essential read for those intersted in potential vorticity theory and potential vorticity budgets.

37.1 Variations on the impermeability theorem **666**
37.1.1 Impermeability for the Haynes-McIntyre PV flux 666
37.1.2 A kinematic derivation of impermeability 667
37.1.3 Impermeability for all components of absolute vorticity 668
37.1.4 Comments 669
37.2 Impermeability theorem for a realistic seawater EOS **669**
37.2.1 Ocean PV in terms of potential density 669
37.2.2 A modified PV-substance flux 670
37.2.3 Integral constraints for steady state 671
37.2.4 Further study 672
37.3 Integrated PV **672**
37.3.1 Region bounded by a buoyancy surface 672
37.3.2 Region bounded by two buoyancy surfaces 673
37.3.3 Region bounded by land and a buoyancy surface 674
37.3.4 An ocean layer that outcrops 675
37.4 Layer integrated PV budget **675**
37.4.1 Layer integrated budget 676
37.4.2 Impermeability across interior layer interfaces 676
37.4.3 PV flux at the land-sea boundary 677
37.4.4 PV flux at the air-sea boundary 678
37.4.5 Thought experiments 679

37.4.6 Is there a preferred form of the PV-substance flux? 679

37.1 Variations on the impermeability theorem

In this section we derive the impermeabity 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.

37.1.1 Impermeability for the Haynes-McIntyre PV flux

Recall the Eulerian flux-form evolution equation for PV-substance (36.40)

$$\frac{\partial(\rho\, Q)}{\partial t} + \nabla\cdot \boldsymbol{J}_{\mathrm{Q}} = 0 \qquad \text{with} \qquad \boldsymbol{J}_{\mathrm{Q}} = \rho\, \boldsymbol{v}\, Q - \dot{\theta}\, \boldsymbol{\omega}_{\mathrm{a}} + \nabla\theta \wedge \boldsymbol{F}, \tag{37.1}$$

with the PV-substance flux vector, $\boldsymbol{J}_{\mathrm{Q}}$, 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

$$\boldsymbol{v}_{\parallel} = \boldsymbol{v} - \hat{\boldsymbol{n}}\,(\boldsymbol{v}\cdot\hat{\boldsymbol{n}}) \qquad \text{and} \qquad \boldsymbol{v}_{\perp} = -\frac{\hat{\boldsymbol{n}}\,\partial\theta/\partial t}{|\nabla\theta|} \qquad \Longrightarrow \boldsymbol{v} = \boldsymbol{v}_{\parallel} + \boldsymbol{v}_{\perp} + \frac{\hat{\boldsymbol{n}}\,\dot{\theta}}{|\nabla\theta|} \tag{37.2}$$

where $\hat{\boldsymbol{n}} = \nabla\theta/|\nabla\theta|$ is the normal vector on an isentrope. The velocity $\boldsymbol{v}_{\perp}$, by definition, satisfies

$$\frac{\partial\theta}{\partial t} + \boldsymbol{v}_{\perp}\cdot\nabla\theta = 0. \tag{37.3}$$

Hence, according to the kinematics detailed in Section 14.4.2,

$$\boldsymbol{v}_{\perp}\cdot\hat{\boldsymbol{n}} = \boldsymbol{v}_{\theta}\cdot\hat{\boldsymbol{n}}, \tag{37.4}$$

where $\boldsymbol{v}_{\theta}$ is the velocity of a point fixed on the θ isosurface. We make use of this key identity below.

With the velocity decomposition (37.2), the PV-substance flux vector takes the form

$$\begin{aligned}
\boldsymbol{J}_{\mathrm{Q}} &= \rho\, \boldsymbol{v}\, Q - \dot{\theta}\, \boldsymbol{\omega}_{\mathrm{a}} + \nabla\theta\wedge\boldsymbol{F} && \text{(37.5a)}\\
&= \left[\boldsymbol{v}_{\parallel} + \boldsymbol{v}_{\perp} + \frac{\dot{\theta}\,\nabla\theta}{|\nabla\theta|^2}\right]\rho\, Q - \dot{\theta}\,\boldsymbol{\omega}_{\mathrm{a}} + \nabla\theta\wedge\boldsymbol{F} && \text{(37.5b)}\\
&= (\boldsymbol{v}_{\parallel} + \boldsymbol{v}_{\perp})\,\rho Q - \dot{\theta}\,[\boldsymbol{\omega}_{\mathrm{a}} - (\boldsymbol{\omega}_{\mathrm{a}}\cdot\hat{\boldsymbol{n}})\,\hat{\boldsymbol{n}}] + \nabla\theta\wedge\boldsymbol{F} && \text{(37.5c)}\\
&= (\boldsymbol{v}_{\parallel} + \boldsymbol{v}_{\perp})\,\rho Q - \dot{\theta}\,(\boldsymbol{\omega}_{\mathrm{a}})_{\parallel} + \nabla\theta\wedge\boldsymbol{F} && \text{(37.5d)}\\
&= \boldsymbol{v}_{\perp}\,\rho Q + \left[\rho Q\,\boldsymbol{v}_{\parallel} - \dot{\theta}\,(\boldsymbol{\omega}_{\mathrm{a}})_{\parallel}\right] + \nabla\theta\wedge\boldsymbol{F} && \text{(37.5e)}\\
&\equiv \boldsymbol{J}_{\perp} + \boldsymbol{J}_{\parallel}, && \text{(37.5f)}
\end{aligned}$$

where

$$(\boldsymbol{\omega}_{\mathrm{a}})_{\parallel} = \boldsymbol{\omega}_{\mathrm{a}} - (\boldsymbol{\omega}_{\mathrm{a}}\cdot\hat{\boldsymbol{n}})\,\hat{\boldsymbol{n}} = \boldsymbol{\omega}_{\mathrm{a}} - \left[\frac{\boldsymbol{\omega}_{\mathrm{a}}\cdot\nabla\theta}{|\nabla\theta|^2}\right]\nabla\theta = \boldsymbol{\omega}_{\mathrm{a}} - \frac{\rho\, Q}{|\nabla\theta|}\,\hat{\boldsymbol{n}}. \tag{37.6}$$

The above results motivate us to we write the PV-substance budget equation (37.1) in the alternative form

$$\frac{\partial(\rho\, Q)}{\partial t} + \nabla\cdot(\boldsymbol{v}_{\mathrm{Q}}\,\rho\, Q) = 0, \tag{37.7}$$

where

$$\boldsymbol{v}_{\mathrm{Q}} = \frac{\boldsymbol{J}_{\mathrm{Q}}}{\rho\, Q} = \boldsymbol{v}_{\perp} + \boldsymbol{v}_{\parallel} - \frac{\dot{\theta}\,(\boldsymbol{\omega}_{\mathrm{a}})_{\parallel} + \boldsymbol{F} \wedge \nabla\theta}{\rho\, Q} \tag{37.8}$$

is an effective velocity that advects the PV-substance through the fluid. A direct calculation shows that $\boldsymbol{v}_{\mathrm{Q}}$ satisfies the following key property

$$\boldsymbol{v}_{\mathrm{Q}} \cdot \hat{\boldsymbol{n}} = \boldsymbol{v}_{\perp} \cdot \hat{\boldsymbol{n}} = \boldsymbol{v}_{\theta} \cdot \hat{\boldsymbol{n}}, \tag{37.9}$$

where the final equality made use of the identity (37.4). As a result, the velocity $\boldsymbol{v}_{\mathrm{Q}}$, which advects PV-substance, has a normal component that is identical to that of the velocity of a point fixed on the isentrope

$$\frac{\partial\theta}{\partial t} + \boldsymbol{v}_{\mathrm{Q}} \cdot \nabla\theta = 0. \tag{37.10}$$

We depict this result in Figure 37.1, 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 37.1: *The flux, $\boldsymbol{J}_{\mathrm{Q}}$, of PV-substance does not penetrate a surface of constant potential temperature (isentrope). This kinematic result follows since the effective velocity of PV-substance, $\boldsymbol{v}_{\mathrm{Q}} = \boldsymbol{J}_{\mathrm{Q}}^{\mathrm{HM}}/(\rho\, Q)$, has the same normal component as a point fixed on a potential temperature iso-surface, $\boldsymbol{v}_{\mathrm{Q}} \cdot \hat{\boldsymbol{n}} = \boldsymbol{v}_{\theta} \cdot \hat{\boldsymbol{n}}$. 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.*

37.1.2 A kinematic derivation of impermeability

We here offer a second derivation of the impermability 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

$$\rho\, Q = \boldsymbol{\omega}_{\mathrm{a}} \cdot \nabla\theta = \nabla \cdot (\boldsymbol{\omega}_{\mathrm{a}}\, \theta), \tag{37.11}$$

so that $\rho\, Q$ is a pure divergence. Taking the Eulerian time derivative then leads to

$$\frac{\partial(\rho\, Q)}{\partial t} = -\nabla \cdot \boldsymbol{J}_{\mathrm{Q}}^{\mathrm{kin}}, \tag{37.12}$$

where

$$\boldsymbol{J}_{\mathrm{Q}}^{\mathrm{kin}} = -\frac{\partial(\boldsymbol{\omega}_{\mathrm{a}}\,\theta)}{\partial t} \tag{37.13}$$

is the kinematic form of the PV-substance flux. By construction, this flux vanishes in the steady state

$$\boldsymbol{J}_{\mathrm{Q}}^{\mathrm{kin}} = 0 \qquad \text{in steady state,} \tag{37.14}$$

which contrasts to the Haynes-McIntyre form given by equation (37.1). We return to this point in Section 37.2 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 9.9.1)

$$\boldsymbol{v}_{\mathrm{a}} = \boldsymbol{v} + \boldsymbol{\Omega} \wedge \boldsymbol{x} \tag{37.15}$$

leads to

$$-\frac{\partial \boldsymbol{\omega}_{\mathrm{a}}}{\partial t}\,\theta = -\frac{\partial(\nabla \wedge \boldsymbol{v}_{\mathrm{a}})}{\partial t}\,\theta = -\left[\nabla \wedge \frac{\partial \boldsymbol{v}_{\mathrm{a}}}{\partial t}\right]\theta = -\nabla \wedge \left[\frac{\partial \boldsymbol{v}_{\mathrm{a}}}{\partial t}\,\theta\right] + \nabla\theta \wedge \frac{\partial \boldsymbol{v}_{\mathrm{a}}}{\partial t}. \tag{37.16}$$

Dropping the total curl yields the modified kinematic form for the PV-substance flux

$$\widetilde{\boldsymbol{J}}_{\mathrm{Q}}^{\mathrm{kin}} = -\frac{\partial \boldsymbol{v}_{\mathrm{a}}}{\partial t} \wedge \nabla\theta - \boldsymbol{\omega}_{\mathrm{a}}\,\frac{\partial \theta}{\partial t}. \tag{37.17}$$

It follows that

$$\widetilde{\boldsymbol{v}}_{\mathrm{Q}}^{\mathrm{kin}} \cdot \hat{\boldsymbol{n}} = \frac{\widetilde{\boldsymbol{J}}_{\mathrm{Q}}^{\mathrm{kin}} \cdot \hat{\boldsymbol{n}}}{\rho\, Q} = -\frac{\boldsymbol{\omega}_{\mathrm{a}} \cdot \nabla\theta}{\rho\, Q}\,\frac{\partial \theta}{\partial t}\,\frac{1}{|\nabla\theta|} = -\frac{\partial \theta}{\partial t}\,\frac{1}{|\nabla\theta|} = \boldsymbol{v}_{\theta} \cdot \hat{\boldsymbol{n}}, \tag{37.18}$$

which then leads to the impermeability theorem for $\widetilde{\boldsymbol{J}}_{\mathrm{Q}}^{\mathrm{kin}}$. 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 $\nabla \cdot \boldsymbol{\omega}_{\mathrm{a}} = 0$.

37.1.3 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

$$\rho\, Q_{\lambda} \equiv \boldsymbol{\omega}_{\mathrm{a}} \cdot \nabla\lambda = \nabla \cdot (\boldsymbol{\omega}_{\mathrm{a}}\,\lambda). \tag{37.19}$$

Following the derivation given in Section 37.1.2 leads to

$$\frac{\partial(\rho\, Q_{\lambda})}{\partial t} = \nabla \cdot \left[\frac{\partial \boldsymbol{v}_{\mathrm{a}}}{\partial t} \wedge \nabla\lambda + \boldsymbol{\omega}_{\mathrm{a}}\,\frac{\partial \lambda}{\partial t}\right] \equiv -\nabla \cdot \boldsymbol{J}_{Q_{\lambda}}. \tag{37.20}$$

It follows that

$$\frac{\boldsymbol{J}_{Q_{\lambda}} \cdot \hat{\boldsymbol{n}}_{\lambda}}{\rho\, Q_{\lambda}} = -\frac{\boldsymbol{\omega}_{\mathrm{a}} \cdot \nabla\lambda}{\rho\, Q_{\lambda}}\,\frac{\partial \lambda}{\partial t}\,\frac{1}{|\nabla\lambda|} = -\frac{\partial \lambda}{\partial t}\,\frac{1}{|\nabla\lambda|} = \boldsymbol{v}_{\lambda} \cdot \hat{\boldsymbol{n}}_{\lambda}, \tag{37.21}$$

where

$$\hat{\boldsymbol{n}}_\lambda = \frac{\nabla\lambda}{|\nabla\lambda|} \qquad \text{and} \qquad \frac{\partial\lambda}{\partial t} + \boldsymbol{v}_\lambda \cdot \nabla\lambda = 0. \tag{37.22}$$

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.

37.1.4 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 prefered depends on the application, with *Bretherton and Schär* (1993), *Davies-Jones* (2003), and *Marshall et al.* (2001) proposing criteria favoring one form over another. We pursue an example of such considerations in Section 37.2.

37.2 Impermeability theorem for a realistic seawater EOS

As seen from Section 37.1.3, 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 36.1.4). As shown in Section 36.2, 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.

37.2.1 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

$$Q^{\text{ocn}} = \frac{\nabla b \cdot \boldsymbol{\omega}_{\text{a}}}{\rho}, \tag{37.23}$$

where the globally defined buoyancy field, b, is approximated by a chosen potential density. As shown in Section 36.2, a globally defined buoyancy does not annihilate the baroclinicity vector for a realistic seawater equation of state

$$\boldsymbol{B} \cdot \nabla b = [-\nabla(1/\rho) \wedge \nabla p] \cdot \nabla b \neq 0. \tag{37.24}$$

Consequently, $\mathrm{D}Q^{\mathrm{ocn}}/\mathrm{D}t \neq 0$ even in the absence of irreversible processes. Nonetheless, the Eulerian budget for PV-substance satisfies

$$\frac{\partial(\rho\, Q^{\mathrm{ocn}})}{\partial t} = -\nabla \cdot \widetilde{\boldsymbol{J}}_{\mathrm{Q}}^{\mathrm{ocn}}, \tag{37.25}$$

and $\widetilde{\boldsymbol{J}}_{\mathrm{Q}}^{\mathrm{ocn}}$ satisfies the impermeability theorem for b-surfaces. Both properties greatly facilitate the study of budgets for PV-substance even within an ocean with a realistic equation of state. Proof of this assertion follows from the discussion in Section 37.1.2, where we know that the kinematic flux

$$\widetilde{\boldsymbol{J}}_{\mathrm{Q}}^{\mathrm{ocn}} = -\frac{\partial \boldsymbol{v}_{\mathrm{a}}}{\partial t} \wedge \nabla b - \boldsymbol{\omega}_{\mathrm{a}} \frac{\partial b}{\partial t} = -\frac{\partial \boldsymbol{v}}{\partial t} \wedge \nabla b - \boldsymbol{\omega}_{\mathrm{a}} \frac{\partial b}{\partial t} \tag{37.26}$$

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 (37.26) follows since

$$\frac{\partial \boldsymbol{v}_{\mathrm{a}}}{\partial t} = \frac{\partial(\boldsymbol{v} + \boldsymbol{\Omega} \wedge \boldsymbol{x})}{\partial t} = \frac{\partial \boldsymbol{v}}{\partial t}, \tag{37.27}$$

given that the Eulerian time derivative is computed at a fixed position, $\boldsymbol{x}$, and the planetary rotation is assumed constant.

37.2.2 A modified PV-substance flux

The kinematic flux (37.26) 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 we make use of the vector-invariant velocity equation (equation (34.21))

$$\frac{\partial \boldsymbol{v}}{\partial t} + \boldsymbol{\omega}_{\mathrm{a}} \wedge \boldsymbol{v} = -\rho^{-1} \nabla p - \nabla \mathcal{M} + \boldsymbol{F}, \tag{37.28}$$

where

$$\mathcal{M} = \frac{1}{2}\, \boldsymbol{v}^2 + \Phi \tag{37.29}$$

is the mechanical energy per mass of a fluid element. Bringing the pressure term inside of the gradient operator leads to

$$\frac{\partial \boldsymbol{v}}{\partial t} + \boldsymbol{\omega}_{\mathrm{a}} \wedge \boldsymbol{v} = p\, \nabla(1/\rho) - \nabla(\mathcal{M} + p/\rho) + \boldsymbol{F}. \tag{37.30}$$

In the following, we write

$$B = \mathcal{M} + p/\rho, \tag{37.31}$$

which is a Bernoulli function.[1] The vector-invariant velocity equation (37.30) leads to the cross-product

$$\frac{\partial \boldsymbol{v}}{\partial t} \wedge \nabla b = -(\boldsymbol{\omega}_{\mathrm{a}} \wedge \boldsymbol{v}) \wedge \nabla b + [p\, \nabla(1/\rho) - \nabla(\mathcal{M} + p/\rho) + \boldsymbol{F}] \wedge \nabla b \tag{37.32a}$$

$$= -(\nabla b \cdot \boldsymbol{\omega}_{\mathrm{a}})\, \boldsymbol{v} + (\nabla b \cdot \boldsymbol{v})\, \boldsymbol{\omega}_{\mathrm{a}} + [p\, \nabla(1/\rho) - \nabla(\mathcal{M} + p/\rho) + \boldsymbol{F}] \wedge \nabla b \tag{37.32b}$$

$$= -\boldsymbol{v}\, \rho\, Q^{\mathrm{ocn}} + (\dot{b} - \partial_t b)\, \boldsymbol{\omega}_{\mathrm{a}} + [p\, \nabla(1/\rho) - \nabla(\mathcal{M} + p/\rho) + \boldsymbol{F}] \wedge \nabla b. \tag{37.32c}$$

[1] The Bernoulli potential, $\mathcal{B}$, arises from an analysis of the total energy budget as in equation (23.57), where we see that the Bernoulli potential in a compressible (non-Boussinesq) fluid, $\mathcal{B} = \mathcal{M} + p/\rho + \mathcal{I}$, also includes the internal energy per mass, $\mathcal{I}$. However, the internal energy is missing from equation (37.31), thus motivating our use of the terminology "a Bernoulli function" rather than "the Bernoulli potential".

Use of this result leads to the flux (37.26)

$$\widetilde{\boldsymbol{J}}_{\mathrm{Q}}^{\mathrm{ocn}} = -\frac{\partial \boldsymbol{v}}{\partial t} \wedge \nabla b - \boldsymbol{\omega}_{\mathrm{a}} \frac{\partial b}{\partial t} \tag{37.33a}$$

$$= \boldsymbol{v}\, \rho\, Q^{\mathrm{ocn}} - \dot{b}\, \boldsymbol{\omega}_{\mathrm{a}} - \boldsymbol{F} \wedge \nabla b + [\nabla(\mathcal{M} + p/\rho) - p\, \nabla(1/\rho)] \wedge \nabla b \tag{37.33b}$$

$$= \boldsymbol{J}_{\mathrm{Q}} + [\nabla(\mathcal{M} + p/\rho) - p\, \nabla(1/\rho)] \wedge \nabla b, \tag{37.33c}$$

where $\boldsymbol{J}_{\mathrm{Q}}$ is the Haynes-McIntyre form of the PV-substance flux given by equation (37.1). The term

$$\nabla(\mathcal{M} + p/\rho) \wedge \nabla b = \nabla \wedge [(\mathcal{M} + p/\rho)\, \nabla b] \tag{37.34}$$

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

$$\boldsymbol{J}_{\mathrm{Q}}^{\mathrm{marshall}} = \widetilde{\boldsymbol{J}}_{\mathrm{Q}}^{\mathrm{ocn}} - \nabla(\mathcal{M} + p/\rho) \wedge \nabla b = -\left[\frac{\partial \boldsymbol{v}}{\partial t} + \nabla(\mathcal{M} + p/\rho)\right] \wedge \nabla b - \boldsymbol{\omega}_{\mathrm{a}} \frac{\partial b}{\partial t}. \tag{37.35}$$

Diagnostically desirable features of $\boldsymbol{J}_{\mathrm{Q}}^{\mathrm{marshall}}$ include the following

- $\nabla b \cdot \boldsymbol{J}_{\mathrm{Q}}^{\mathrm{marshall}}/(\rho\, Q) = \partial_t b$, thus satisfying the impermeability theorem.
- $\boldsymbol{J}_{\mathrm{Q}}^{\mathrm{marshall}}$ has no explicit reference to irreversible processes. Consequently, it can in some cases be simpler to diagnose than the Haynes-McIntyre flux $\boldsymbol{J}_{\mathrm{Q}}$.
- In a steady state, the flux is given by

$$\boldsymbol{J}_{\mathrm{Q}}^{\mathrm{marshall}} = \nabla b \wedge \nabla(\mathcal{M} + p/\rho) = \nabla \wedge [b\, \nabla(\mathcal{M} + p/\rho)]. \qquad \text{steady state.} \tag{37.36}$$

  Consequently, $\mathcal{M} + p/\rho$ 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.

37.2.3 Integral constraints for steady state

The steady state PV-substance flux in the form (37.36) can be used to develop some integral constraints on the steady flow. For this purpose consider the steady form of $\boldsymbol{J}_{\mathrm{Q}}^{\mathrm{marshall}}$ and integrate over an arbitrary simply connected area making use of Stokes' Theorem

$$\int_{\mathcal{S}} \nabla \wedge [b\, \nabla B] \cdot \hat{\boldsymbol{n}}\, \mathrm{d}\mathcal{S} = \oint_{\partial \mathcal{S}} b\, \nabla B \cdot \mathrm{d}\boldsymbol{r} = \oint_{\partial \mathcal{S}} b\, \mathrm{d}B = -\oint_{\partial \mathcal{S}} B\, \mathrm{d}b. \tag{37.37a}$$

The first equality made use of Stokes' Theorem; the second make use of the identity for exact differentials

$$\nabla B \cdot \mathrm{d}\boldsymbol{r} = \mathrm{d}B; \tag{37.38}$$

and the final equality made use of

$$b\, \mathrm{d}B = \mathrm{d}(b\, B) - B\, \mathrm{d}b \tag{37.39}$$

and noted that

$$\oint_{\partial \mathcal{S}} \mathrm{d}(B\, b) = 0. \tag{37.40}$$

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

$$\int_{\mathcal{S}} \boldsymbol{J}_{\mathrm{Q}}^{\mathrm{marshall}} \cdot \hat{\boldsymbol{n}} \, \mathrm{d}\mathcal{S} = 0 \qquad \text{area enclosed by contour with } \mathcal{M} + p/\rho \text{ constant or } b \text{ constant.} \tag{37.41}$$

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 $\hat{\boldsymbol{n}} = \hat{\boldsymbol{z}}$) in a Boussinesq fluid.

37.2.4 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 36.5.4.

37.3 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 37.2

$$Q = \frac{\boldsymbol{\omega}_{\mathrm{a}} \cdot \nabla b}{\rho} = \frac{\nabla \cdot (\boldsymbol{\omega}_{\mathrm{a}} \, b)}{\rho}. \tag{37.42}$$

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 propery measuring the amount of PV-substance per unit mass (i.e., the concentration of PV-substance), and correspondingly with $\rho \, Q$ 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 $\rho \, Q$

$$\mathcal{I} = \int_{\mathcal{R}} Q \, \rho \, \mathrm{d}V = \int_{\mathcal{R}} \nabla \cdot (\boldsymbol{\omega}_{\mathrm{a}} \, b) \, \mathrm{d}V = \oint_{\partial \mathcal{R}} \boldsymbol{\omega}_{\mathrm{a}} \, b \cdot \hat{\boldsymbol{n}} \, \mathrm{d}\mathcal{S}, \tag{37.43}$$

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.

37.3.1 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 37.2. 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 (37.43) so that

$$\mathcal{I} = \oint_{\partial \mathcal{R}} \boldsymbol{\omega}_{\mathrm{a}} \, b \cdot \hat{\boldsymbol{n}} \, \mathrm{d}\mathcal{S} = b_2 \oint_{\partial \mathcal{R}} \boldsymbol{\omega}_{\mathrm{a}} \cdot \hat{\boldsymbol{n}} \, \mathrm{d}\mathcal{S}. \tag{37.44}$$

We can now use the divergence theorem to return to the volume integral, only now with b outside of the integral

$$\mathcal{I} = b_2 \int_{\mathcal{R}} \nabla \cdot \boldsymbol{\omega}_{\mathrm{a}} \, \mathrm{d}V = 0, \tag{37.45}$$

[2] Recall our discussion of extensive and intensive fluid properties in Section 15.3.1.

Figure 37.2: *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 $\mathscr{R}$. 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, $\mathscr{R}_{12}$, 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 $\mathscr{R}$ and $\mathscr{R}_{12}$. 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 $\nabla \cdot \boldsymbol{\omega}_{\mathrm{a}} = 0$ led to the final equality. Equivalently, we can use Stokes' Theorem to convert the closed area integral, $\oint_{\partial\mathscr{R}} \boldsymbol{\omega}_{\mathrm{a}} \cdot \hat{\boldsymbol{n}}\, \mathrm{d}\mathcal{S}$, 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 $\mathscr{I} = 0$ (see Section 2.7.4).

The remarkable identity (37.45) 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.

37.3.2 Region bounded by two buoyancy surfaces

The identity (37.45) has a corollary, in which we consider a region bounded by two b-surfaces such as $\mathscr{R}_{12}$ shown in the right panel of Figure 37.2. 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

$$\mathscr{I} = \int_{\mathscr{R}_{12}} \nabla \cdot (\boldsymbol{\omega}_{\mathrm{a}}\, b)\, \mathrm{d}V = \int_{\mathscr{R}_2} \nabla \cdot (\boldsymbol{\omega}_{\mathrm{a}}\, b)\, \mathrm{d}V - \int_{\mathscr{R}_1} \nabla \cdot (\boldsymbol{\omega}_{\mathrm{a}}\, b)\, \mathrm{d}V, \tag{37.46}$$

where the domain $\mathscr{R}_1$ extends from the ground up to b_1 and $\mathscr{R}_2$ 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 $\mathscr{R}$ in Figure 37.2. Invoking the single isentrope result we see that

Figure 37.3: *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, $b_B(\boldsymbol{x}, t)$.*

both integrals separately vanish. We are thus led to a vanishing integral for the layer

$$\mathcal{I} = \int_{\mathcal{R}_{12}} \nabla \cdot (\boldsymbol{\omega}_{\mathrm{a}}\, b)\, \mathrm{d}V = 0. \tag{37.47}$$

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.

37.3.3 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 37.2. The integrated PV is given by

$$\begin{aligned}
\mathcal{I} &= \int_{\mathcal{R}_{AB}} \nabla \cdot (\boldsymbol{\omega}_{\mathrm{a}}\, b)\, \mathrm{d}V && (37.48\mathrm{a})\\
&= \int_{\mathcal{S}_A} \boldsymbol{\omega}_{\mathrm{a}}\, b \cdot \hat{\boldsymbol{n}}\, \mathrm{d}\mathcal{S} + \int_{\mathcal{S}_B} \boldsymbol{\omega}_{\mathrm{a}}\, b \cdot \hat{\boldsymbol{n}}\, \mathrm{d}\mathcal{S} && (37.48\mathrm{b})\\
&= \theta_A \int_{\mathcal{S}_A} \boldsymbol{\omega}_{\mathrm{a}} \cdot \hat{\boldsymbol{n}}\, \mathrm{d}\mathcal{S} + \int_{\mathcal{S}_B} b_B\, \boldsymbol{\omega}_{\mathrm{a}} \cdot \hat{\boldsymbol{n}}\, \mathrm{d}\mathcal{S} && (37.48\mathrm{c})\\
&= \theta_A \left[\int_{\mathcal{S}_A} \boldsymbol{\omega}_{\mathrm{a}} \cdot \hat{\boldsymbol{n}}\, \mathrm{d}\mathcal{S} + \int_{\mathcal{S}_B} \boldsymbol{\omega}_{\mathrm{a}} \cdot \hat{\boldsymbol{n}}\, \mathrm{d}\mathcal{S} \right] + \int_{\mathcal{S}_B} (b_B - b_A)\, \boldsymbol{\omega}_{\mathrm{a}} \cdot \hat{\boldsymbol{n}}\, \mathrm{d}\mathcal{S} && (37.48\mathrm{d})\\
&= b_A \int_{\mathcal{R}_{AB}} \nabla \cdot \boldsymbol{\omega}_{\mathrm{a}}\, \mathrm{d}V + \int_{\mathcal{S}_B} (b_B - b_A)\, \boldsymbol{\omega}_{\mathrm{a}} \cdot \hat{\boldsymbol{n}}\, \mathrm{d}\mathcal{S} && (37.48\mathrm{e})\\
&= \int_{\mathcal{S}_B} (b_B - b_A)\, \boldsymbol{\omega}_{\mathrm{a}} \cdot \hat{\boldsymbol{n}}\, \mathrm{d}\mathcal{S}, && (37.48\mathrm{f})
\end{aligned}$$

where we made use $\nabla \cdot \boldsymbol{\omega}_{\mathrm{a}} = 0$ 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 37.4: *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 $\mathcal{S}_B$.*

37.3.4 An ocean layer that outcrops

Figure 37.4 depicts a buoyancy layer in the ocean that outcrops at the ocean surface at both of its ends. Following the derivation in Section 37.3.3 we have the following integrated PV contents

$$\mathcal{I}_A \equiv \int_{\mathcal{S}_A} b\,\boldsymbol{\omega}_{\mathrm{a}} \cdot \hat{\boldsymbol{n}}\,\mathrm{d}\mathcal{S} + \int_{\mathcal{S}_1+\mathcal{S}_2+\mathcal{S}_3} b\,\boldsymbol{\omega}_{\mathrm{a}} \cdot \hat{\boldsymbol{n}}\,\mathrm{d}\mathcal{S} = \int_{\mathcal{S}_1+\mathcal{S}_2+\mathcal{S}_3} (b - b_A)\,\boldsymbol{\omega}_{\mathrm{a}} \cdot \hat{\boldsymbol{n}}\,\mathrm{d}\mathcal{S} \tag{37.49a}$$

$$\mathcal{I}_B \equiv \int_{\mathcal{S}_B} b\,\boldsymbol{\omega}_{\mathrm{a}} \cdot \hat{\boldsymbol{n}}\,\mathrm{d}\mathcal{S} + \int_{\mathcal{S}_2} b\,\boldsymbol{\omega}_{\mathrm{a}} \cdot \hat{\boldsymbol{n}}\,\mathrm{d}\mathcal{S} = \int_{\mathcal{S}_2} (b - b_B)\,\boldsymbol{\omega}_{\mathrm{a}} \cdot \hat{\boldsymbol{n}}\,\mathrm{d}\mathcal{S}. \tag{37.49b}$$

Taking the difference in the integrals leads to the integrated PV within the layer $\mathcal{R}_{AB}$

$$\mathcal{I}_{AB} \equiv \int_{\mathcal{R}_{AB}} \nabla \cdot (\boldsymbol{\omega}_{\mathrm{a}}\, b)\,\mathrm{d}V \tag{37.50a}$$

$$= \mathcal{I}_A - \mathcal{I}_B \tag{37.50b}$$

$$= \int_{\mathcal{S}_1} (b - b_A)\,\boldsymbol{\omega}_{\mathrm{a}} \cdot \hat{\boldsymbol{n}}\,\mathrm{d}\mathcal{S} + (b_B - b_A) \int_{\mathcal{S}_2} \boldsymbol{\omega}_{\mathrm{a}} \cdot \hat{\boldsymbol{n}}\,\mathrm{d}\mathcal{S} + \int_{\mathcal{S}_3} (b - b_A)\,\boldsymbol{\omega}_{\mathrm{a}} \cdot \hat{\boldsymbol{n}}\,\mathrm{d}\mathcal{S}. \tag{37.50c}$$

37.4 Layer integrated PV budget

In Section 37.3 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 37.1 by considering a buoyancy (isopycnal) layer within the ocean that intersects the bottom on one side and the atmosphere on the other (Figure 37.5). 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, entrainment, detrainment). 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 37.3. As per the discussion of Section 37.2, 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 37.5: *An isopycnal layer of seawater denoted by $\mathscr{R}$, with the layer intersecting bottom topography on one side and the atmosphere on the other.*

37.4.1 Layer integrated budget

In addition to waves, currents, entrainment, and detrainment 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 formulational framework, we derive the isopycnal layer PV budget making use of the Leibniz-Reynolds transport theorem derived in Section 15.3.4. Just as for the layer integrated tracer budget considered in Section 15.4, applying Leibniz-Reynolds to the layer integrated PV budget renders

$$\frac{\mathrm{d}}{\mathrm{d}t}\left[\int_{\mathscr{R}} \rho\, Q\, \mathrm{d}V\right] = \int_{\mathscr{R}} \left[\frac{\partial(\rho\, Q)}{\partial t} + \nabla\cdot(\rho\, Q\, \dot{\boldsymbol{x}})\right] \mathrm{d}V, \tag{37.51}$$

where $\mathscr{R}$ is the domain defined by the layer and

$$\dot{\boldsymbol{x}} = \frac{\mathrm{d}\boldsymbol{x}}{\mathrm{d}t} \tag{37.52}$$

is the velocity for a point on the layer boundary. Making use of the PV equation, $\partial(\rho\, Q)/\partial t = -\nabla\cdot\boldsymbol{J}_{\mathrm{Q}}$, and the divergence theorem renders

$$\frac{\mathrm{d}}{\mathrm{d}t}\left[\int_{\mathscr{R}} \rho\, Q\, \mathrm{d}V\right] = \int_{\mathscr{R}} \nabla\cdot(-\boldsymbol{J}_{\mathrm{Q}} + \rho\, Q\, \dot{\boldsymbol{x}})\, \mathrm{d}V = \int_{\partial\mathscr{R}} (-\boldsymbol{J}_{\mathrm{Q}} + \rho\, Q\, \dot{\boldsymbol{x}})\cdot\hat{\boldsymbol{n}}\, \mathrm{d}\mathcal{S}. \tag{37.53}$$

This result holds around the full domain boundary. Now we decompose that boundary into portions defined by isopyncal layer interfaces and those along the air-sea and land-sea boundaries.

37.4.2 Impermeability across interior layer interfaces

Rather than invoking the impermeability theorem derived in Section 37.1, we rederive it within the present context to further our confidence in its validity. We thus consider the following for interior layer interfaces

$$[-\boldsymbol{J}_{\mathrm{Q}} + \rho\, Q\, \dot{\boldsymbol{x}}]\cdot\hat{\boldsymbol{n}} = \left[\rho\, Q\, (\dot{\boldsymbol{x}} - \boldsymbol{v}) + \dot{b}\, \boldsymbol{\omega}_{\mathrm{a}} - \nabla b \wedge \boldsymbol{F}\right]\cdot\hat{\boldsymbol{n}} \tag{37.54a}$$

$$= \left[(\boldsymbol{\omega}_{\mathrm{a}}\cdot\nabla b)\, (\dot{\boldsymbol{x}} - \boldsymbol{v}) + \dot{b}\, \boldsymbol{\omega}_{\mathrm{a}}\right]\cdot\hat{\boldsymbol{n}} \tag{37.54b}$$

$$= [(\boldsymbol{\omega}_{\mathrm{a}}\cdot\nabla b)\, (\dot{\boldsymbol{x}} - \boldsymbol{v}) + (\partial b/\partial t + \boldsymbol{v}\cdot\nabla b)\, \boldsymbol{\omega}_{\mathrm{a}}]\cdot\hat{\boldsymbol{n}} \tag{37.54c}$$

$$= [(\boldsymbol{\omega}_{\mathrm{a}}\cdot\nabla b)\, \dot{\boldsymbol{x}} + (\partial b/\partial t)\, \boldsymbol{\omega}_{\mathrm{a}}]\cdot\hat{\boldsymbol{n}}, \tag{37.54d}$$

where

$$\hat{\boldsymbol{n}} = \frac{\nabla b}{|\nabla b|} \tag{37.55}$$

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 (37.9) (here applied to isopycnals)

$$\dot{\boldsymbol{x}} \cdot \hat{\boldsymbol{n}} = -\frac{\partial b/\partial t}{|\nabla b|}. \tag{37.56}$$

This result then leads to the impermeability statement for isopycnal interfaces in the fluid interior

$$(-\boldsymbol{J}_{\mathrm{Q}} + \rho\, Q\, \dot{\boldsymbol{x}}) \cdot \hat{\boldsymbol{n}} = 0. \tag{37.57}$$

We thus conclude that changes to the layer integrated PV occur only via transfer across the land-sea boundary and the air-sea boundary

$$\frac{\mathrm{d}}{\mathrm{d}t}\left[\int_{\mathcal{R}} \rho\, Q\, \mathrm{d}V\right] = \int_{\text{land-sea}} [-\boldsymbol{J}_{\mathrm{Q}} + \rho\, Q\, \dot{\boldsymbol{x}}] \cdot \hat{\boldsymbol{n}}\, \mathrm{d}\mathcal{S} + \int_{\text{air-sea}} [-\boldsymbol{J}_{\mathrm{Q}} + \rho\, Q\, \dot{\boldsymbol{x}}] \cdot \hat{\boldsymbol{n}}\, \mathrm{d}\mathcal{S}. \tag{37.58}$$

37.4.3 PV flux at the land-sea boundary

We here evaluate the PV flux from equation (37.5a) at the land-sea boundary

$$-\boldsymbol{J}_{\mathrm{Q}} + \rho\, Q\, \dot{\boldsymbol{x}} = \rho\, Q\, (\dot{\boldsymbol{x}} - \boldsymbol{v}) + \dot{b}\, \boldsymbol{\omega}_{\mathrm{a}} + \nabla b \wedge \boldsymbol{F}. \tag{37.59}$$

At a solid boundary, the no-normal flow boundary condition means that $\hat{\boldsymbol{n}} \cdot \boldsymbol{v} = 0$. Likewise, the velocity of a point along the boundary moves along the tangent to the boundary so that $\dot{\boldsymbol{x}} \cdot \hat{\boldsymbol{n}} = 0$. The bottom boundary condition is thus affected just by irreversible processes

$$(-\boldsymbol{J}_{\mathrm{Q}} + \rho\, Q\, \dot{\boldsymbol{x}}) \cdot \hat{\boldsymbol{n}} = (\dot{b}\, \boldsymbol{\omega}_{\mathrm{a}} + \nabla b \wedge \boldsymbol{F}) \cdot \hat{\boldsymbol{n}}. \tag{37.60}$$

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

$$\text{no geothermal heating} \Longrightarrow (-\boldsymbol{J}_{\mathrm{Q}} + \rho\, Q\, \dot{\boldsymbol{x}}) \cdot \hat{\boldsymbol{n}} = (\nabla b \wedge \boldsymbol{F}) \cdot \hat{\boldsymbol{n}} = (\hat{\boldsymbol{n}} \wedge \nabla b) \cdot \boldsymbol{F}. \tag{37.61}$$

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

$$\text{no geothermal heating} \Longrightarrow \hat{\boldsymbol{n}} \cdot \nabla b = 0. \tag{37.62}$$

Buoyancy isolines thus intersect the bottom parallel to the bottom outward normal, as shown in Figure 37.6. Correspondingly, $(\hat{\boldsymbol{n}} \wedge \nabla b) \cdot \boldsymbol{F}$ projects out that component of the friction vector pointing parallel to the bottom with constant z. Assuming buoyancy increases upward along the sloping bottom, as per a stably stratified bottom fluid, then $\hat{\boldsymbol{n}} \wedge \nabla b$ points clockwise around bowls and counter-clockwise around bumps (see Figure 37.6).

Within the bottom boundary layer, quadratic bottom drag is a common parameterization of the acceleration associated with turbulent frictional processes

$$\boldsymbol{F} = -C_{\mathrm{d}}\, |\boldsymbol{u}|\, \boldsymbol{u}, \tag{37.63}$$

where C_{d} is a non-dimensional drag coefficient and $\boldsymbol{u}$ is the horizontal velocity. In this case the boundary condition for PV takes the form

$$(\nabla b \wedge \boldsymbol{F}) \cdot \hat{\boldsymbol{n}} = -C_{\mathrm{d}}\, |\boldsymbol{u}|\, (\hat{\boldsymbol{n}} \wedge \nabla b) \cdot \boldsymbol{u}. \tag{37.64}$$

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 $\nabla b \wedge \boldsymbol{F} = 0$. This result is expected from the discussion in Section 36.4 and Figure 36.3, where we note that friction changes PV by rotating buoyancy surfaces, with that rotation realized only when friction is not aligned with ∇b.

Figure 37.6: *Buoyancy isosurfaces that intersect the bottom. As discussed in Section 15.4 and depicted in Figure 15.4, in the absence of geothermal heating, a buoyancy isosurface satisfies the no-normal flux bottom boundary condition, $\hat{\boldsymbol{n}} \cdot \nabla b = 0$. This boundary condition requires buoyancy isosurfaces to be orthogonal to the bottom. Assuming buoyancy increases upward along the sloping bottom, then $\hat{\boldsymbol{n}} \wedge \nabla b$ 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 37.4.3.*

37.4.4 PV flux at the air-sea boundary

We make use of the kinematic boundary condition derived in Section 14.4.3 for the permeable air-sea boundary, where the boundary condition (14.60) leads to

$$\rho\, \hat{\boldsymbol{n}} \cdot (\dot{\boldsymbol{x}} - \boldsymbol{v}) = \mathcal{Q}_{\mathrm{m}} \qquad \text{air-sea boundary} \tag{37.65}$$

with $\mathcal{Q}_{\mathrm{m}}$ the mass per time per surface area crossing the boundary. We are thus led to the air-sea boundary condition

$$(-\boldsymbol{J}_{\mathrm{Q}} + \rho\, Q\, \dot{\boldsymbol{x}}) \cdot \hat{\boldsymbol{n}} = Q\, \mathcal{Q}_{\mathrm{m}} + (\dot{b}\, \boldsymbol{\omega}_{\mathrm{a}} + \nabla b \wedge \boldsymbol{F}) \cdot \hat{\boldsymbol{n}}. \tag{37.66}$$

Besides the irreversible processes, PV is affected at the air-sea interface by the transfer of matter across the boundary via the term $Q\, \mathcal{Q}_{\mathrm{m}}$. 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

$$\hat{\boldsymbol{n}} \approx \hat{\boldsymbol{z}}, \tag{37.67}$$

in which case

$$(-\boldsymbol{J}_{\mathrm{Q}} + \rho\, Q\, \dot{\boldsymbol{x}}) \cdot \hat{\boldsymbol{n}} \approx Q\, \mathcal{Q}_{\mathrm{m}} + \dot{b}\, (\zeta + f) + (\nabla b \wedge \boldsymbol{F}) \cdot \hat{\boldsymbol{z}}. \tag{37.68}$$

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 34.4.

37.4.5 Thought experiments

The surface PV flux (37.68) 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 (37.68) 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.

37.4.6 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 (37.1))

$$\boldsymbol{J}_{\mathrm{Q}} = \rho\,\boldsymbol{v}\,Q - \dot{b}\,\boldsymbol{\omega}_{\mathrm{a}} + \nabla b \wedge \boldsymbol{F}. \tag{37.69}$$

We could have instead chosen to work with the Marshall form (equation (37.35))

$$\boldsymbol{J}_{\mathrm{Q}}^{\mathrm{marshall}} = -\left[\frac{\partial \boldsymbol{v}}{\partial t} + \nabla(\mathcal{M} + p/\rho)\right] \wedge \nabla b - \boldsymbol{\omega}_{\mathrm{a}}\,\frac{\partial b}{\partial t}, \tag{37.70}$$

or the modified kinematic form (equation (37.17))

$$\widetilde{\boldsymbol{J}}_{\mathrm{Q}}^{\mathrm{kin}} = -\frac{\partial \boldsymbol{v}_{\mathrm{a}}}{\partial t} \wedge \nabla\theta - \boldsymbol{\omega}_{\mathrm{a}}\,\frac{\partial \theta}{\partial t}. \tag{37.71}$$

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

$$\boldsymbol{J}_{\mathrm{Q}} = \rho\, \boldsymbol{v}\, Q - (\boldsymbol{v} \cdot \nabla b)\, \boldsymbol{\omega}_{\mathrm{a}} + \nabla b \wedge \boldsymbol{F} \tag{37.72}$$

$$\boldsymbol{J}_{\mathrm{Q}}^{\mathrm{marshall}} = -\nabla(\mathcal{M} + p/\rho) \wedge \nabla b \tag{37.73}$$

$$\widetilde{\boldsymbol{J}}_{\mathrm{Q}}^{\mathrm{kin}} = 0. \tag{37.74}$$

The physical picture for $\widetilde{\boldsymbol{J}}_{\mathrm{Q}}^{\mathrm{kin}}$ 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 $\boldsymbol{J}_{\mathrm{Q}}$ or $\boldsymbol{J}_{\mathrm{Q}}^{\mathrm{marshall}}$ 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 $\boldsymbol{J}_{\mathrm{Q}}^{\mathrm{marshall}}$ for steady budgets, and we explore facets of that approach in Section 36.5 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)).

38

General vertical coordinate dynamics

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 8 and the corresponding kinematics in Chapter 17. We make particular use of the layer integrated notions introduced for mass continuity and the tracer equations in Sections 17.9 and 17.10. 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.

38.1 Equations of motion . . . **682**
38.1.1 Notation . . . 682
38.1.2 Mass and tracer equations . . . 682
38.1.3 Momentum equation . . . 682
38.1.4 Eulerian flux-form horizontal momentum equation . . . 683
38.1.5 Vector-invariant horizontal momentum equation . . . 683
38.1.6 Hydrostatic balance with constant gravitational acceleration . . . 684
38.2 Concerning the pressure force . . . **684**
38.2.1 Computing the horizontal pressure gradient . . . 685
38.2.2 Integrated pressure force on the cell faces . . . 686
38.2.3 Net vertical pressure force . . . 687
38.2.4 Net horizontal pressure force . . . 687
38.2.5 Comments . . . 687
38.3 Hydrostatic vorticity and potential vorticity . . . **687**
38.3.1 Basic manipulations . . . 688
38.3.2 Vorticity and potential vorticity equation . . . 689
38.3.3 Boussinesq fluid . . . 690
38.4 Exercise . . . **690**

38.1 Equations of motion

We here derive the equations of motion based on generalized vertical coordinates. The scalar equations were already discussed in Sections 17.8, 17.9, and 17.10, so our main focus concerns the momentum equation.

38.1.1 Notation

For much of this chapter we focus on the continuous formulation of the generalized vertical coordinates. Following the discussion in Section 17.2, we encounter the specific thickness throughout the equations

$$\hbar = \frac{\partial z}{\partial \sigma} = z_\sigma. \tag{38.1}$$

Using this notation, we note that the dia-surface transport velocity is written

$$w^{(\dot\sigma)} = \frac{\partial z}{\partial \sigma}\,\dot\sigma = \hbar\,\dot\sigma. \tag{38.2}$$

Likewise, the dia-surface advection operator is given by

$$w^{(\dot\sigma)}\,\partial_z = \dot\sigma\,\partial_\sigma. \tag{38.3}$$

38.1.2 Mass and tracer equations

The mass and tracer equations were derived in Sections 17.8, 17.9, and 17.10, with their continous vertical coordinate formulation given by

$$\frac{\partial(\rho\,\hbar)}{\partial t} + \nabla_\sigma \cdot (\rho\,\hbar\,\boldsymbol{u}) + \partial_\sigma\,(\rho\,\hbar\,\dot\sigma) = 0 \tag{38.4a}$$

$$\frac{\partial(\hbar\,\rho\,C)}{\partial t} + \nabla_\sigma \cdot (\hbar\,\rho\,C\,\boldsymbol{u} + \hbar\,\boldsymbol{J}^h) + \partial_\sigma(\rho\,\hbar\,\dot\sigma\,C + \hbar\,\nabla\sigma \cdot \boldsymbol{J}) = 0. \tag{38.4b}$$

Compatibility is maintained between the mass continuity equation (38.4a) and the tracer equation (38.4b) 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, $\boldsymbol{J}$, 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 15.1.

38.1.3 Momentum equation

From Section 23.6, the horizontal and vertical components to the momentum equation are

$$\rho\,\frac{\mathrm{D}\boldsymbol{u}}{\mathrm{D}t} + 2\,\rho\,\boldsymbol{\Omega} \wedge \boldsymbol{u} = -\rho\,\nabla_z\Phi - \nabla_z p + \rho\boldsymbol{F}^h \tag{38.5a}$$

$$\rho\,\frac{\mathrm{D}w}{\mathrm{D}t} = -\rho\,\frac{\partial\Phi}{\partial z} - \frac{\partial p}{\partial z} + \rho F^z. \tag{38.5b}$$

Note that for the traditional form of the geopotential (Section 9.12), $\Phi = g\,z$ so that the horizontal gradient of the geopotential vanishes

$$\Phi = g\,z \Longrightarrow \nabla_z\Phi = 0. \tag{38.6}$$

This gradient is nonzero in the presence of astronomical tide forcing (Chapter 43). It is also nonzero when computed along surfaces of constant generalized vertical coordinate, $\nabla_\sigma\Phi \neq 0$, as encountered in the following.

Horizontal momentum equation

We transform the horizontal derivatives from geopotential coordinates to generalized vertical coordinates according to (see equation (8.61))

$$\nabla_z = \nabla_\sigma - (\nabla_\sigma z)\,\partial_z \tag{38.7}$$

thus leading to the horizontal momentum equation

$$\rho\,\frac{\mathrm{D}\boldsymbol{u}}{\mathrm{D}t} + 2\,\rho\,\boldsymbol{\Omega}\wedge\boldsymbol{u} = -\rho\,[\nabla_\sigma - (\nabla_\sigma z)\,\partial_z]\,\Phi - [\nabla_\sigma - (\nabla_\sigma z)\,\partial_z]\,p + \rho\boldsymbol{F}^h. \tag{38.8}$$

In Section 38.1.6 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

$$\rho\,\frac{\mathrm{D}w}{\mathrm{D}t} = -\frac{\partial\sigma}{\partial z}\left[\rho\,\frac{\partial\Phi}{\partial\sigma} + \frac{\partial p}{\partial\sigma}\right] + \rho\,F^z, \tag{38.9}$$

with the hydrostatic form given by

$$\frac{\partial p}{\partial\sigma} = -\rho\,\frac{\partial\Phi}{\partial\sigma}. \tag{38.10}$$

38.1.4 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 17.10)

$$\hbar\,\rho\,\frac{\mathrm{D}u}{\mathrm{D}t} = \left[\frac{\partial(\hbar\,\rho\,u)}{\partial t}\right]_\sigma + \nabla_\sigma\cdot(\hbar\,\rho\,u\,\boldsymbol{u}) + \partial_\sigma(\hbar\,\rho\,u\,\dot\sigma) \tag{38.11a}$$

$$\hbar\,\rho\,\frac{\mathrm{D}v}{\mathrm{D}t} = \left[\frac{\partial(\hbar\,\rho\,v)}{\partial t}\right]_\sigma + \nabla_\sigma\cdot(\hbar\,\rho\,v\,\boldsymbol{u}) + \partial_\sigma(\hbar\,\rho\,v\,\dot\sigma). \tag{38.11b}$$

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, $w = \mathrm{D}z/\mathrm{D}t$. 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 38.1.5.

38.1.5 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 38.3). Here, we start with the material time derivative in the form (17.49c) appropriate for generalized vertical coordinates, in which case the horizontal acceleration is given by

$$\frac{\mathrm{D}\boldsymbol{u}}{\mathrm{D}t} = \left[\frac{\partial\boldsymbol{u}}{\partial t}\right]_\sigma + (\boldsymbol{u}\cdot\nabla_\sigma)\,\boldsymbol{u} + (\dot\sigma\,\partial_\sigma)\,\boldsymbol{u}. \tag{38.12}$$

Now make use of the vector identity (see Section 2.3.4)

$$(\boldsymbol{u}\cdot\nabla_\sigma)\,\boldsymbol{u} = \nabla_\sigma K + (\nabla_\sigma\wedge\boldsymbol{u})\wedge\boldsymbol{u}, \tag{38.13}$$

where

$$K = \boldsymbol{u}\cdot\boldsymbol{u}/2 \tag{38.14}$$

is the kinetic energy per mass of the horizontal flow. Introducing the GVC version of the relative vorticity (see Section 46.2.1)

$$\widetilde{\zeta} \equiv \hat{\boldsymbol{z}}\cdot(\nabla_\sigma\wedge\boldsymbol{u}) = \left[\frac{\partial v}{\partial x}\right]_\sigma - \left[\frac{\partial u}{\partial y}\right]_\sigma \tag{38.15}$$

renders

$$\frac{\mathrm{D}\boldsymbol{u}}{\mathrm{D}t} = \left[\frac{\partial\boldsymbol{u}}{\partial t}\right]_\sigma + \nabla_\sigma K + \widetilde{\zeta}\,\hat{\boldsymbol{z}}\wedge\boldsymbol{u} + \dot{\sigma}\,\partial_\sigma\,\boldsymbol{u}, \tag{38.16}$$

so that the horizontal momentum equation takes the form

$$\left[\frac{\partial\boldsymbol{u}}{\partial t}\right]_\sigma + \dot{\sigma}\,\frac{\partial\boldsymbol{u}}{\partial\sigma} + (2\,\boldsymbol{\Omega} + \hat{\boldsymbol{z}}\,\widetilde{\zeta})\wedge\boldsymbol{u} = -\nabla_\sigma K - \nabla_z\Phi - (1/\rho)\,\nabla_z p + \boldsymbol{F}^h, \tag{38.17}$$

where again $\nabla_z = \nabla_\sigma - (\nabla_\sigma z)\,\partial_z$ as per equation (38.7). This equation is form-invariant regardless the horizontal coordinates, thus motivating the name *vector-invariant*.[1]

38.1.6 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 9.12) with g assumed to be a constant effective gravitational acceleration, then the horizontal momentum equation (38.8) becomes

$$\rho\,\frac{\mathrm{D}\boldsymbol{u}}{\mathrm{D}t} + 2\,\rho\,\boldsymbol{\Omega}\wedge\boldsymbol{u} = -[\nabla_\sigma - (\nabla_\sigma z)\,\partial_z]\,p + \rho\boldsymbol{F}^h. \tag{38.18}$$

Furthermore, assuming a hydrostatic balance (and corresponding simplification of the Coriolis acceleration as per Section 24.1.3) allows us to write $\partial p/\partial z = -g\,\rho$ so that

$$\rho\,\frac{\mathrm{D}\boldsymbol{u}}{\mathrm{D}t} + \rho\,f\,\hat{\boldsymbol{z}}\wedge\boldsymbol{u} = -[\nabla_\sigma p + \rho\,\nabla_\sigma\Phi] + \rho\boldsymbol{F}^h, \tag{38.19}$$

which also takes on the vector-invariant form

$$\left[\frac{\partial\boldsymbol{u}}{\partial t}\right]_\sigma + \dot{\sigma}\,\frac{\partial\boldsymbol{u}}{\partial\sigma} + (f + \widetilde{\zeta})\,\hat{\boldsymbol{z}}\wedge\boldsymbol{u} = -\nabla_\sigma(K+\Phi) - (1/\rho)\,\nabla_\sigma p + \boldsymbol{F}^h. \tag{38.20}$$

38.2 Concerning the pressure force

As seen in Section 19.1.3, the pressure force acting on a fluid region is given by the integral

$$\boldsymbol{F}^{\text{pressure}} = -\oint_{\partial\mathcal{R}} p\,\hat{\boldsymbol{n}}\,\mathrm{d}\mathcal{S} = -\int_{\mathcal{R}}\nabla p\,\mathrm{d}V, \tag{38.21}$$

[1] See Section 4.4.4 of *Griffies* (2004) for a detailed derivation using arbitrary horizontal coordinates.

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 38.1.6 and 38.2.1. 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 (38.21) 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 38.2. 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.

38.2.1 Computing the horizontal pressure gradient

Figure 38.1: *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 (38.7) can lead to numerical difficulties. For example, with a simple geopotential and a hydrostatic fluid, equation (38.19) shows that the horizontal pressure gradient takes the form

$$\nabla_z p = \nabla_\sigma p + \rho\,\nabla_\sigma \Phi = \nabla_\sigma p + g\,\rho\,\nabla_\sigma z, \tag{38.22}$$

with this decomposition illustrated in Figure 38.1 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 (38.21) 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.

38.2.2 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 38.2 the pressure force acts in the positive $\hat{\boldsymbol{y}}$ direction

$$\boldsymbol{F}^{\text{press}}_{\text{left}} = \hat{\boldsymbol{y}} \int_{\text{left}} p \, \mathrm{d}x \, \mathrm{d}z \tag{38.23}$$

whereas pressure on the right wall acts in the opposite direction

$$\boldsymbol{F}^{\text{press}}_{\text{right}} = -\hat{\boldsymbol{y}} \int_{\text{right}} p \, \mathrm{d}x \, \mathrm{d}z. \tag{38.24}$$

Similar expressions appear for the front and back vertical boundaries acting in the $\hat{\boldsymbol{x}}$ 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

$$\hat{\boldsymbol{n}} = \frac{\nabla(z - \eta)}{|\nabla(z - \eta)|} = \frac{\hat{\boldsymbol{z}} - \nabla\eta}{\sqrt{1 + |\nabla\eta|^2}}. \tag{38.25}$$

Following our discussion of dia-surface transport in Section 17.3.5, we know that the product of the normal direction and the area element can be written

$$\hat{\boldsymbol{n}} \, \mathrm{d}\mathcal{S} = (\hat{\boldsymbol{z}} - \nabla\eta) \, \mathrm{d}A, \tag{38.26}$$

where $\mathrm{d}A = \mathrm{d}x \, \mathrm{d}y$ is the horizontal projection of the area element (see Figure 17.1). Hence, the net pressure force acting on the top face is given by

$$\boldsymbol{F}^{\text{press}}_{\text{top}} = -\hat{\boldsymbol{z}} \int_{\text{top}} p \, \mathrm{d}x \, \mathrm{d}y + \hat{\boldsymbol{x}} \int_{\text{top}} p \, (\partial z/\partial x)_\sigma \, \mathrm{d}x \, \mathrm{d}y + \hat{\boldsymbol{y}} \int_{\text{top}} p \, (\partial z/\partial y)_\sigma \, \mathrm{d}x \, \mathrm{d}y, \tag{38.27}$$

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

$$\boldsymbol{F}^{\text{press}}_{\text{bott}} = \hat{\boldsymbol{z}} \int_{\text{bott}} p \, \mathrm{d}x \, \mathrm{d}y - \hat{\boldsymbol{x}} \int_{\text{bott}} p \, (\partial z/\partial x)_\sigma \, \mathrm{d}x \, \mathrm{d}y - \hat{\boldsymbol{y}} \int_{\text{bott}} p \, (\partial z/\partial y)_\sigma \, \mathrm{d}x \, \mathrm{d}y. \tag{38.28}$$

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.

38.2.3 Net vertical pressure force

Bringing the pieces together leads to the net vertical pressure force acting on the grid cell

$$\boldsymbol{F}^{\text{press}}_{\text{vertical}} = -\hat{\boldsymbol{z}} \left[\int_{\text{top}} p \, \mathrm{d}x \, \mathrm{d}y - \int_{\text{bott}} p \, \mathrm{d}x \, \mathrm{d}y \right]. \tag{38.29}$$

If the fluid is in hydrostatic balance, then this vertical force is given by the weight of fluid within the cell

$$\boldsymbol{F}^{\text{press}}_{\text{vertical}} = \hat{\boldsymbol{z}} \, M g, \tag{38.30}$$

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.

38.2.4 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

$$\boldsymbol{F}^{\text{press}}_{\text{merid}} = \left[\int_{\text{left}} p \, \mathrm{d}x \, \mathrm{d}z - \int_{\text{right}} p \, \mathrm{d}x \, \mathrm{d}z \right] + \left[\int_{\text{top}} p \, (\partial z / \partial y)_\sigma \, \mathrm{d}x \, \mathrm{d}y - \int_{\text{bott}} p \, (\partial z / \partial y)_\sigma \, \mathrm{d}x \, \mathrm{d}y \right]. \tag{38.31}$$

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 $(\partial z / \partial y)_\sigma \, \mathrm{d}y = \mathrm{d}z$ on the top and bottom faces, so that

$$\boldsymbol{F}^{\text{press}}_{\text{merid}} = - \oint p \, \mathrm{d}x \, \mathrm{d}z. \tag{38.32}$$

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

$$\boldsymbol{F}^{\text{press}}_{\text{merid}} = - \oint p \, \mathrm{d}x \, \mathrm{d}z = - \oint \mathrm{d}x \, [\mathrm{d}(p \, z) - z \, \mathrm{d}p] = g^{-1} \oint \Phi \, \mathrm{d}x \, \mathrm{d}p, \tag{38.33}$$

where $\oint \mathrm{d}x \, \mathrm{d}(p \, z) = 0$. This form is useful with compressible / non-Boussinesq models, in which pressure is a natural vertical coordinate (e.g., see the caption to Figure 38.1).

38.2.5 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.

38.3 Hydrostatic vorticity and potential vorticity

Generalized vertical coordinates are most commonly used to study hydrostatic fluids. In particular, we develop the equation for the vertical component of vorticity, $\widetilde{\zeta}$, in a hydrostatic fluid as written using generalized vertical coordinates.

Figure 38.2: *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 $\sigma(x, y, z, t) =$ 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.*

38.3.1 Basic manipulations

Our starting point is the vector-invariant velocity equation given by equation (38.20)

$$\left[\frac{\partial \boldsymbol{u}}{\partial t}\right]_\sigma + \dot{\sigma}\,\frac{\partial \boldsymbol{u}}{\partial \sigma} + \widetilde{\zeta}_{\mathrm{a}}\,\hat{\boldsymbol{z}} \wedge \boldsymbol{u} = -\nabla_\sigma(K + \Phi) - (1/\rho)\,\nabla_\sigma p + \boldsymbol{F}^h, \tag{38.34}$$

where $\widetilde{\zeta}_{\mathrm{a}} = \widetilde{\zeta} + f$ is the absolute vorticity. Taking the curl of this equation and projecting onto the vertical direction leads to

$$\frac{\mathrm{D}\widetilde{\zeta}_{\mathrm{a}}}{\mathrm{D}t} = -\widetilde{\zeta}_{\mathrm{a}}\,\nabla_\sigma \cdot \boldsymbol{u} + \frac{\hat{\boldsymbol{z}} \cdot (\nabla_\sigma \rho \wedge \nabla_\sigma p)}{\rho^2} + \hat{\boldsymbol{z}} \cdot \left[\frac{\partial \boldsymbol{u}}{\partial \sigma} \wedge \nabla_\sigma \dot{\sigma} + \nabla_\sigma \wedge \boldsymbol{F}^h\right] \tag{38.35}$$

where we noted that the planetary vorticity, f, is independent of time and vertical position.

Making use of the mass conservation equation

We now make use of mass conservation written in the form of equation (17.85)

$$\frac{1}{\rho\,\hbar}\,\frac{\mathrm{D}(\rho\,\hbar)}{\mathrm{D}t} = -(\nabla_\sigma \cdot \boldsymbol{u} + \partial\dot{\sigma}/\partial\sigma), \tag{38.36}$$

to render

$$\rho\,\hbar\,\frac{\mathrm{D}}{\mathrm{D}t}\left[\frac{\widetilde{\zeta}_{\mathrm{a}}}{\rho\,\hbar}\right] = \frac{\hat{\boldsymbol{z}} \cdot (\nabla_\sigma \rho \wedge \nabla_\sigma p)}{\rho^2} + \widetilde{\zeta}_{\mathrm{a}}\,\frac{\partial \dot{\sigma}}{\partial \sigma} + \hat{\boldsymbol{z}} \cdot \left[\frac{\partial \boldsymbol{u}}{\partial \sigma} \wedge \nabla_\sigma \dot{\sigma} + \nabla_\sigma \wedge \boldsymbol{F}^h\right]. \tag{38.37}$$

Massaging the $\dot{\sigma}$ terms

The terms containing $\dot{\sigma}$ can be written in the form

$$\widetilde{\zeta}_{\mathrm{a}}\,\partial_\sigma\dot{\sigma} + \hat{\boldsymbol{z}}\cdot(\partial_\sigma\boldsymbol{u}\wedge\nabla_\sigma\dot{\sigma}) = \widetilde{\zeta}_{\mathrm{a}}\,\partial_\sigma\dot{\sigma} + \hat{\boldsymbol{z}}\cdot[-\nabla_\sigma\wedge(\dot{\sigma}\,\partial_\sigma\boldsymbol{u}) + \dot{\sigma}\,\nabla_\sigma\wedge\partial_\sigma\boldsymbol{u}] \tag{38.38a}$$

$$= \widetilde{\zeta}_{\mathrm{a}}\,\partial_\sigma\dot{\sigma} + \dot{\sigma}\,\partial_\sigma\widetilde{\zeta}_{\mathrm{a}} - \hat{\boldsymbol{z}}\cdot[\nabla_\sigma\wedge(\dot{\sigma}\,\partial_\sigma\boldsymbol{u})] \tag{38.38b}$$

$$= \partial_\sigma(\dot{\sigma}\,\widetilde{\zeta}_{\mathrm{a}}) - \hat{\boldsymbol{z}}\cdot[\nabla_\sigma\wedge(\dot{\sigma}\,\partial_\sigma\boldsymbol{u})]\,. \tag{38.38c}$$

38.3.2 Vorticity and potential vorticity equation

The above results allow us to write equation (38.37) in the form

$$\rho\,\hbar\,\frac{\mathrm{D}Q}{\mathrm{D}t} = \frac{\hat{\boldsymbol{z}}\cdot(\nabla_\sigma\rho\wedge\nabla_\sigma p)}{\rho^2} + \partial_\sigma(\dot{\sigma}\,\widetilde{\zeta}_{\mathrm{a}}) + \nabla_\sigma\cdot\left[\hat{\boldsymbol{z}}\wedge\dot{\sigma}\,\partial_\sigma\boldsymbol{u} - \hat{\boldsymbol{z}}\wedge\boldsymbol{F}^h\right], \tag{38.39}$$

where we introduced the potential vorticity defined according to the generalized vertical coordinates

$$Q = \frac{\widetilde{\zeta}_{\mathrm{a}}}{\hbar\,\rho}. \tag{38.40}$$

The potential vorticity equation (38.39) has a generally nonzero baroclinicity contribution (see Section 34.4 for more on baroclinicity)

$$\frac{\hat{\boldsymbol{z}}\cdot(\nabla_\sigma\rho\wedge\nabla_\sigma p)}{\rho^2} \tag{38.41}$$

so that the potential vorticity (38.40) is generally not materially invariant even for a perfect fluid.

Pressure coordinates

The baroclinicity (38.41) vanishes when choosing $\sigma = p$. We already noted this property in our general discussion of baroclinicity in Section 34.4. This choice, however, does not offer us a materially invariant potential vorticity since $\dot{\sigma} = \dot{p}$ does not generally vanish. 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 $\partial_\sigma(\dot{\sigma}\,\widetilde{\zeta}_{\mathrm{a}})$ term does not.

Eulerian flux-form equation

Expanding the material time derivative in equation (38.39) allows us to cancel the $\partial_\sigma(\dot{\sigma}\,\widetilde{\zeta}_{\mathrm{a}})$ 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

$$\left[\frac{\partial(\rho\,\hbar\,Q)}{\partial t}\right]_\sigma = \frac{\hat{\boldsymbol{z}}\cdot(\nabla_\sigma\rho\wedge\nabla_\sigma p)}{\rho^2} - \nabla_\sigma\cdot\left[\rho\,\hbar\,\boldsymbol{u}\,Q + \hat{\boldsymbol{z}}\wedge\dot{\sigma}\,\partial_\sigma\boldsymbol{u} + \hat{\boldsymbol{z}}\wedge\boldsymbol{F}^h\right], \tag{38.42}$$

which, since $\widetilde{\zeta}_{\mathrm{a}} = \hbar\,\rho\,Q$, is equivalent to the absolute vorticity equation

$$\left[\frac{\partial\widetilde{\zeta}_{\mathrm{a}}}{\partial t}\right]_\sigma = \frac{\hat{\boldsymbol{z}}\cdot(\nabla_\sigma\rho\wedge\nabla_\sigma p)}{\rho^2} - \nabla_\sigma\cdot\left[\boldsymbol{u}\,\widetilde{\zeta}_{\mathrm{a}} + \hat{\boldsymbol{z}}\wedge\dot{\sigma}\,\partial_\sigma\boldsymbol{u} + \hat{\boldsymbol{z}}\wedge\boldsymbol{F}^h\right]. \tag{38.43}$$

As a sanity check, we note that setting $\sigma = z$ so that $\hbar = 1$ reduces the vorticity equation (38.43) to the vertical component of the vorticity equation (34.30) (see Exercise 38.1).

38.3.3 Boussinesq fluid

Recall our discussion of vorticity for the Boussinesq fluid in Section 34.7, 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 (38.34) and setting the factor $1/\rho$ multiplying the pressure gradient to $1/\rho_0$ as part of the Boussinesq approximation

$$(1/\rho)\,\nabla_\sigma p \longrightarrow (1/\rho_0)\,\nabla_\sigma p, \tag{38.44}$$

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

$$Q = \frac{\widetilde{\zeta}_{\rm a}}{\hbar} \tag{38.45}$$

which satisfies the material and Eulerian evolution equations

$$\hbar\,\frac{\mathrm{D}Q}{\mathrm{D}t} = \partial_\sigma(\dot\sigma\,\widetilde{\zeta}_{\rm a}) + \nabla_\sigma\cdot\left[\hat{\boldsymbol{z}}\wedge\dot\sigma\,\partial_\sigma\boldsymbol{u} - \hat{\boldsymbol{z}}\wedge\boldsymbol{F}^h\right] \tag{38.46}$$

$$\left[\frac{\partial(\hbar\,Q)}{\partial t}\right]_\sigma = -\nabla_\sigma\cdot\left[\hbar\,\boldsymbol{u}\,Q + \hat{\boldsymbol{z}}\wedge\dot\sigma\,\partial_\sigma\boldsymbol{u} + \hat{\boldsymbol{z}}\wedge\boldsymbol{F}^h\right]. \tag{38.47}$$

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.

38.4 Exercise

EXERCISE 38.1: CHECKING THE VORTICITY EQUATION

Verify that for $2\,\boldsymbol{\Omega} = f\,\hat{\boldsymbol{z}}$ the choice $\sigma = z$ reduces the vorticity equation (38.43) to the vertical component of the vorticity equation (34.30).

Part VII

Balanced models

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.

Perhaps the simplest balance model of geophysical relevance is the two-dimensional barotropic model. We develop the theory for this model in Chapter 39. The nuts and bolts of the next two chapters 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 different aspects of ocean and atmospheric 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 including geostrophic turbulence. PG and QG represent two examples of *balance models*, in which knowledge of potential vorticity is sufficient to determine the stratification, pressure, and velocity.

39

Two-dimensional barotropic flow

A single layer of shallow water fluid is among the simplest models available for the study of rotating flows. However, the presence of divergence adds analytical complexity that may not be necessary physically nor desirable analytically. Additionally, we may choose to focus on the low frequency planetary (Rossby wave) modes rather than the divergent and higher frequency gravity wave mode (Section 31.3). For these reasons, we introduce in this chapter a fluid dynamical model in which the two-dimensional velocity has zero divergence. The resulting set of equations forms the *two-dimensional barotropic model*, which is a well used theoretical model of geophysical flows. Vorticity is the primary dynamical field for the two-dimensional barotropic model, with its knowledge sufficient to fully determine the velocity.

With no vertical stratification and a vanishing horizontal divergence, the vertical velocity in the two-dimensional barotropic model is formally zero. As stated in the discussion of shallow water gravity waves in Section 31.3, gravity waves are supported by a nonzero horizontal flow convergence. Hence, the two-dimensional barotropic model does not support gravity waves given the absence of horizontal divergence. The gravity wave speed has effectively been set to infinite. We emphasize that the absence of vertical stratification for the two-dimensional barotropic fluid is *not* the reason for the absence of gravity waves. Indeed, a single layer of shallow water fluid has no stratification given that it has uniform density and yet it supports gravity waves due to the horizonally divergent nature of the shallow water flow (Section 31.3).

READER'S GUIDE FOR THIS CHAPTER

The two-dimensional barotropic model offers a suitable starting point for the study of balanced flow. Additionally, some work with the two-dimensional barotropic model was motivated by studies of coherent vortex structures, such as those found by the simulations documented in *McWilliams* (1984), with Chapter 3 of *McWilliams* (2006) exploring analytical vortex solutions.

39.1 Loose threads **694**
39.2 Governing equations **694**
39.3 Vorticity and potential vorticity **695**
39.3.1 Vorticity equation 695
39.3.2 Jacobian form of vorticity advection 695
39.3.3 Potential vorticity equation 696
39.3.4 Poisson equation for the streamfunction 696
39.4 Pressure **696**

39.4.1 Poisson equation for pressure 696
39.4.2 Self-advection source in the pressure equation 697
39.4.3 Sample pressure gradients 698
39.4.4 Further study 700
39.5 Exercises **700**

39.1 Loose threads

- Rossby waves and wave-mean interactions
- Eliassan-Palm flux
- Analytic vortex solutions

39.2 Governing equations

The governing equations for the 2d-barotropic model are based on the shallow water equations yet with the horizontal velocity assumed to be non-divergent

$$\frac{\mathrm{D}\boldsymbol{u}}{\mathrm{D}t} + f\,\hat{\boldsymbol{z}} \wedge \boldsymbol{u} = -\nabla_z \varphi \qquad \text{and} \qquad \nabla \cdot \boldsymbol{u} = 0, \tag{39.1}$$

where the pressure is normalized according to

$$\varphi = p/\rho_0 \tag{39.2}$$

with ρ_0 a constant reference density, and where material evolution occurs with the two-dimensional non-divergent flow

$$\frac{\mathrm{D}}{\mathrm{D}t} = \frac{\partial}{\partial t} + \boldsymbol{u} \cdot \nabla. \tag{39.3}$$

With the horizontal flow $\boldsymbol{u} = (u, v, 0)$ non-divergent, we can introduce a streamfunction

$$\boldsymbol{u} = \hat{\boldsymbol{z}} \wedge \nabla\psi \qquad \Longrightarrow \qquad u = -\frac{\partial \psi}{\partial y} \qquad \text{and} \qquad v = \frac{\partial \psi}{\partial x}. \tag{39.4}$$

We close this section by noting that the the velocity self-advection can be written as the divergence of a 2×2 symmetric tensor

$$-\left(\boldsymbol{u} \cdot \nabla\right) \boldsymbol{u} = \nabla \cdot \mathbb{E}, \tag{39.5}$$

where

$$-\,\mathbb{E} = \begin{bmatrix} u^2 & u\,v \\ u\,v & v^2 \end{bmatrix}, \tag{39.6}$$

thus bringing the momentum equation (39.1) to the form

$$\frac{\partial \boldsymbol{u}}{\partial t} + f\,\hat{\boldsymbol{z}} \wedge \boldsymbol{u} = -\nabla_z \varphi + \nabla \cdot \mathbb{E}. \tag{39.7}$$

The tensor $\rho_0\,\mathbb{E}$ is the *kinetic stress tensor* arising from velocity self-advection. We introduced the three-dimensional form of this stress in Section 19.4.

39.3 Vorticity and potential vorticity

The vertical component of the relative vorticity is given by

$$\zeta = \hat{\boldsymbol{z}} \cdot (\nabla \wedge \boldsymbol{u}) = \frac{\partial v}{\partial x} - \frac{\partial u}{\partial y} = \nabla^2 \psi. \tag{39.8}$$

There are many applications of the 2d barotropic model on the sphere. We are concerned here with the case of a tangent plane configuration in which the Coriolis parameter is given by the β-plane approximation

$$f = f_0 + \beta\,(y - y_0). \tag{39.9}$$

39.3.1 Vorticity equation

To form the vorticity equation, take the zonal derivative of the meridional momentum equation and meridional derivative of the zonal momentum equation (see equation (39.1)) to arrive at

$$\frac{\partial}{\partial t}\left(\frac{\partial v}{\partial x} - \frac{\partial u}{\partial y}\right) + \frac{\partial}{\partial x}\left[\nabla \cdot (\boldsymbol{u}\, v)\right] - \frac{\partial}{\partial y}\left[\nabla \cdot (\boldsymbol{u}\, u)\right] + f \nabla \cdot \boldsymbol{u} + \beta\, v = 0. \tag{39.10}$$

Notice how the pressure gradient force dropped out as a result of there being zero baroclinicity for the barotropic flow (recall Section 34.4 for a study of how baroclinicity affects vorticity). Straightforward manipulations, and use of the non-divergence condition $\nabla \cdot \boldsymbol{u} = 0$, then leads to the identity

$$\frac{\partial}{\partial x}\left[\nabla \cdot (\boldsymbol{u}\, v)\right] - \frac{\partial}{\partial y}\left[\nabla \cdot (\boldsymbol{u}\, u)\right] = \boldsymbol{u} \cdot \nabla \zeta, \tag{39.11}$$

so that the 2d vorticity equation takes the form

$$\frac{\partial \zeta}{\partial t} + \boldsymbol{u} \cdot \nabla \zeta + \beta\, v = 0, \tag{39.12}$$

which takes on the material evolution form

$$\frac{\mathrm{D}\zeta}{\mathrm{D}t} = -\beta\, v. \tag{39.13}$$

The material evolution of vorticity in the 2d-barotropic fluid is only affected by meridional advection of planetary vorticity. Since $\beta > 0$ over the globe, we see that northward flow ($v > 0$) produces a negative source for relative vorticity following a fluid particle. This source term is the β-effect that we return to in Section 39.3.3 when studying the potential vorticity.

39.3.2 Jacobian form of vorticity advection

It is sometimes convenient to write the advection operator in the form

$$\begin{aligned}
\boldsymbol{u} \cdot \nabla \zeta &= u\, \partial_x \zeta + v\, \partial_y \zeta && (39.14a)\\
&= -\partial_y \psi\, \partial_x \zeta + \partial_x \psi\, \partial_y \zeta && (39.14b)\\
&= \hat{\boldsymbol{z}} \cdot (\nabla \psi \wedge \nabla \zeta) && (39.14c)\\
&\equiv J(\psi, \zeta), && (39.14d)
\end{aligned}$$

where J is the Jacobian operator

$$J(A, B) = \hat{\boldsymbol{z}} \cdot (\nabla A \wedge \nabla B) = \frac{\partial A}{\partial x}\frac{\partial B}{\partial y} - \frac{\partial A}{\partial y}\frac{\partial B}{\partial x}. \tag{39.15}$$

39.3.3 Potential vorticity equation

For 2d non-divergent flow, the potential vorticity

$$q = \zeta + f \tag{39.16}$$

is the same as absolute vorticity. Since f is time independent, we can write the vorticity equation in the alternative form

$$\frac{\partial \zeta}{\partial t} = -\nabla \cdot (\boldsymbol{u}\, q). \tag{39.17}$$

Similarly, we have the potential vorticity equation

$$\frac{\partial q}{\partial t} = -\nabla \cdot (\boldsymbol{u}\, q) \Longleftrightarrow \frac{\mathrm{D}q}{\mathrm{D}t} = 0. \tag{39.18}$$

To maintain a materially constant potential vorticity requires the relative vorticity to 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 (39.13). Furthermore, this behavior is a reflection of the β-effect studied in Section 34.6.2.

39.3.4 Poisson equation for the streamfunction

Given initial and boundary conditions, the barotropic vorticity equation (39.17) allows us to determine the evolution of vorticity. We can in turn invert the Poission equation (see Section 3.4 for discussion of the Poisson equation)

$$\nabla^2 \psi = \zeta \tag{39.19}$$

to determine the streamfunction and then the velocity field $\boldsymbol{u} = \hat{\boldsymbol{z}} \wedge \nabla \psi$. Hence, time integration of the barotropic potential vorticity equation is sufficient to fully specify time evolution of the 2d barotropic flow field. We do not need to explicitly determine pressure to determine the flow.

39.4 Pressure

The velocity field in a two-dimensional barotropic model is highly constrained. Notably, pressure forces are present yet we do not need their explicit expression to determine the flow evolution. Furthermore, in the absence of two-dimensional divergence, the free surface is absolutely flat even in the presence of topography. It is as if there was a flat rigid lid placed on the surface with a corresponding *lid pressure* maintaining the flat free surface. That is, pressure gradients exist in the barotropic fluid even though the surface is flat, with vertically homogeneous pressure gradients instantaneously maintaining the non-divergence constraint. This *rigid lid* approximation is commonly employed for studies of large-scale ocean flow.

39.4.1 Poisson equation for pressure

Even though we do not need the pressure to time step velocity, it is of interest to examine the pressure field to better understand the dynamical forces acting in the barotropic fluid. This analysis also provides insights into the three-dimensional incompressible Boussinesq fluid studied in Chapter 25.

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 (39.1) 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 pressure[1]

$$-\nabla^2\varphi = \partial_x[\nabla\cdot(\boldsymbol{u}\,u)] + \partial_y[\nabla\cdot(\boldsymbol{u}\,v)] - f\,\zeta + \beta\,u, \tag{39.20}$$

where we set

$$\nabla\cdot\partial_t\boldsymbol{u} = \partial_t\nabla\cdot\boldsymbol{u} = 0. \tag{39.21}$$

Given sufficient boundary conditions, the elliptic partial differential equation (39.20) can be inverted to find the pressure field (Section 3.4). In Exercise 25.2, we encountered a similar elliptic problem for the pressure in a three-dimensional incompressible fluid. 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 (39.20) 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 fall into shallow minima, making it difficult to find the true 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.

39.4.2 Self-advection source in the pressure equation

Following our discussion of three-dimensional incompressible flow in Exercise 25.2, we can write the self-advection source appearing in the pressure equation (39.20) in the form

$$\partial_x[\nabla\cdot(\boldsymbol{u}\,u)] + \partial_y[\nabla\cdot(\boldsymbol{u}\,v)] = \mathbb{S}_{mn}\,\mathbb{S}_{mn} - \mathbb{A}_{mn}\,\mathbb{A}_{mn}, \tag{39.22}$$

where the rate of strain tensor, $\mathbb{S}$, and rotation tensor, $\mathbb{A}$, have components given by equations (13.22a) and (13.22b)

$$\mathbb{S}_{mn} = \frac{1}{2}\left[\frac{\partial v_m}{\partial x_n} + \frac{\partial v_n}{\partial x_m}\right] = \mathbb{S}_{nm} \qquad \text{rate of strain tensor} \tag{39.23}$$

$$\mathbb{A}_{mn} = \frac{1}{2}\left[\frac{\partial v_m}{\partial x_n} - \frac{\partial v_n}{\partial x_m}\right] = -\mathbb{A}_{nm} \qquad \text{rotation tensor.} \tag{39.24}$$

For the two-dimensional flow the rotation tensor is related to the relative vorticity via

$$\mathbb{A}_{mn} = -\epsilon_{mn}\,\zeta/2, \tag{39.25}$$

where ϵ_{mn} is the anti-symmetric Levi-Civita permutation symbol (see Section 1.4.1)

$$\epsilon_{mn} = \begin{bmatrix} \epsilon_{11} & \epsilon_{12} \\ \epsilon_{21} & \epsilon_{22} \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}. \tag{39.26}$$

The identity (39.22) indicates that strain provides a positive source to the Poisson equation (39.20) whereas vorticity provides a negative source. We support this result by considering two idealized flows in Section 39.4.3.

[1] We maintain the minus sign on the left hand side of equation (39.20) 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.

Figure 39.1: *Left panel: example flow field from solid-body rotation,* $\boldsymbol{u} = \boldsymbol{\Omega} \wedge \boldsymbol{x} = (-y, x)$*, which has vorticity* $\nabla \wedge \boldsymbol{u} = 2\,\boldsymbol{\Omega}$ *and zero strain. Right panel: corresponding acceleration from self-advection,* $-(\boldsymbol{u} \cdot \nabla)\,\boldsymbol{u} = \Omega^2\,\boldsymbol{x}$*. The units are arbitrary.*

39.4.3 Sample pressure gradients

We here offer two examples to help support our understanding of how pressure gradients arise in a 2d-barotropic flow in a manner that maintains, at each time instance, a non-divergent flow field. For simplicity we just consider the non-rotating case so that it is sufficient to examine how pressure responds to the self-advection term.

Solid-body flow

Consider a velocity field in a flat domain that is initialized in circular solid-body motion (Figure 39.1)

$$\boldsymbol{u} = \boldsymbol{\Omega} \wedge \boldsymbol{x} = \Omega\,(-y\,\hat{\boldsymbol{x}} + x\,\hat{\boldsymbol{y}}), \tag{39.27}$$

where $\boldsymbol{\Omega} = \Omega\,\hat{\boldsymbol{z}}$ is a constant angular velocity. This flow has zero strain and yet a non-zero vorticity

$$\hat{\boldsymbol{z}} \cdot (\nabla \wedge \boldsymbol{u}) = \zeta = 2\,\Omega \Longrightarrow -\mathbb{A}_{mn}\,\mathbb{A}_{mn} = -\Omega^2. \tag{39.28}$$

The velocity time tendency from self-advection acceleration is given by

$$-\,(\boldsymbol{u} \cdot \nabla)\,\boldsymbol{u} = \Omega^2\,\boldsymbol{x}, \tag{39.29}$$

which is the outward directed centrifugal acceleration associated with the circular solid-body motion.[2] 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, the solid-body flow is an exact steady solution for the two-dimensional non-rotating barotropic system.

The pressure field is found by solving the Poisson equation

$$-\,\nabla^2\varphi = -\Omega^2. \tag{39.30}$$

[2] We considered the more general case in Section 27.1.2 when decomposing the material acceleration for two-dimensional flow into natural coordinates.

Figure 39.2: *Left panel: example purely strained flow field with zero vorticity,* $\boldsymbol{u} = \boldsymbol{\Omega}\,(-x\,\hat{\boldsymbol{x}} + y\,\hat{\boldsymbol{y}})$*. Right panel: corresponding converging acceleration from self-advection,* $-(\boldsymbol{u}\cdot\nabla)\,\boldsymbol{u} = -\Omega^2\,\boldsymbol{x}$*. The units are arbitrary.*

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

$$\varphi = (\Omega\, r)^2/2, \tag{39.31}$$

where we used the cylindrical expression for the Laplacian operator (equation (7.83b)) and set $\varphi(r = 0) = 0$. To help understand the physics of this parabolic pressure field, recall the analysis in Section 29.3 of a solid-body rotating shallow water layer in a cylindrical tank. In contrast to the barotropic system, the horizontal velocity in a shallow water layer is divergent so that the layer thickness is not constrained to remain flat. Hence, the centrifugal acceleration causes the shallow water 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 shallow water layer has a parabolic free surface with a minimum at the center (see equation (29.57)). The parabolic shaped free surface creates a pressure field that precisely corresponds to the pressure field (39.31) in the two-dimensional barotropic system. The shallow water transient adjustment (whose linear form occurs through the gravity waves discussed in Chapter 31) required to reach a steady state occurs instantaneously in the two-dimensional barotropic fluid.

Irrotational flow

Now consider the following pure strain flow (Figure 39.2)

$$\boldsymbol{u} = \Omega\,(-x\,\hat{\boldsymbol{x}} + y\,\hat{\boldsymbol{y}}), \tag{39.32}$$

whose vorticity vanishes and whose self-advection acceleration is given by

$$-\,(\boldsymbol{u}\cdot\nabla)\,\boldsymbol{u} = -\Omega^2\,\boldsymbol{x} \Longrightarrow \mathbb{S}_{mn}\,\mathbb{S}_{mn} = \Omega^2. \tag{39.33}$$

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 to exactly counteracting the converging acceleration from the strained flow. 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 two-dimensional non-rotating barotropic flow.

39.4.4 Further study

Bryan (1969) provided the first working numerical algorithm to determine the ocean general circulation model. Bryan's method made use of the rigid lid approximation of Section 39.4 so that the depth integrated velocity is non-divergent. 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)).

The discussion of pressure in Section 39.4.3 is based on a similar presentation in Appendix B of *Jeevanjee and Romps* (2015).

39.5 Exercises

EXERCISE 39.1: 2D BAROTROPIC MODEL AND GRAVITY WAVES

Are there gravity waves for the 2d barotropic model described in Section 39.2. Why? Hint: recall the discussion of gravity waves for the shallow water system in Section 31.3. A one-sentence answer is sufficient.

EXERCISE 39.2: INTEGRAL PROPERTIES OF THE INVISCID 2D BAROTROPIC MODEL

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).

(a) Show that the domain integrated kinetic energy per mass remains constant in time

$$\mathcal{K} = \frac{1}{2} \int_{\mathcal{S}} \boldsymbol{u} \cdot \boldsymbol{u} \, \mathrm{d}\mathcal{S}, \tag{39.34}$$

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

$$\mathcal{C} = \int_{\mathcal{S}} \zeta \, \mathrm{d}\mathcal{S}. \tag{39.35}$$

(d) Show that the domain integrated enstrophy is constant in time for f-plane motion ($\beta = 0$)

$$\mathrm{Z}^{(\zeta)} = \int_{\mathcal{S}} \zeta^2 \, \mathrm{d}\mathcal{S}. \tag{39.36}$$

(e) Show that the domain integrated potential enstrophy is constant in time even with $\beta \neq 0$

$$\mathrm{Z}^{(q)} = \int_{\mathcal{S}} q^2 \, \mathrm{d}\mathcal{S}. \tag{39.37}$$

EXERCISE 39.3: CIRCULATION IN A 2D BAROTROPIC FLOW

Consider a two-dimensional barotropic flow on a β-plane in the presence of a biharmonic friction operator, where the governing vorticity equation is

$$\frac{\partial \zeta}{\partial t} + J(\psi, \zeta + \beta y) = -\nu \nabla^4 \zeta, \tag{39.38}$$

with $\nu > 0$ a constant biharmonic viscosity with dimensions of $\mathrm{L}^4\ \mathrm{T}^{-1}$. Show that the circulation around a fixed material area, $\mathcal{S}$, in the fluid evolves according to

$$\frac{\mathrm{d}\mathcal{C}}{\mathrm{d}t} = -\oint_{\partial \mathcal{S}} \left[\psi \frac{\partial q}{\partial s} + \nu \frac{\partial (\nabla^2 \zeta)}{\partial n} \right] \mathrm{d}s, \tag{39.39}$$

where s is the arc-length along the boundary of the region and n is a coordinate normal to the boundary.

EXERCISE 39.4: 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 12.1. Here, we develop a similar equation for the gradient of relative vorticity in a two-dimensional barotropic flow. For this purpose, consider the inviscid barotropic vorticity equation on an f-plane

$$\frac{\partial \zeta}{\partial t} + J(\psi, \zeta) = 0. \tag{39.40}$$

(a) Show that the material evolution of the vorticity gradient is given by

$$\frac{\mathrm{D}(\nabla \zeta)}{\mathrm{D}t} = -J(\nabla \psi, \zeta). \tag{39.41}$$

(b) Show that the material evolution of the squared vorticity gradient is given by

$$\frac{\mathrm{D}|\nabla \zeta|^2}{\mathrm{D}t} = 2\, J(\zeta, \nabla \psi) \cdot \nabla \zeta. \tag{39.42}$$

EXERCISE 39.5: 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 30.5.

As in Section 30.5.1, the relative angular momentum for a region of fluid is given by

$$\boldsymbol{L} = \int \mathrm{d}\mathcal{S} \int (\boldsymbol{x} \wedge \boldsymbol{v})\, \rho\, \mathrm{d}z, \tag{39.43}$$

where $\boldsymbol{x}$ 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

$$\boldsymbol{L} = \rho\, H \int_{\mathcal{S}} (\boldsymbol{x} \wedge \boldsymbol{u})\, \mathrm{d}\mathcal{S}, \tag{39.44}$$

with $\boldsymbol{u}$ 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

$$L^z = \rho\, H \int_{\mathcal{S}} \hat{\boldsymbol{z}} \cdot (\boldsymbol{x} \wedge \boldsymbol{u})\, \mathrm{d}\mathcal{S}. \tag{39.45}$$

Show for a simply connected and bounded region, L^z can be written

$$L^z = 2\,\rho\,H \int_{\mathcal{S}} (\psi_b - \psi)\,\mathrm{d}\mathcal{S} \tag{39.46}$$

where ψ is the streamfunction satisfying $\boldsymbol{u} = \hat{\boldsymbol{z}} \wedge \nabla\psi$, and ψ_b is the value of the streamfunction evaluated on the region boundary. Hint: note that $\nabla \cdot \boldsymbol{x} = 2$ for a horizontal position vector. Also recall from Section 16.4.2 that the streamfunction equals to a spatial constant when evaluated along the domain boundary.

40

Shallow water planetary and quasi-geostrophy

READER'S GUIDE FOR THIS CHAPTER

We here derive the mechanical equations for planetary geostrophy (PG) and quasi-geostrophy (QG) within the shallow water fluid system. The derivation requires elements from asymptotic analysis, with salient details introduced here. This chapter is largely technical in nature, with certain details put to use in the continuously stratified systems discussed in Chapters 41 and 42.

40.1 Scaling analysis and the Buckingham-Π theorem **703**
40.2 Shallow water equations **704**
40.2.1 Dimensional scales 704
40.2.2 Physical dimensions 705
40.2.3 Number of non-dimensional parameters 705
40.2.4 Choosing the non-dimensional parameters 706
40.2.5 Assumed values for the non-dimensional parameters 707
40.2.6 Deformation radius and the free surface undulation scale 708
40.2.7 Non-dimensional shallow water equations 709
40.3 Shallow water planetary geostrophy **711**
40.4 Shallow water quasi-geostrophy **711**
40.4.1 Quasi-geostrophic scaling 711
40.4.2 Outlining the asymptotic method 712
40.4.3 Zeroth order asymptotic equations 713
40.4.4 First order asymptotic equations 714
40.4.5 Dimensional equations and quasi-geostrophic PV 715
40.5 Exercises **716**

40.1 Scaling analysis and the Buckingham-Π theorem

Scaling analysis is ubiquitous in physics, with the Buckingham-Π theorem providing a useful framework for scaling. This 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-Π states that the number

of dimensionless parameters is

$$N_{\text{dimensionless}} = K - R. \tag{40.1}$$

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-Π 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. We focus here on the regime of large-scale atmospheric and oceanic flow where the 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.

40.2 Shallow water equations

A single-layer of inviscid shallow water fluid is governed by the momentum and thickness equations (Chapter 29)

$$\frac{\partial \boldsymbol{u}}{\partial t} + (\boldsymbol{u} \cdot \nabla)\, \boldsymbol{u} + \boldsymbol{f} \wedge \boldsymbol{u} = -g \nabla \eta \tag{40.2a}$$

$$\frac{\partial h}{\partial t} + \nabla \cdot (h\, \boldsymbol{u}) = 0, \tag{40.2b}$$

where (see Figure 29.1)

$$\eta = \eta_b + h = H + \Delta\eta. \tag{40.3}$$

40.2.1 Dimensional scales

We identify nine dimensional parameters for the shallow water system.

- LENGTH SCALES
  - ⋆ H = depth scale of the fluid; for the shallow water system, H is the area average fluid thickness (see Figure 29.1).
  - ⋆ 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 zonal (east-west) length scales can be longer than meridional (north-south) scales. Nonetheless, this choice does not preclude the dynamical emergence of anisotropic length scales. Indeed, by not *a priori* introducing anisotropic length scales, we ensure that the emergence of anisotropy naturally arises from the dynamics.
  - ⋆ 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 free surface height undulations, $\Delta\eta$.
- ⋆ $\mathcal{B}$ = scale for undulations of the bottom topography, η_b.

- VELOCITY SCALES
  - ⋆ U = velocity scale for fluid particle motion via advection; 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
  $$c = \sqrt{g\,H}. \tag{40.4}$$
  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.

- 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. But for our purposes, nine is all we are interested in for a single layer of inviscid shallow water fluid.

40.2.2 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 so that mass is described by area times height

$$M = \int \rho\,\mathrm{d}V\ [\equiv]\ L^2 H\,\rho. \tag{40.5}$$

Relatedly, we are unconcerned with the fluid density since it is uniform and does not explicitly appear in any of the governing equations.

40.2.3 Number of non-dimensional parameters

The Buckingham-Π theorem then says we have

$$N_{\text{dimensionless}} = 9 - 2 = 7 \tag{40.6}$$

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-Π 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 moving onward with the analysis.

40.2.4 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 set externally to the flow state itself.

1. VERTICAL TO HORIZONTAL ASPECT RATIO: The ratio of the vertical scale to the horizontal scale defines the aspect ratio

$$\delta_{\text{vertical/horizontal}} = \frac{\text{vertical depth scale}}{\text{horizontal length scale}} = \frac{H}{L}. \tag{40.7}$$

2. RATIO OF HORIZONTAL SCALE TO PLANETARY SCALE: The ratio of the lateral length scale to the planet radius is

$$\delta_{\text{horizontal/planet}} = \frac{\text{lateral length scale}}{\text{planetary length scale}} = \frac{L}{R}. \tag{40.8}$$

3. RATIO OF FREE SURFACE UNDULATION TO DEPTH: The ratio of the free surface undulation scale to the vertical depth scale is

$$\delta_{\text{free surface/depth}} = \frac{\text{free surface undulation scale}}{\text{vertical depth scale}} = \frac{\mathcal{H}}{H}. \tag{40.9}$$

4. RATIO OF BOTTOM TOPOGRAPHY UNDULATION TO FREE SURFACE UNDULATION: The ratio of the bottom topography undulation scale to the free surface undulation scale is

$$\delta_{\text{bottom/free surface}} = \frac{\text{bottom topography undulation scale}}{\text{free surface undulation scale}} = \frac{\mathcal{B}}{\mathcal{H}}. \tag{40.10}$$

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

$$Fr = \frac{U}{c} = \frac{U}{\sqrt{g\,H}}. \tag{40.11}$$

   Note that the Froude number is not directly utilized in the following, though its introduction is useful in other contexts.

6. ROSSBY NUMBER: The Rossby number is the ratio of the fluid particle acceleration scale to the Coriolis acceleration

$$Ro = \frac{\text{parcel acceleration}}{\text{Coriolis acceleration}}. \tag{40.12}$$

   The particle acceleration scale is determined by the local time tendency and advection

$$\text{particle acceleration} = \frac{\partial \boldsymbol{u}}{\partial t} + (\boldsymbol{u}\cdot\nabla)\,\boldsymbol{u} \tag{40.13a}$$

$$\sim \frac{U}{T} + \frac{U^2}{L}. \tag{40.13b}$$

   We assume that the time scale is determined by advection, so that

$$T \sim \frac{L}{U} \Rightarrow \frac{U^2}{L} = \frac{U}{T}, \tag{40.14}$$

in which case the Rossby number is given by

$$Ro = \frac{1}{f\,T} = \frac{U}{f\,L}. \tag{40.15}$$

Another interpretation for the Rossby number is the ratio of the relative vorticity to the planetary vorticity

$$Ro = \frac{\text{relative vorticity}}{\text{planetary vorticity}} \tag{40.16}$$

With the relative vorticity scaling as U/L and the planetary vorticity scaling as f, we recover the expression (40.15) for the Rossby number.

7. GEOSTROPHIC NUMBER: We define the geostrophic number as the ratio of the Coriolis acceleration to the pressure gradient acceleration[1]

$$Ge = \frac{\text{Coriolis acceleration}}{\text{pressure gradient acceleration}}. \tag{40.17}$$

The Coriolis acceleration scales as

$$\text{Coriolis acceleration} \sim f\,U \tag{40.18}$$

whereas the pressure gradient acceleration, $-g\,\nabla\eta$, scales as

$$\text{pressure gradient acceleration} \sim \frac{g\,\mathcal{H}}{L}, \tag{40.19}$$

so that

$$Ge = \frac{\text{Coriolis acceleration}}{\text{pressure gradient acceleration}} = \frac{f\,U}{(g/L)\,\mathcal{H}}. \tag{40.20}$$

40.2.5 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

$$\delta_{\text{vertical/horizontal}} \ll 1. \tag{40.21}$$

This assumption was made when formulating the shallow water system, which is based on hydrostatic balance (see Section 29.1). 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

$$\delta_{\text{horizontal/planet}} \ll 1 \qquad \text{quasi-geostrophy} \tag{40.22a}$$

$$\delta_{\text{horizontal/planet}} \sim 1 \qquad \text{planetary geostrophy.} \tag{40.22b}$$

[1] The geostrophic number is generally not introduced in the literature, since it will later be assumed equal to unity. We find it pedagogical to introduce it in order to enumerate the seven non-dimensional parameters available for the shallow water system.

3. RATIO OF FREE SURFACE UNDULATION TO DEPTH: This ratio will be implied by other scaling assumptions along with the dynamical equations.

4. RATIO OF BOTTOM TOPOGRAPHY UNDULATION TO FREE SURFACE UNDULATION: We assume that gradients in the bottom topography are small relative to gradients in the surface height. This assumption means that topographic effects on the flow are order Rossby number or smaller, which can be ensured if the scales for the bottom topography are much smaller than scales for the surface height
$$\delta_{\text{bottom/free surface}} \ll 1. \tag{40.23}$$
In Exercise 40.4 we remove this restriction on the bottom flow and consider the resulting constraint on the geostrophic flow.

5. FROUDE NUMBER: The Froude number is implied by sizes assumed for the other non-dimensional numbers.

6. SMALL ROSSBY NUMBER: The Rossby number is assumed small
$$Ro = \frac{U}{f\,L} \ll 1, \tag{40.24}$$
which means that the Coriolis acceleration is a leading order term in the horizontal momentum equation (40.2a).

7. UNIT GEOSTROPHIC NUMBER: The geostrophic number is assumed to be order unity
$$Ge \sim 1. \tag{40.25}$$
This assumption means that the Coriolis acceleration and pressure gradient acceleration scale together
$$f\,U \sim (g/L)\,\mathcal{H}. \tag{40.26}$$
Making use of the momentum equation (40.2a), we see that this scaling is consistent only so long as the Rossby number is small, $Ro \ll 1$. Furthermore, this scaling constrains the scale of the free surface undulation, $\mathcal{H}$, as we discuss in Section 40.2.6.

40.2.6 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, (40.26), to express the free surface undulation scale according to
$$\Delta\eta \sim \mathcal{H} = \frac{f\,U\,L}{g} = Ro\,\frac{f^2\,L^2}{g} = Ro\,H\,\frac{f^2\,L^2}{g\,H} = Ro\,H\left[\frac{L}{L_d}\right]^2. \tag{40.27}$$
In the final equality we introduced the deformation radius
$$L_d = \frac{\sqrt{g\,H}}{f}. \tag{40.28}$$
We encountered the deformation radius when discussing geostrophic adjustment in Section 31.5. Furthermore, note that the deformation radius is the scale whereby the relative vorticity and the surface height (vortex stretching) make equal contributions to the potential vorticity (see page 92

of *Pedlosky* (1987)). 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

$$Fr = \frac{U}{\sqrt{g\,H}} = \frac{U}{f\,L_d} = Ro\,\frac{L}{L_d}. \tag{40.29}$$

Furthermore, the squared ratio of the deformation radius to the lateral length scale is termed the Burger number

$$F^{-1} = Bu = \left[\frac{L_d}{L}\right]^2. \tag{40.30}$$

Use of the Burger number allows us to write the Froude number as

$$Fr = \frac{Ro}{\sqrt{Bu}} = Ro\,\sqrt{F} \tag{40.31}$$

and the free surface height undulation scale as

$$\mathcal{H} = H\,Ro\left[\frac{L}{L_d}\right]^2 = H\,\frac{Ro}{Bu} = H\,Ro\,F = H\,\frac{Fr^2}{Ro}. \tag{40.32}$$

Hence, the ratio of the free surface undulations to the depth scale is given by

$$\delta_{\text{free surface/depth}} = \frac{\mathcal{H}}{H} = Ro\left[\frac{L}{L_d}\right]^2 = \frac{Ro}{Bu} = Ro\,F = \frac{Fr^2}{Ro}. \tag{40.33}$$

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.

40.2.7 Non-dimensional shallow water equations

To non-dimensionalize the shallow water equations, introduce non-dimensional variables, denoted by a hat, according to

$$t = T\,\hat{t} \qquad (x,y) = L\,(\hat{x},\hat{y}) \qquad \partial_t = T^{-1}\partial_{\hat{t}} \qquad \nabla = L^{-1}\,\hat{\nabla} \tag{40.34a}$$

$$(u,v) = U\,(\hat{u},\hat{v}) \qquad f = f_o\,\hat{f} \qquad \Delta\eta = \mathcal{H}\,\hat{\eta} \qquad \eta = H + \Delta\eta = H + \mathcal{H}\,\hat{\eta}. \tag{40.34b}$$

where f_o is a typical scale for the Coriolis parameter. Importantly, we assume that the non-dimensional variables (the hat-variables) are order unity. Also note that the non-dimensional hatted variables should not be confused with the unit vector notation used elsewhere in this book.

Non-dimensional momentum equation

Introducing the above variables into the shallow water momentum equation (40.2a) renders

$$\frac{U}{T}\frac{\partial\hat{\boldsymbol{u}}}{\partial\hat{t}} + \frac{U^2}{L}\,(\hat{\boldsymbol{u}}\cdot\hat{\nabla})\,\hat{\boldsymbol{u}} + f_o\,U\,(\hat{\boldsymbol{f}}\wedge\hat{\boldsymbol{u}}) = -\frac{g\,\mathcal{H}}{L}\,\hat{\nabla}\hat{\eta}. \tag{40.35}$$

As before, we assume the time scale is given by the advection time

$$T = \frac{L}{U} = \frac{1}{R_o\,f_0}, \tag{40.36}$$

so that dividing by $f_0\,U$ leads to

$$Ro\left[\frac{\partial\hat{\boldsymbol{u}}}{\partial\hat{t}}+(\hat{\boldsymbol{u}}\cdot\hat{\nabla})\,\hat{\boldsymbol{u}}\right]+(\hat{\boldsymbol{f}}\wedge\hat{\boldsymbol{u}})=-\left[\frac{g\,H}{f_o\,L\,U}\frac{Ro}{Bu}\right]\hat{\nabla}\hat{\eta}, \tag{40.37}$$

where we set $\mathcal{H}=H\,(Ro/Bu)$ according to equation (40.33). We reduce the factor on the right hand side according to

$$\frac{g\,H}{f_o\,L\,U}\frac{Ro}{Bu}=\frac{g\,H}{f_o\,L\,U}\frac{U}{f_o\,L}\frac{L^2}{L_d^2}=\frac{g\,H}{f_o\,L\,U}\frac{U}{f_o\,L}\frac{L^2\,f_o^2}{g\,H}=1. \tag{40.38}$$

Hence, the non-dimensional inviscid shallow water momentum equation takes on the rather elegant form

$$Ro\left[\frac{\partial\hat{\boldsymbol{u}}}{\partial\hat{t}}+(\hat{\boldsymbol{u}}\cdot\hat{\nabla})\,\hat{\boldsymbol{u}}\right]+\hat{\boldsymbol{f}}\wedge\hat{\boldsymbol{u}}=-\hat{\nabla}\hat{\eta}. \tag{40.39}$$

Introducing the non-dimensional material time derivative

$$\frac{\mathrm{D}}{\mathrm{D}\hat{t}}=\frac{\partial}{\partial\hat{t}}+\hat{\boldsymbol{u}}\cdot\hat{\nabla} \tag{40.40}$$

brings the momentum equation to

$$Ro\,\frac{\mathrm{D}\hat{\boldsymbol{u}}}{\mathrm{D}\hat{t}}+\hat{\boldsymbol{f}}\wedge\hat{\boldsymbol{u}}=-\hat{\nabla}\hat{\eta}. \tag{40.41}$$

We see that 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 material acceleration.

Non-dimensional continuity (thickness) equation

The continuity equation (40.2b) can be written as

$$\frac{\partial\Delta\eta}{\partial t}+(H+\Delta\eta-\eta_{\mathrm{b}})\,\nabla\cdot\boldsymbol{u}+(\boldsymbol{u}\cdot\nabla)\,(H+\Delta\eta-\eta_{\mathrm{b}})=0, \tag{40.42}$$

where we wrote (see Figure 29.1)

$$h=H+\Delta\eta-\eta_{\mathrm{b}}. \tag{40.43}$$

Our assumed scaling for the bottom topography, (40.23), allows us to drop η_{b} from this equation. Also note that H is a constant and so it has a zero gradient. Hence, the continuity equation takes the form

$$\frac{\partial\Delta\eta}{\partial t}+(H+\Delta\eta)\,\nabla\cdot\boldsymbol{u}+(\boldsymbol{u}\cdot\nabla)\,\Delta\eta=0. \tag{40.44}$$

Introduction of the dimensionless variables leads to

$$H\,\frac{Ro}{Bu}\left[\frac{1}{T}\frac{\partial}{\partial\hat{t}}+\frac{U}{L}\hat{\boldsymbol{u}}\cdot\hat{\nabla}\right]\hat{\eta}+\frac{U\,H}{L}\left[1+\frac{Ro}{Bu}\,\hat{\eta}\right]\hat{\nabla}\cdot\hat{\boldsymbol{u}}=0. \tag{40.45}$$

With the time scale set by advection, $T=L/U$, we have

$$\frac{Ro}{Bu}\,\frac{\mathrm{D}\hat{\eta}}{\mathrm{D}\hat{t}}+\left[1+\frac{Ro}{Bu}\,\hat{\eta}\right]\hat{\nabla}\cdot\hat{\boldsymbol{u}}=0. \tag{40.46}$$

Equivalently, introducing the inverse Burger number, $F=Bu^{-1}$ renders

$$F\,Ro\,\frac{\mathrm{D}\hat{\eta}}{\mathrm{D}\hat{t}}+(1+F\,Ro\,\hat{\eta})\,\hat{\nabla}\cdot\hat{\boldsymbol{u}}=0. \tag{40.47}$$

40.3 Shallow water planetary geostrophy

We now make use of the non-dimensional equations derived in Section 40.2.7 to derive the dynamical equations for planetary geostrophy and quasi-geostrophy. For planetary geostrophy, we drop the fluid particle acceleration from the momentum equation (40.41), given that it is order Rossby number smaller than the Coriolis and pressure gradient accelerations. This assumption means that we are only interested in flows that satisfy the geostrophic balance

$$f\,\hat{\boldsymbol{z}} \wedge \hat{\boldsymbol{u}} = -\hat{\nabla}\hat{\eta}. \tag{40.48}$$

For the mass continuity equation (40.47), we must make an assumption about the ratio of the Rossby number to the Burger number. For planetary geostrophy, we assume

$$Ro \sim Bu. \tag{40.49}$$

Hence, the full mass continuity equation is retained; no terms are dropped. Since the Rossby number is small, $Ro \sim Bu = (L_d/L)^2 \ll 1$ means the horizontal length scale for the planetary geostrophic flows is much larger than the deformation radius

$$L \gg L_d. \tag{40.50}$$

The dimensional planetary geostrophic equations take the form

$$f\,\hat{\boldsymbol{z}} \wedge \boldsymbol{u} = -g\,\nabla\eta \tag{40.51a}$$

$$\frac{\mathrm{D}h}{\mathrm{D}t} = -h\nabla\cdot\boldsymbol{u} \tag{40.51b}$$

$$\eta = \eta_{\mathrm{b}} + h. \tag{40.51c}$$

As shown in Exercise 40.1, these equations are equivalent to

$$f\,\hat{\boldsymbol{z}} \wedge \boldsymbol{u} = -g\,\nabla\eta \qquad \text{and} \qquad \frac{\mathrm{D}Q}{\mathrm{D}t} = 0 \qquad \text{with} \qquad Q = \frac{f}{h}. \tag{40.52}$$

We encountered this form for the potential vorticity in Section 33.3, in which the shallow water PV, $(f+\zeta)/h$, is approximed by f/h in the small Rossby number limit.

40.4 Shallow water quasi-geostrophy

In this section we develop the quasi-geostrophic equations for a single shallow water fluid layer. We make use of asymptotic methods to derive these equations using the Rossby number as the small parameter.

40.4.1 Quasi-geostrophic scaling

Quasi-geostrophic scaling is based on the following assumptions.

1. $Ro \ll 1$, which is fundamental to geostrophic scaling.

2. $T \sim L/U$; that is, the time scale is determined by advection, which is how time has scaled throughout this chapter.

3. $F^{-1} = Bu \sim 1$, which means that the horizontal scales of motion for the quasi-geostrophic flows are on the order of the deformation radius, $L \sim L_d$. From equation (40.33), it furthermore means that undulations of the free surface height scale according to the Rossby number: $\mathcal{H} \sim H\,Ro$, meaning that the free surface height undulations are small. Finally, when developing the vorticity equation in Section 40.4.4, a unit Froude number means that the advection of vorticity contributes roughly the same as vortex stretching (see equation (40.68)).

4. $|\beta\, L| \ll |f_o|$, which means that the Coriolis frequency does not vary much from its central value.

The third and fourth assumptions are distinct from planetary geostrophy.

40.4.2 Outlining the asymptotic method

To derive the quasi-geostrophic shallow water model, we employ an asymptotic expansion in the Rossby number and stop at the first nontrivial order. For this purpose, recall the non-dimensional momentum and continuity equations from Section 40.2.7, and make use of the assumed $Bu \sim 1$ scaling

$$Ro\,\frac{\mathrm{D}\hat{\boldsymbol{u}}}{\mathrm{D}\hat{t}} + (\hat{\boldsymbol{f}} \wedge \hat{\boldsymbol{u}}) = -\hat{\nabla}\hat{\eta} \tag{40.53a}$$

$$F\,Ro\,\frac{\mathrm{D}\hat{\eta}}{\mathrm{D}\hat{t}} + (1 + F\,Ro\,\hat{\eta})\,\hat{\nabla}\cdot\hat{\boldsymbol{u}} = 0. \tag{40.53b}$$

We maintain the dimensionless parameter

$$F \equiv Bu^{-1} \tag{40.54}$$

in the continuity equation for later discussion. Again, for QG it is assumed to have unit scale and so will not play a role in the following asymptotic expansion.

Asymptotic methods are ideally suited for non-dimensional equations since we can unambiguously determine scales via the size of non-dimensional parameters. We here take the Rossby number to be small. It therefore makes sense to perform an asymptotic expansion of the prognostic fields in terms of the Rossby number. There are three prognostic fields, $\hat{u}, \hat{v}, \hat{\eta}$, in which we assume can be written

$$\hat{u} = \hat{u}_0 + Ro\,\hat{u}_1 + Ro^2\,\hat{u}_2 + \ldots \tag{40.55a}$$

$$\hat{v} = \hat{v}_0 + Ro\,\hat{v}_1 + Ro^2\,\hat{v}_2 + \ldots \tag{40.55b}$$

$$\hat{\eta} = \hat{\eta}_0 + Ro\,\hat{\eta}_1 + Ro^2\,\hat{\eta}_2 + \ldots \tag{40.55c}$$

In addition to expanding the prognostic variables, we expand the non-dimensional Coriolis parameter in terms of the Rossby number

$$\hat{\boldsymbol{f}} = \frac{\boldsymbol{f}}{f_0} = \frac{(f_0 + \beta\, y)\,\hat{\boldsymbol{z}}}{f_0} \equiv (\hat{f}_0 + Ro\,\hat{\beta}\,\hat{y})\,\hat{\boldsymbol{z}}. \tag{40.56}$$

In this equation, $\hat{\boldsymbol{z}}$ is the unit vector in the vertical, the time scaling in equation (40.36) renders

$$\hat{\beta}\,\hat{y} = \frac{\beta\, y}{R_o\, f_0} = T\,\beta\, y, \tag{40.57}$$

and

$$\hat{f}_0 = \frac{f_0}{f_0} = 1. \tag{40.58}$$

Although $\hat{f}_0 = 1$, it is useful to retain this term as a placeholder. The Coriolis expression (40.56), in particular the assumed scaling (40.57), is 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. The full spherical dependence of the Coriolis parameter has been reduced down to a mere constant plus a linear term (i.e., the β-plane approximation discussed in Section 24.2).

The practical goal of asymptotic analysis is to develop a closed set of prognostic equations for functions appearing in the asymptotic expansions (40.55a)-(40.56). 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 in something desired by the analyst.

40.4.3 Zeroth order asymptotic equations

At this point the setup has been done, the philosophy exposed, so we are ready to enter the "turn the crank" stage. To do so, we insert the asymptotic expansions (40.55a)-(40.55c) into the non-dimensional partial differential equations (40.53a) and (40.53b). There is a need to pay careful attention to detail here while organizing terms according to the Rossby number power. Since Ro is arbitrarily small, and all non-dimensional fields are order unity regardless their order, the only means to maintain self-consistency is that terms of equal order in Rossby number must balance. That is, we do not mix terms from different orders of Rossby number. This observation is basic to asymptotic methods.

Again, our overall 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

$$Ro\,\frac{\mathrm{D}\hat{\boldsymbol{u}}_0}{\mathrm{D}\hat{t}} + (\hat{f}_0 + Ro\,\hat{\beta}\,\hat{y})\,\hat{\boldsymbol{z}} \wedge (\hat{\boldsymbol{u}}_0 + Ro\,\hat{\boldsymbol{u}}_1) = -\hat{\nabla}(\hat{\eta}_0 + Ro\,\hat{\eta}_1) \tag{40.59a}$$

$$F\,Ro\,\frac{\mathrm{D}\hat{\eta}_0}{\mathrm{D}\hat{t}} + \hat{\nabla}\cdot\hat{\boldsymbol{u}}_0 + Ro\,\hat{\nabla}\cdot\hat{\boldsymbol{u}}_1 + F\,Ro\,\hat{\eta}_0\,\hat{\nabla}\cdot\hat{\boldsymbol{u}}_0 = 0. \tag{40.59b}$$

Terms balancing at order Ro^0 are given by

$$\hat{f}_0\,\hat{\boldsymbol{z}} \wedge \hat{\boldsymbol{u}}_0 = -\hat{\nabla}\hat{\eta}_0 \tag{40.60a}$$

$$\hat{\nabla}\cdot\hat{\boldsymbol{u}}_0 = 0, \tag{40.60b}$$

with the momentum balance reduced to the f-plane geostrophic balance. Fortunately, these two equations are self-consistent, since the curl of the f-plane geostrophic balance (40.60a) leads to the

non-divergence condition (40.60b). Given the non-divergence condition (40.60b), the zeroth order velocity field can be written in terms of a streamfunction

$$\hat{u}_0 = -\frac{\partial \hat{\psi}_0}{\partial \hat{y}} \qquad \hat{v}_0 = \frac{\partial \hat{\psi}_0}{\partial \hat{x}} \qquad \hat{\zeta}_0 = \hat{\nabla}^2 \hat{\psi}_0, \tag{40.61}$$

where the zeroth order streamfunction is the ratio of the zeroth order surface height to zeroth order Coriolis parameter

$$\hat{\psi}_0 = \frac{\hat{\eta}_0}{\hat{f}_0}, \tag{40.62}$$

and we introduced the zeroth order vorticity, $\hat{\zeta}_0$, which will appear in subsequent steps. The expression (40.62) for the streamfunction, with the division by $\hat{f}_0$, provides the motivation for retaining $\hat{f}_0$ in the equations since the dimensional form of the streamfunction will have the dimensional f_0 appearing (see equation (40.71)).

40.4.4 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

$$\frac{\mathrm{D}_0 \hat{\boldsymbol{u}}_0}{\mathrm{D}\hat{t}} + \hat{f}_0\, \hat{\boldsymbol{z}} \wedge \hat{\boldsymbol{u}}_1 + \hat{\beta}\, \hat{y}\, \hat{\boldsymbol{z}} \wedge \hat{\boldsymbol{u}}_0 = -\hat{\nabla} \hat{\eta}_1 \tag{40.63a}$$

$$F\, \frac{\mathrm{D}_0 \hat{\eta}_0}{\mathrm{D}\hat{t}} + \hat{\nabla} \cdot \hat{\boldsymbol{u}}_1 = 0. \tag{40.63b}$$

At this order, the material time derivative makes use of *only* the zeroth order geostrophic horizontal velocity

$$\frac{\mathrm{D}_0}{\mathrm{D}\hat{t}} = \frac{\partial}{\partial \hat{t}} + \hat{\boldsymbol{u}}_0 \cdot \hat{\nabla}. \tag{40.64}$$

The set of first order equations (40.63a) and (40.63b) are not closed, because evolution of zeroth order terms are functions of first order terms. However, the first order terms can be eliminated using two steps. First, we produce the vorticity equation from the momentum equation; second, we combine the vorticity equation and continuity equation. Although the details are specific to shallow water quasi-geostrophy, similar steps are frequently encountered in other geophysical fluid dynamical systems.

Taking the curl of the momentum equation (40.63a) eliminates the pressure gradient, $\hat{\nabla}\hat{\eta}_1$, thus producing the vorticity equation

$$\frac{\partial \hat{\zeta}_0}{\partial \hat{t}} + (\hat{\boldsymbol{u}}_0 \cdot \hat{\nabla})\, (\hat{\zeta}_0 + \hat{\beta}\, \hat{y}) = -\hat{f}_0\, \hat{\nabla} \cdot \hat{\boldsymbol{u}}_1. \tag{40.65}$$

To this order, vortex stretching on the right hand side arises just from the planetary vorticity, since relative vorticity stretching occurs at a higher order. Note that since $\hat{\beta}\, \hat{y}$ is time independent, we can write the vorticity equation in the material form

$$\frac{\mathrm{D}_0\, (\hat{\zeta}_0 + \hat{\beta}\, \hat{y})}{\mathrm{D}\hat{t}} = -\hat{f}_0\, \hat{\nabla} \cdot \hat{\boldsymbol{u}}_1. \tag{40.66}$$

As anticipated, we need one more step to close the system, since the evolution of zeroth order vorticity in equations (40.65) and (40.66) is a function of vortex stretching induced by convergence

of the first order velocity. We can substitute for the ageostrophic term $\hat{\nabla} \cdot \hat{\boldsymbol{u}}_1$ through use of the continuity equation (40.63b), thus leading to a prognostic equation involving just zeroth order terms

$$\frac{\partial(\hat{\zeta}_0 + \hat{\beta}\,\hat{y} - F\,\hat{f}_0\,\hat{\eta}_0)}{\partial \hat{t}} + (\hat{\boldsymbol{u}}_0 \cdot \hat{\nabla})\,(\hat{\zeta}_0 + \hat{\beta}\,\hat{y} - F\,\hat{f}_0\,\hat{\eta}_0) = 0, \tag{40.67}$$

which can be written in the material form

$$\frac{\mathrm{D}_0}{\mathrm{D}\hat{t}}\left[\hat{\zeta}_0 + \hat{\beta}\,\hat{y} - F\,\hat{f}_0\,\hat{\eta}_0\right] = 0. \tag{40.68}$$

Finally, we introduce the geostrophic streamfunction $\hat{\psi}_0 = \hat{\eta}_0/\hat{f}_0$ (equation (40.62)) to render

$$\frac{\mathrm{D}_0}{\mathrm{D}\hat{t}}\left[\hat{\nabla}^2\,\hat{\psi}_0 + \hat{\beta}\,\hat{y} - F\,\hat{f}_0^2\,\hat{\psi}_0\right] = 0. \tag{40.69}$$

Equations (40.68) and (40.69) are statements of the material conservation of quasi-geostrophic potential vorticity (QGPV), where material conservation is defined by the zeroth order horizontal geostrophic currents (equation (40.64)).

40.4.5 Dimensional equations and quasi-geostrophic PV

The material conservation equation (40.69) represents the culmination of our quest to realize a self-consistent closed prognostic equation via an asymptotic expansion to first order in Rossby number. We now gather the pieces, and in so doing transform the non-dimensional equations into their dimensional form. Since we are not interested in higher order terms, we drop all 0 subscripts, except for the Coriolis parameter. To proceed, invert the process started in Section 40.2.7, so that

$$\hat{t} = T^{-1}\,t \qquad (\hat{x}, \hat{y}) = L^{-1}\,(x, y) \qquad \partial_{\hat{t}} = T\,\partial_t \qquad \hat{\nabla} = L\,\nabla \tag{40.70a}$$

$$(\hat{u}, \hat{v}) = U^{-1}\,(u, v) \qquad \hat{f}_0 = \frac{f_0}{f_0} \qquad \hat{\beta}\,\hat{y} = \frac{\beta\,y}{Ro\,f_0} = \frac{L}{U}\,\beta\,y \tag{40.70b}$$

$$\hat{\eta} = \mathcal{H}^{-1}\,\Delta\eta \qquad \mathcal{H} = H\,F\,Ro \qquad \hat{\zeta} = \frac{L}{U}\,\zeta = L^2\,\nabla^2\psi, \tag{40.70c}$$

where the quasi-geostrophic streamfunction is

$$\psi = \frac{g\,\Delta\eta}{f_0}. \tag{40.71}$$

Note that *Vallis* (2017) defines the streamfunction as

$$\psi_{\text{vallis}} = \frac{g\,\eta}{f_0} = \frac{g\,(H + \Delta\eta)}{f_0}, \tag{40.72}$$

which differs by the constant $g\,H/f_0$. However, there is no difference in the dynamics since a streamfunction is defined only up to a constant. Also note that we take $F \sim 1$ for QG.

We now make use of these relations in the non-dimensional quasi-geostrophic potential vorticity

(QGPV) equation (40.68)

$$\hat{q} = \hat{\zeta}_0 + \hat{\beta}\,\hat{y} - \hat{f}_0\,\hat{\eta}_0 \tag{40.73a}$$
$$= \frac{L}{U}\,(\zeta + \beta\,y) - \frac{\Delta\eta}{\mathcal{H}} \tag{40.73b}$$
$$= \frac{L}{U}\,(\zeta + \beta\,y) - \frac{\Delta\eta}{H\,Ro} \tag{40.73c}$$
$$= \frac{L}{U}\left[\zeta + \beta\,y - \frac{f_0\,\Delta\eta}{H}\right] \tag{40.73d}$$
$$= \frac{L}{U}\left[\zeta + \beta\,y - \frac{g\,\Delta\eta}{f_0}\,\frac{1}{L_{\mathrm{d}}^2}\right] \tag{40.73e}$$
$$= \frac{L}{U}\left[\zeta + \beta\,y - L_{\mathrm{d}}^{-2}\,\psi\right]. \tag{40.73f}$$

Multiplying both sides by f_0 leads to the dimensionful QGPV for the shallow water fluid layer

$$q = Ro\,f_0\,\hat{q} = \zeta + \beta\,y - L_{\mathrm{d}}^{-2}\,\psi = \beta\,y + (\nabla^2 - L_{\mathrm{d}}^{-2})\,\psi. \tag{40.74}$$

40.5 Exercises

EXERCISE 40.1: PV CONSERVATION FOR PLANETARY GEOSTROPHY
Show that the planetary geostrophic equations

$$\hat{\boldsymbol{z}}\,f \wedge \boldsymbol{u} = -g\,\nabla\eta \qquad \text{and} \qquad \frac{\mathrm{D}h}{\mathrm{D}t} = -h\nabla\cdot\boldsymbol{u} \qquad \text{with} \qquad \eta = \eta_{\mathrm{b}} + h \tag{40.75}$$

are equivalent to

$$\hat{\boldsymbol{z}}\,f \wedge \boldsymbol{u} = -g\,\nabla\eta \qquad \text{and} \qquad \frac{\mathrm{D}Q}{\mathrm{D}t} = 0 \qquad \text{with} \qquad Q = \frac{f}{h}. \tag{40.76}$$

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 40.3. 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 29.1 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.

EXERCISE 40.3: LINEARIZED SHALLOW WATER PV

The potential vorticity for a shallow water layer is materially invariant when the flow is inviscid and adiabatic

$$\frac{\mathrm{D}Q}{\mathrm{D}t} = 0 \qquad \text{with} \qquad Q = \frac{\zeta + f}{h}. \tag{40.77}$$

Suppose that deviations of the free surface height, $\Delta\eta$, from its equilibrium position are small compared to the thickness, H, of the resting layer. Also assume the Rossby number is small so that $|\zeta| \ll |f|$. Consider flow on a β-plane so that $f = f_0 + \beta\, y$.

(a) Show that the evolution equation for potential vorticity can be approximated as

$$\frac{\mathrm{D}}{\mathrm{D}t}\left[\zeta + \beta\, y - \frac{f_0 \Delta\eta}{H}\right] = 0. \tag{40.78}$$

Hint: read Section 42.4.1.

(b) Using f-plane geostrophic balance, obtain an expression for ζ in terms of η.

EXERCISE 40.4: TOPOGRAPHIC CONSTRAINTS ON $f-$PLANE SHALLOW WATER FLOW

Throughout the quasi-geostrophic scaling for the shallow water in Section 40.2, we assumed the bottom topography has a tiny amplitude relative to undulations of the free surface height (see equation (40.23)). Return to the shallow water quasi-geostrophic scaling and now assume the topographic undulations are on the same order as resting fluid thickness, $\eta_{\mathrm{b}}/H \sim 1$. In this case we are formally moving outside of the QG regime. But we can still deduce interesting constraints on the leading order (geostrophic) flow.

(a) What constraint does topography impose on the geostrophic flow?

(b) How does the geostrophic streamfunction (and hence the free surface undulations $\Delta\eta$) relate to the bottom topography?

Hint: the answer is provided in Section 33.6.5. However, formulate the answer using notions presented in the current chapter.

41

Continuously stratified planetary geostrophy

Continuously stratified planetary geostrophy forms the foundation for many theories of the large-scale ocean circulation. The planetary geostrophic equations consist of linear steady frictional geostrophic balance 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 33.6. In this chaper 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.

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 34, and potential vorticity from Chapter 36. 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 **720**
41.1.1 Dimensional parameters 721
41.1.2 Physical dimensions and non-dimensional parameters 722
41.1.3 Choosing the non-dimensional parameters 722
41.1.4 Relating the buoyancy scale to the Coriols acceleration scale 723
41.1.5 Richardson number and QG/PG flow regimes 723
41.1.6 The Rossby deformation radius 724
41.1.7 Assumed values for the non-dimensional parameters 724
41.1.8 Non-dimensional Boussinesq equations 726
41.2 Summary equations for planetary geostrophy **727**
41.3 Planetary geostrophic potential vorticity **728**
41.3.1 Derivation 728
41.3.2 Impermeability theorem 730
41.3.3 Kinematic fluxes satisfying impermeability 730
41.4 Depth integrated vorticity budget **731**
41.4.1 The β-effect, stretching, and meridional transport 731
41.4.2 Bottom kinematics and dynamics 732
41.4.3 Surface kinematics and dynamics 734
41.4.4 Summary of force curls driving depth integrated meridional flow . . 735
41.5 Vorticity of the depth integrated velocity **737**
41.5.1 Depth integrated velocity equation 737
41.5.2 Vorticity budget 738
41.6 Vorticity equation for the depth averaged velocity **738**
41.6.1 Formulation 739
41.6.2 Rigid lid approximation and the role of JEBAR 739
41.6.3 Further study 741

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 25.2)

$$\frac{\mathrm{D}\boldsymbol{u}}{\mathrm{D}t} + f\,\hat{\boldsymbol{z}} \wedge \boldsymbol{u} = -\nabla_z \varphi \tag{41.1a}$$

$$\frac{\partial \varphi}{\partial z} = b \tag{41.1b}$$

$$\frac{\mathrm{D}b}{\mathrm{D}t} = 0 \tag{41.1c}$$

$$\nabla \cdot \boldsymbol{v} = 0, \tag{41.1d}$$

where $\boldsymbol{v} = (\boldsymbol{u}, w)$ 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 $\mathrm{L}^2\ \mathrm{T}^{-2}$), 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

$$b = \tilde{b}(z) + b'(x, y, z, t), \tag{41.2}$$

and introduce the corresponding background squared buoyancy frequency

$$N^2 = \frac{\partial \tilde{b}}{\partial z}. \tag{41.3}$$

With this decomposition, the buoyancy equation (41.1c) takes the form

$$\frac{\mathrm{D}b'}{\mathrm{D}t} + w\, N^2 = 0. \tag{41.4}$$

We also introduce an associated decomposition of the hydrostatic pressure

$$\varphi = \tilde{\varphi}(z) + \varphi'(x, y, z, t) \tag{41.5}$$

where $\tilde{\varphi}$ is hydrostatically balanced by $\tilde{b}$

$$\frac{\mathrm{d}\tilde{\varphi}}{\mathrm{d}z} = \tilde{b}, \tag{41.6}$$

and the fluctuating pressure, φ', is hydrostatically balanced by b'

$$\frac{\partial \varphi'}{\partial z} = b'. \tag{41.7}$$

41.1.1 Dimensional parameters

Following the shallow water discussion in Section 40.2.1, 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 × 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 40.2.1, 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-Π theorem then says there are

$$N_{\text{dimensionless}} = 9 - 2 = 7 \tag{41.8}$$

non-dimensional parameters.

41.1.3 Choosing the non-dimensional parameters

Following the shallow water discussion in Section 40.2.4, 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

$$\delta_{\text{verical/horizontal}} = \frac{\text{vertical depth scale}}{\text{horizontal length scale}} = \frac{H}{L}. \tag{41.9}$$

2. RATIO OF HORIZONTAL SCALE TO PLANETARY SCALE: The ratio of the lateral length scale to the planet radius is given by

$$\delta_{\text{horizontal/planet}} = \frac{\text{lateral length scale}}{\text{planetary length scale}} = \frac{L}{R}. \tag{41.10}$$

3. RATIO VERTICAL TO HORIZONTAL VELOCITY SCALE: The ratio of the vertical to horizontal velocity is given by

$$\frac{\text{vertical velocity scale}}{\text{horizontal velocity scale}} = \frac{W}{U}. \tag{41.11}$$

4. HYDROSTATIC NUMBER: The hydrostatic number is the ratio of the pressure gradient scale to the buoyancy scale. For the hydrostatic fluid fluctuations

$$\frac{\Phi}{H} = B, \tag{41.12}$$

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

$$Ro = \frac{\text{parcel acceleration}}{\text{Coriolis acceleration}} = \frac{U}{f\,L}, \tag{41.13}$$

where we again assume time scales advectively

$$T \sim \frac{L}{U} \Longrightarrow Ro = \frac{U}{f\,L} = \frac{1}{f\,T}. \tag{41.14}$$

6. GEOSTROPHIC NUMBER: The ratio of the Coriolis acceleration to the pressure gradient acceleration defines the geostrophic number

$$Ge = \frac{\text{Coriolis acceleration}}{\text{pressure gradient acceleration}}. \tag{41.15}$$

The Coriolis acceleration scales as

$$\text{Coriolis acceleration} \sim f\,U \tag{41.16}$$

whereas the pressure gradient acceleration from the fluctuating pressure, φ', scales as

$$\text{pressure gradient acceleration} \sim \frac{\Phi}{L}, \tag{41.17}$$

so that

$$Ge = \frac{\text{Coriolis acceleration}}{\text{pressure gradient acceleration}} = \frac{f\,U}{(\Phi/L)}. \tag{41.18}$$

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

$$\frac{\text{fluctuating buoyancy frequency}}{\text{background buoyancy frequency}} = \frac{B/H}{N^2}. \tag{41.19}$$

41.1.4 Relating the buoyancy scale to the Coriols acceleration scale

The fluctuating buoyancy (b') and fluctuating pressure (φ') have scales related through the hydrostatic balance. From equation (41.12) we have

$$B = \frac{\Phi}{H}. \tag{41.20}$$

Additionally, assuming geostrophic scaling, equation (41.18) means that the fluctuating pressure has a scale related to the Coriolis acceleration scale according to

$$\Phi = f\,U\,L. \tag{41.21}$$

Hence, the scale for the fluctuating buoyancy is given by

$$B = \frac{f\,U\,L}{H}. \tag{41.22}$$

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

$$Ri = \frac{N^2}{|\partial_z \mathbf{u}|^2}. \tag{41.23}$$

In regions where $Ri << 1$, 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-Helmholz instability*. In contrast, for large-scale highly stratified flow, the Richardson number is quite large, with $Ri \sim 100$ 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

$$Ri = \frac{N^2}{(U/H)^2}, \tag{41.24}$$

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

$$Bu = \left[\frac{L_{\mathrm{d}}}{L}\right]^2 = \left[\frac{N\,H}{f\,L}\right]^2 = \frac{U^2\,Ri}{U^2/(Ro)^2} = (Ro)^2\,Ri. \tag{41.25}$$

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

$$Ri \sim (Ro)^{-2} \qquad \text{QG flow.} \tag{41.26}$$

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 characeterized 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 continously stratified planetary geostrophic and quasi-geostrophic motions. Hence, the buoyancy frequency and the Coriolis parameter a 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 $T_f = 2\pi/f$ than buoyancy oscillations with period $T_b = 2\pi/N$.

Letting the squared buoyancy frequency N^2 refer to a value typical of a particular flow regime, one can define the Rossby deformation radius

$$L_{\mathrm{d}} = H\,\frac{N}{f}. \tag{41.27}$$

The ratio f/N appears frequently in rotating/stratified fluids, and is sometimes called the Prandtl ratio

$$\frac{f}{N} = \text{Prandtl ratio.} \tag{41.28}$$

With $H \approx 1$ km and $N/f \approx 100$, the Rossby radius is roughly 100 km. The Rossby radius is a crucial scale in geophysical fluids. For example, it sets the scale for unstable baroclinic waves leading to baroclinically unstable flow (see Chapter 6 of *Vallis* (2017)).

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 40.2.5. 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

$$\delta_{\text{vertical/horizontal}} \ll 1. \tag{41.29}$$

This assumption is part of the hydrostatic approximation (Section 24.3).

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

$$\delta_{\text{horizontal/planet}} \ll 1 \qquad \text{quasi-geostrophy} \tag{41.30a}$$

$$\delta_{\text{horizontal/planet}} \sim 1 \qquad \text{planetary geostrophy.} \tag{41.30b}$$

3. SMALL RATIO VERTICAL TO HORIZONTAL VELOCITY SCALE: The continuity equation implies

$$\frac{W}{H} = \frac{U}{L}, \tag{41.31}$$

so that

$$W = U\,\frac{H}{L}. \tag{41.32}$$

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))

$$\Phi = H\,B. \tag{41.33}$$

5. SMALL ROSSBY NUMBER: The Rossby number is assumed small

$$Ro = \frac{U}{f\,L} = \frac{1}{f\,T} \ll 1, \tag{41.34}$$

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

$$Ge \sim 1, \tag{41.35}$$

which means that the Coriolis acceleration and pressure gradient acceleration scale together

$$f\,U \sim \frac{\Phi}{L}. \tag{41.36}$$

This scaling is consistent with the momentum equation (41.1a) so long as the Rossby number is small, $Ro \ll 1$.

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

$$\frac{B/H}{N^2} = \frac{\Phi}{H^2\,N^2} = \frac{f\,U\,L}{H^2\,N^2} = \frac{U}{f\,L}\,\frac{L^2\,f^2}{H^2\,N^2} = Ro\,\frac{L^2}{L_{\mathrm{d}}^2}, \tag{41.37}$$

where we introduced the deformation radius for the continously stratified system

$$L_{\mathrm{d}} = H\,\frac{N}{f}. \tag{41.38}$$

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. Notably, since N^2 is a function of depth, we must keep this in mind when returning to dimensional fields, particularly for the QG system in Section 42.2.3.

41.1.8 Non-dimensional Boussinesq equations

Following the shallow water approach in Section 40.2.7, we introduce non-dimensional variables according to

$$t = T\,\hat{t} \qquad (x,y) = L\,(\hat{x},\hat{y}) \qquad \partial_t = T^{-1}\,\partial_{\hat{t}} \qquad \nabla_z = L^{-1}\hat{\nabla}_z \qquad \partial_z = H^{-1}\partial_{\hat{z}} \tag{41.39a}$$

$$(u,v) = U\,(\hat{u},\hat{v}) \qquad w = W\,\hat{w} \qquad f = f_0\,\hat{f} \qquad \varphi' = f_0\,U\,L\,\hat{\varphi} \qquad b' = B\,\hat{b} = \left[\frac{f_0\,U\,L}{H}\right]\hat{b}. \tag{41.39b}$$

For the second equality in the buoyancy scale, we made use of the relation (41.22) to connect the buoyancy fluctuation scale to the Coriolis acceleration scale. We also make use of the following relations between scales

$$T = \frac{L}{U} \qquad \text{advective scaling for time} \tag{41.40}$$

$$W = \frac{U\,H}{L} \qquad \text{continuity scaling for vertical velocity} \tag{41.41}$$

$$Ro = \frac{U}{f_0\,L} = \frac{1}{T\,f_0} \qquad \text{a result of advective time scaling.} \tag{41.42}$$

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 fine that W instead scales like $W = Ro\,U\,(H/L)$.

Non-dimensional momentum equation

Introducing the above dimensionless variables and dimensionful scales into the Boussinesq momentum equation (41.1a) renders

$$\frac{U}{T}\frac{\partial\hat{\boldsymbol{u}}}{\partial\hat{t}} + \frac{U^2}{L}\,(\hat{\boldsymbol{u}}\cdot\hat{\nabla}_z)\,\hat{\boldsymbol{u}} + \frac{W\,U}{H}\,\hat{\boldsymbol{w}}\,\frac{\partial\hat{\boldsymbol{u}}}{\partial\hat{z}} + f_0\,U\,(\hat{\boldsymbol{f}}\wedge\hat{\boldsymbol{u}}) = -f_0\,U\,\hat{\nabla}_z\,\hat{\varphi}. \tag{41.43}$$

Hence, dividing by $f_0\,U$ leads to

$$Ro\left[\frac{\partial\hat{\boldsymbol{u}}}{\partial\hat{t}}+(\hat{\boldsymbol{u}}\cdot\hat{\nabla}_{\hat{z}})\,\hat{\boldsymbol{u}}+\hat{w}\frac{\partial\hat{\boldsymbol{u}}}{\partial\hat{z}}\right]+(\hat{\boldsymbol{f}}\wedge\hat{\boldsymbol{u}})=-\hat{\nabla}_z\,\hat{\varphi}. \tag{41.44}$$

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

$$\frac{\partial\hat{\varphi}}{\partial\hat{z}}=\hat{b}, \tag{41.45}$$

and the non-dimensional continuity equation is

$$\hat{\nabla}\cdot\hat{\boldsymbol{v}}=0. \tag{41.46}$$

Non-dimensional buoyancy equation

The buoyancy equation (41.4) requires a bit more work to non-dimensionalize that for the momentum equation. The material time derivative takes the form

$$\frac{\mathrm{D}b'}{\mathrm{D}t}=\frac{B}{T}\,\frac{\mathrm{D}\hat{b}}{\mathrm{D}\hat{t}}=\frac{U}{L}\,\frac{f_0\,U\,L}{H}\,\frac{\mathrm{D}\hat{b}}{\mathrm{D}\hat{t}}=\frac{f_0\,U^2}{H}\,\frac{\mathrm{D}\hat{b}}{\mathrm{D}\hat{t}}, \tag{41.47}$$

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

$$N^2\,w=N^2\,W\,\hat{w}=N^2\,U\,(H/L)\,\hat{w}=L_{\mathrm{d}}^2\,\frac{U\,f_0^2}{H\,L}\,\hat{w}, \tag{41.48}$$

where we introduced the deformation scale, $L_{\mathrm{d}}=HN/f$, from equation (41.38). Bringing these two pieces together leads to

$$Ro\,F\,\frac{\mathrm{D}\hat{b}}{\mathrm{D}\hat{t}}+\hat{w}=0, \tag{41.49}$$

where we introduced the Burger number

$$Bu=F^{-1}=\left[\frac{L_{\mathrm{d}}}{L}\right]^2. \tag{41.50}$$

41.2 Summary equations for planetary geostrophy

Just like for the shallow water model in Section 40.3, 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

$$F\,Ro=Ro/Bu\sim 1, \tag{41.51}$$

or

$$L^2\sim L_{\mathrm{d}}^2\,Ro^{-1}. \tag{41.52}$$

With this scaling, and with the Rossby number small, the momentum equation (41.44) 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

$$\frac{\mathrm{D}b'}{\mathrm{D}t} + w\,N^2 = 0 \tag{41.53a}$$

$$f\,\hat{\boldsymbol{z}} \wedge \boldsymbol{u} = -\nabla_z \varphi' \tag{41.53b}$$

$$\frac{\partial \varphi'}{\partial z} = b' \tag{41.53c}$$

$$\nabla \cdot \boldsymbol{v} = 0. \tag{41.53d}$$

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, $\boldsymbol{F}$. The following form for the frictional and diadiabatic planetary geostrophic flow will form the basis for many discussions in this chapter

$$f\,\hat{\boldsymbol{z}} \wedge \boldsymbol{u} = -\rho_0^{-1}\,\nabla_z p + \boldsymbol{F} \qquad \text{frictional geostrophy} \tag{41.54a}$$

$$\frac{\partial b}{\partial t} + \boldsymbol{u} \cdot \nabla_z b + N^2\,w = \dot{b} \qquad \text{diabatic buoyancy equation} \tag{41.54b}$$

$$\frac{\partial p}{\partial z} = -\rho\,g \qquad \text{hydrostatic} \tag{41.54c}$$

$$\nabla \cdot \boldsymbol{v} = \nabla_z \cdot \boldsymbol{u} + \frac{\partial w}{\partial z} = 0 \qquad \text{incompressible} \tag{41.54d}$$

$$N^2 = \frac{\partial b}{\partial z} \qquad \text{squared buoyancy frequency.} \tag{41.54e}$$

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

$$\dot{b} = \frac{\partial b}{\partial S}\,\dot{S} + \frac{\partial b}{\partial \Theta}\,\dot{\Theta}. \tag{41.55}$$

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, $\boldsymbol{v} = (\boldsymbol{u}, w)$, as seen in the buoyancy equation (41.54b), with the horizontal velocity components determined by the frictional geostrophic balance (41.53b). 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 36.5 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 (41.54a)-(41.54e).

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 (41.54a). The vertical component of this vorticity equation is given by (see also Section 26.4.2)

$$\beta\, v = f\,\frac{\partial w}{\partial z} + \hat{\boldsymbol{z}}\cdot(\nabla_z \wedge \boldsymbol{F}) \qquad \text{with} \qquad \beta = \frac{\partial f}{\partial y}. \tag{41.56}$$

Next make use of the thermal wind relation in the presence of friction

$$f\,\frac{\partial \boldsymbol{u}}{\partial z} = \hat{\boldsymbol{z}} \wedge \nabla_z b - \frac{\partial(\hat{\boldsymbol{z}} \wedge \boldsymbol{F})}{\partial z} \tag{41.57}$$

as well as the identities

$$N^2\,\frac{\mathrm{D}f}{\mathrm{D}t} = N^2\,\beta\, v \tag{41.58a}$$

$$f\,\frac{\mathrm{D}N^2}{\mathrm{D}t} = f\,\frac{\partial \dot{b}}{\partial z} - f\,\nabla b\cdot\frac{\partial \boldsymbol{v}}{\partial z} \tag{41.58b}$$

$$f\,\nabla b\cdot\frac{\partial \boldsymbol{v}}{\partial z} = f\,N^2\,\frac{\partial w}{\partial z} - \frac{\partial(\hat{\boldsymbol{z}} \wedge \boldsymbol{F})}{\partial z}\cdot\nabla_z b, \tag{41.58c}$$

to render

$$\frac{\mathrm{D}(f\,N^2)}{\mathrm{D}t} = N^2\,\beta\, v + f\,\frac{\partial \dot{b}}{\partial z} - f\,\nabla b\cdot\frac{\partial \boldsymbol{v}}{\partial z} \tag{41.59a}$$

$$= N^2\left[f\,\frac{\partial w}{\partial z} + \hat{\boldsymbol{z}}\cdot(\nabla_z \wedge \boldsymbol{F})\right] + f\,\frac{\partial \dot{b}}{\partial z} - f\,N^2\,\frac{\partial w}{\partial z} + \frac{\partial(\hat{\boldsymbol{z}} \wedge \boldsymbol{F})}{\partial z}\cdot\nabla_z b \tag{41.59b}$$

$$= f\,\frac{\partial \dot{b}}{\partial z} + \nabla b\cdot(\nabla \wedge \boldsymbol{F}) \tag{41.59c}$$

$$= \nabla\cdot(f\,\dot{b}\,\hat{\boldsymbol{z}} + \boldsymbol{F} \wedge \nabla b). \tag{41.59d}$$

We thus identify the planetary geostrophic potential vorticity

$$Q_{\mathrm{PG}} = f\,N^2, \tag{41.60}$$

which is materially invariant in the absence of diabatic processes and friction

$$\frac{\mathrm{D}Q_{\mathrm{PG}}}{\mathrm{D}t} = 0 \qquad \text{when } \dot{b} = 0 \text{ and } \boldsymbol{F} = 0. \tag{41.61}$$

We can write the general budget equation in the form of an Eulerian flux-form expression

$$\frac{\partial Q_{\mathrm{PG}}}{\partial t} + \nabla\cdot\boldsymbol{J}_{\mathrm{PG}} = 0, \tag{41.62}$$

where the potential vorticity flux is given by

$$\boldsymbol{J}_{\mathrm{PG}} = \boldsymbol{v}\,Q_{\mathrm{PG}} - \dot{b}\,f\,\hat{\boldsymbol{z}} + \nabla b \wedge \boldsymbol{F} + \nabla \wedge \boldsymbol{A}. \tag{41.63}$$

The vector $\boldsymbol{A}$ is an arbitrary gauge function that has no impact on the potential vorticity evolution. Comparing to the hydrostatic Boussinesq case from Section 36.5.2, 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 37.1.2, we here verify that the potential vorticity flux vector (41.63) satisfies the impermeability theorem for buoyancy isosurfaces. We do so here for the particular case of a zero gauge function ($\boldsymbol{A} = 0$), in which case

$$\boldsymbol{v}_{\rm PG} \cdot \nabla b = \frac{\boldsymbol{J}_{\rm PG} \cdot \nabla b}{Q} = \boldsymbol{v} \cdot \nabla b - \dot{b} = -\frac{\partial b}{\partial t}, \tag{41.64}$$

reveals that

$$\frac{\partial b}{\partial t} + \boldsymbol{v}_{\rm PG} \cdot \nabla b = 0. \tag{41.65}$$

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 37.1.2, 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

$$\frac{\partial Q}{\partial t} = \frac{\partial}{\partial t} \nabla \cdot (f\, b\, \hat{\boldsymbol{z}}) = \nabla \cdot \left[f\, \frac{\partial b}{\partial t}\, \hat{\boldsymbol{z}} \right] \equiv -\nabla \cdot \tilde{\boldsymbol{J}}_{\rm PG}, \tag{41.66}$$

where

$$\tilde{\boldsymbol{J}}_{\rm PG} = -f\, \frac{\partial b}{\partial t}\, \hat{\boldsymbol{z}}. \tag{41.67}$$

This form of the PV-substance flux also satisfies impermeability since

$$\tilde{\boldsymbol{v}}_{\rm PG} \cdot \nabla b = \frac{\tilde{\boldsymbol{J}}_{\rm PG} \cdot \nabla b}{Q} = -\frac{\partial b}{\partial t}, \tag{41.68}$$

so that

$$\frac{\partial b}{\partial t} + \tilde{\boldsymbol{v}}_{\rm PG} \cdot \nabla b = 0. \tag{41.69}$$

The flux $\tilde{\boldsymbol{J}}_{\rm PG}$ vanishes in the steady state, whereas the steady state form of $\boldsymbol{J}_{\rm PG}$ does not vanish. Following the discussion in Section 37.2.2, we may choose to introduce a gauge transformation to the kinematic flux $\tilde{\boldsymbol{J}}_{\rm PG}$ so that it does not vanish in the steady state. Taking the small Rossby number limit of the flux (37.35) renders

$$\boldsymbol{J}_{\rm Q}^{\rm marshall\ PG} = -\nabla(g\, z + \varphi) \wedge \nabla b - f\, \frac{\partial b}{\partial t}\, \hat{\boldsymbol{z}}. \tag{41.70}$$

This flux differs from $\tilde{\boldsymbol{J}}_{\rm PG}$ by a curl

$$\nabla(g\, z + \varphi) \wedge \nabla b = \nabla \wedge [(g\, z + \varphi)\, \nabla b], \tag{41.71}$$

and it also satisfies the impermeability theorem. As discussed in Section 37.4.6, 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 $\boldsymbol{J}_{\rm Q}^{\rm marshall\ PG}$ 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 34.3.4, 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 is physics refers to a torque as affecting changes to angular momentum, with angular momentum generally distinct from vorticity (see Chapter 35 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

$$\boldsymbol{F} = \frac{\partial \boldsymbol{\tau}}{\partial z}. \tag{41.72}$$

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 26.4.2 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 (41.72)), the vertical component of the PG vorticity equation takes the form

$$\rho_0 \, \beta \, v = \frac{\partial}{\partial z} \left[\rho_0 \, f \, w + \hat{\boldsymbol{z}} \cdot (\nabla_z \wedge \boldsymbol{\tau}) \right]. \tag{41.73}$$

Vertical integration from the ocean bottom at $z = \eta_{\mathrm{b}}(x, y)$ to sea surface at $z = \eta(x, y, t)$ leads to

$$\rho_0 \, \beta \, V = \underbrace{\rho_0 \, f \, [w(\eta) - w(\eta_{\mathrm{b}})]}_{\text{column stretching}} + \underbrace{\hat{\boldsymbol{z}} \cdot (\nabla_z \wedge \Delta \boldsymbol{\tau})}_{\text{boundary stresses}}, \tag{41.74}$$

where

$$\Delta \boldsymbol{\tau} = \boldsymbol{\tau}(\eta) - \boldsymbol{\tau}(\eta_{\mathrm{b}}) \tag{41.75}$$

is the difference in boundary stresses applied at the ocean surface and ocean bottom, and

$$V = \int_{\eta_{\mathrm{b}}}^{\eta} v \, \mathrm{d}z \tag{41.76}$$

is the depth-integrated meridional flow. For a planetary geostrophic flow, absolute vorticity is approximated by just the planetary vorticity: $\zeta_{\mathrm{a}} = \zeta + f \approx f$. As revealed by the vorticity equation (41.74), 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 (41.74) 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, $w(\eta_{\mathrm{b}}) < 0$, 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 (41.74) 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, $\hat{\boldsymbol{z}} \cdot [\nabla_z \wedge \boldsymbol{\tau}(\eta)] > 0$, or by the curl of bottom stresses, $\hat{\boldsymbol{z}} \cdot [\nabla_z \wedge \boldsymbol{\tau}(\eta_{\mathrm{b}})] < 0$, lead to poleward motion of the fluid column, with opposite motion for the converse curls.

The β-effect (Section 34.6.2) is a fundamental element of the depth integrated vorticity equation (41.74). 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 (41.74) is fundamentally related to vortex stretching.

Equation (41.74) 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 $w(\eta) = w(\eta_{\mathrm{b}}) = 0$. 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 $z = \eta_{\mathrm{b}}(x, y)$ (Section 14.4.1) is given by

$$w = \boldsymbol{u} \cdot \nabla_z \eta_{\mathrm{b}} \qquad \text{at } z = \eta_{\mathrm{b}}(x, y). \tag{41.77}$$

This relation expresses the no-normal flow condition $\hat{\boldsymbol{n}} \cdot \boldsymbol{v} = 0$ at the ocean bottom, with $\hat{\boldsymbol{n}} = -\nabla(z - \eta_{\mathrm{b}})/|\nabla(z - \eta_{\mathrm{b}})|$ 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 (41.74), 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 (41.54a). Evaluating the horizontal components of this equation at the ocean bottom yields

$$\rho_0 \, f \, \hat{\boldsymbol{z}} \wedge \boldsymbol{u} = -\nabla_z p_{\mathrm{b}} + \boldsymbol{F}_{\mathrm{b}}, \tag{41.78}$$

where $p_{\rm b}(x, y, t)$ is the bottom pressure and $\boldsymbol{F}_{\rm b}$ is the bottom friction. For the special case of a no-slip bottom, all velocity components vanish at $z = \eta_{\rm b}$. In that case, we consider $\boldsymbol{u}$ in equation (41.78) 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 Ekman[1] components via

$$\rho_0 \, f \, \hat{\boldsymbol{z}} \wedge \boldsymbol{u}_{\rm g} = -\nabla_z p_{\rm b} \qquad \text{and} \qquad \rho_0 \, f \, \hat{\boldsymbol{z}} \wedge \boldsymbol{u}_{\rm e} = \boldsymbol{F}_{\rm b}, \tag{41.79}$$

so that

$$\boldsymbol{u}_{\rm g} = \frac{\hat{\boldsymbol{z}} \wedge \nabla_z p_{\rm b}}{\rho_0 \, f} \qquad \text{and} \qquad \boldsymbol{u}_{\rm e} = -\frac{\hat{\boldsymbol{z}} \wedge \boldsymbol{F}_{\rm b}}{\rho_0 \, f}. \tag{41.80}$$

The corresponding bottom vertical velocity components are determined by inserting equations (41.80) into the bottom kinematic boundary condition (41.77)

$$w_{\rm g} = \frac{\hat{\boldsymbol{z}} \cdot (-\nabla_z \eta_{\rm b} \wedge \nabla_z p_{\rm b})}{\rho_0 \, f} \qquad \text{and} \qquad w_{\rm e} = \frac{\hat{\boldsymbol{z}} \cdot (\nabla_z \eta_{\rm b} \wedge \boldsymbol{F}_b)}{\rho_0 \, f}. \tag{41.81}$$

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 (41.77). As seen by these equations, is only the projection of $\nabla_z p_{\rm b}$ and $\boldsymbol{F}_b$ 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

$$w_{\rm g} = \frac{\hat{\boldsymbol{z}} \cdot (-\nabla_z \eta_{\rm b} \wedge \nabla_z p_{\rm b})}{\rho_0 \, f} = \frac{\hat{\boldsymbol{z}} \cdot [\nabla_z \wedge p_{\rm b} \nabla_z \eta_{\rm b}]}{\rho_0 \, f}. \tag{41.82}$$

The numerator is the curl of the horizontal projection of the pressure contact force along the bottom, $p_{\rm b} \nabla_z \eta_{\rm b}$. This term is the topographic form stress discussed in Section 20.2 for a general fluid and in Section 33.6.5 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 (41.80) arises from changes in bottom depth. However, that portion of the bottom pressure gradient has no impact on $w_{\rm g}$, since it only drives horizontal flow along isobaths. We see this property by writing

$$p_{\rm b} = -\rho_0 \, g \, \eta_{\rm b} + p'_{\rm b} \Longrightarrow w_{\rm g} = \frac{\hat{\boldsymbol{z}} \cdot (-\nabla_z \eta_{\rm b} \wedge \nabla_z p'_{\rm b})}{\rho_0 \, f}. \tag{41.83}$$

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 34.4).

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

[1] Recall our discussion of Ekman mechanics in Chapter 28.

Figure 41.1: *Depicting how bottom pressure gradients create vertical motion in planetary geostrophic flow next to a sloping bottom according to equations (41.81) and (41.83). 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 $\boldsymbol{F}_b$ that are aligned parallel to the topographic slope contribute to vertical motion. We illustrate here a case where the bottom pressure gradient leads to $-\nabla\eta_b \cdot \boldsymbol{u}_g > 0$ so that $w_g < 0$ in the northern hemisphere and $w_g > 0$ in the southern hemisphere (see equations (41.81) and (41.83)).*

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, $\boldsymbol{F}_b$, 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 (41.74) and rewriting it as an expression for vertical motion

$$\rho_0 f [w(\eta_b) - w(\eta)] = -\rho_0 \beta V + \hat{\boldsymbol{z}} \cdot (\nabla_z \wedge \Delta\boldsymbol{\tau}). \tag{41.84}$$

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 16.2) is given by

$$w = -\frac{Q_{\mathrm{m}}}{\rho_0} + \frac{\partial \eta}{\partial t} + \boldsymbol{u} \cdot \nabla_z \eta \qquad \text{at } z = \eta(x, y, t). \tag{41.85}$$

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 (26.24a) at the ocean surface renders

$$\rho_0 \, f \, \hat{\boldsymbol{z}} \wedge \boldsymbol{u} = -\nabla_z p_{\mathrm{a}} + \boldsymbol{F}_\eta, \tag{41.86}$$

where p_{a} is the pressure applied to the ocean surface and $\boldsymbol{F}_\eta$ is the corresponding horizontal friction vector. Like the bottom, we decompose the horizontal velocity into a geostrophic component and an Ekman component via

$$\rho_0 \, f \, \hat{\boldsymbol{z}} \wedge \boldsymbol{u}_{\mathrm{g}} = -\nabla_z p_{\mathrm{a}} \qquad \text{and} \qquad \rho_0 \, f \, \hat{\boldsymbol{z}} \wedge \boldsymbol{u}_{\mathrm{e}} = \boldsymbol{F}_\eta, \tag{41.87}$$

so that

$$\boldsymbol{u}_{\mathrm{g}} = \frac{\hat{\boldsymbol{z}} \wedge \nabla_z p_{\mathrm{a}}}{\rho_0 \, f} \qquad \text{and} \qquad \boldsymbol{u}_{\mathrm{e}} = -\frac{\hat{\boldsymbol{z}} \wedge \boldsymbol{F}_\eta}{\rho_0 \, f}. \tag{41.88}$$

The corresponding vertical velocity components are determined by inserting into the surface kinematic boundary condition (41.85)

$$w_{\mathrm{Q}\eta} = -\frac{Q_{\mathrm{m}}}{\rho_0} + \frac{\partial \eta}{\partial t} \qquad \text{and} \qquad w_{\mathrm{g}} = \frac{\hat{\boldsymbol{z}} \cdot (-\nabla_z \eta \wedge \nabla_z p_{\mathrm{a}})}{\rho_0 \, f} \qquad \text{and} \qquad w_{\mathrm{e}} = \frac{\hat{\boldsymbol{z}} \cdot (\nabla_z \eta \wedge \boldsymbol{F}_\eta)}{\rho_0 \, f}, \tag{41.89}$$

where we introduced a vertical velocity, $w_{\mathrm{Q}\eta}$, 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 (41.85). Furthermore, it is only the projection of $\nabla_z p_{\mathrm{a}}$ and $\boldsymbol{F}_\eta$ 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 (41.81), (41.83), and (41.89) into equation (41.74) renders the depth integrated planetary vorticity balance

$$\rho_0 \, \beta \, V = f \left(-Q_{\mathrm{m}} + \rho_0 \, \partial_t \eta \right) + \hat{\boldsymbol{z}} \cdot \left[\nabla_z \eta \wedge (\boldsymbol{F}_\eta - \nabla_z p_{\mathrm{a}}) - \nabla_z \eta_{\mathrm{b}} \wedge (\boldsymbol{F}_{\mathrm{b}} - \nabla_z p'_{\mathrm{b}}) + \nabla_z \wedge \Delta \boldsymbol{\tau} \right]. \tag{41.90}$$

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 (41.90) shows that the depth integrated meridional flow, within the planetary geostrophic regime, is driven by the following processes.

[2] See Section 3.3 of *Samelson* (2011) for more details on this scaling of the planetary geostrophic system.

Surface mass transport plus sea surface fluctuations

The term $f\,(-Q_{\rm m} + \rho_0\,\partial_t \eta)$ 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 Goldsbrough-Stommel circulation (see *Huang and Schmitt* (1993) for a review).

Curl of turbulent boundary stresses

The term $\hat{\boldsymbol{z}} \cdot (\nabla_z \wedge \Delta \boldsymbol{\tau})$ 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

$$\nabla_z p_{\rm a} \wedge \nabla_z \eta = \nabla_z \wedge (p_{\rm a} \nabla_z \eta) = -\nabla_z \wedge (\eta \nabla_z p_{\rm a}) \tag{41.91}$$

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 34.4, with these misalignments driving vertical motion and a corresponding depth integrated meridional flow.

Bottom pressure torque

The term

$$\nabla_z p'_{\rm b} \wedge (-\nabla_z \eta_{\rm b}) = \nabla_z \wedge (-p'_{\rm b} \nabla_z \eta_{\rm b}) = \nabla_z \wedge (\eta_{\rm b} \nabla_z p'_{\rm b}) \tag{41.92}$$

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 33.6.5.

Surface frictional acceleration

The term

$$\hat{\boldsymbol{z}} \cdot (\nabla_z \eta \wedge \boldsymbol{F}_\eta) = \hat{\boldsymbol{z}} \cdot \left[\nabla_z \eta \wedge \frac{\partial \boldsymbol{\tau}_\eta}{\partial z} \right] \tag{41.93}$$

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

$$\hat{\boldsymbol{z}} \cdot \int_{-h_{\text{e-surf}}}^{\eta} \left[\nabla_z \eta \wedge \frac{\partial \boldsymbol{\tau}}{\partial z} \right] \mathrm{d}z = \hat{\boldsymbol{z}} \cdot [\nabla_z \eta \wedge \boldsymbol{\tau}(\eta)], \tag{41.94}$$

where we assumed $\boldsymbol{\tau}$ is negligible at the base of the surface Ekman layer, $z = -h_{\text{e-surf}}(x, y, t)$. The term $\hat{\boldsymbol{z}} \cdot [\nabla_z \eta \wedge \boldsymbol{\tau}(\eta)]$ 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

$$\hat{\boldsymbol{z}} \cdot (-\nabla_z \eta_{\mathrm{b}} \wedge \boldsymbol{F}_{\mathrm{b}}) = \hat{\boldsymbol{z}} \cdot \left[-\nabla_z \eta_{\mathrm{b}} \wedge \frac{\partial \boldsymbol{\tau}_{\mathrm{b}}}{\partial z} \right] \tag{41.95}$$

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

$$\hat{\boldsymbol{z}} \cdot \int_{\eta_{\mathrm{b}}}^{h_{\text{e-bot}}} \left[-\nabla_z \eta_{\mathrm{b}} \wedge \frac{\partial \boldsymbol{\tau}}{\partial z} \right] \mathrm{d}z = \hat{\boldsymbol{z}} \cdot [-\nabla_z \eta_{\mathrm{b}} \wedge \boldsymbol{\tau}(\eta_{\mathrm{b}})], \tag{41.96}$$

where we assumed $\boldsymbol{\tau}$ is negligible at the top of the bottom Ekman layer, $z = -\eta_{\mathrm{b}} + h_{\text{e-bot}}$. The term $\hat{\boldsymbol{z}} \cdot [-\nabla_z \eta_{\mathrm{b}} \wedge \boldsymbol{\tau}(\eta_{\mathrm{b}})]$ 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 (41.54a) is given by

$$\rho_0 \, f \, \hat{\boldsymbol{z}} \wedge \boldsymbol{U} = -\int_{\eta_{\mathrm{b}}}^{\eta} \nabla_z p \, \mathrm{d}z + \Delta \boldsymbol{\tau} \tag{41.97}$$

where

$$\boldsymbol{U} = \int_{\eta_{\mathrm{b}}}^{\eta} \boldsymbol{u} \, \mathrm{d}z \tag{41.98}$$

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 (41.72). For the depth integrated pressure gradient, we introduce the decomposition

$$\int_{\eta_{\mathrm{b}}}^{\eta} p \, \mathrm{d}z = \int_{\eta_{\mathrm{b}}}^{\eta} [\mathrm{d}(p \, z) - z \, \mathrm{d}p] = p_{\mathrm{a}} \, \eta - p_{\mathrm{b}} \, \eta_{\mathrm{b}} + \mathscr{P}, \tag{41.99}$$

where we used the hydrostatic balance to write $\mathrm{d}p = -g \, \rho \, \mathrm{d}z$. We also introduced the potential energy per horizontal area of a fluid column

$$\mathscr{P} = \int_{\eta_{\mathrm{b}}}^{\eta} g \, \rho \, z \, \mathrm{d}z = (g \, \overline{\rho}/2) \, (\eta^2 - \eta_{\mathrm{b}}^2) + \int_{\eta_{\mathrm{b}}}^{\eta} g \, \rho' \, z \, \mathrm{d}z, \tag{41.100}$$

and defined the area averaged density, $\overline{\rho}$, and its anomaly, $\rho' = \rho - \overline{\rho}$

$$\overline{\rho} = \frac{\int \rho \, \mathrm{d}A}{\int \mathrm{d}A} \qquad \text{and} \qquad \rho = \overline{\rho} + \rho'. \tag{41.101}$$

These results then lead to the depth integrated horizontal pressure gradient

$$\int_{\eta_b}^{\eta} \nabla_z p \, dz = \nabla_z \left[\int_{\eta_b}^{\eta} p \, dz \right] - p_a \nabla_z \eta + p_b \nabla_z \eta_b \tag{41.102a}$$

$$= \nabla_z \left[p_a \, \eta - p_b \, \eta_b + \mathscr{P} \right] - p_a \nabla_z \eta + p_b \nabla_z \eta_b \tag{41.102b}$$

$$= \eta \nabla_z p_a - \eta_b \nabla_z p_b + \nabla_z \mathscr{P}, \tag{41.102c}$$

which in turn renders the depth integrated planetary geostrophic momentum balance

$$\rho_0 \, f \, \hat{\boldsymbol{z}} \wedge \boldsymbol{U} = -\eta \, \nabla_z p_a + \eta_b \nabla_z p_b - \nabla_z \mathscr{P} + \Delta \boldsymbol{\tau}. \tag{41.103}$$

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, $\Delta \boldsymbol{\tau} = \boldsymbol{\tau}(\eta) - \boldsymbol{\tau}(\eta_b)$.

41.5.2 Vorticity budget

Taking the curl of the depth integrated balance (41.103) annihilates the potential energy term, thus leaving

$$\rho_0 \, \beta \, V = -\rho_0 \, f \, \nabla \cdot \boldsymbol{U} + \nabla \wedge [p_a \nabla \eta + \boldsymbol{\tau}(\eta) - p_b \nabla \eta_b - \boldsymbol{\tau}(\eta_b)]. \tag{41.104}$$

From Section 16.3, we know that the divergence of the depth-integrated flow for a steady Boussinesq fluid is given by

$$\rho_0 \nabla \cdot \boldsymbol{U} = Q_m, \tag{41.105}$$

so that

$$\rho_0 \, \beta \, V = -f \, Q_m + \nabla \wedge [p_a \nabla \eta + \boldsymbol{\tau}(\eta) - p_b \nabla \eta_b - \boldsymbol{\tau}(\eta_b)]. \tag{41.106}$$

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 (33.74) (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

$$\overline{\boldsymbol{u}} = \frac{1}{D} \int_{\eta_b}^{\eta} \boldsymbol{u} \, dz \qquad \text{with} \qquad D = \eta - \eta_b. \tag{41.107}$$

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 Formulation

We start by deriving the momentum equation for the depth averaged flow. For that purpose, rearrange the depth integrated momentum budget (41.103) according to

$$\rho_0 \, f \, \hat{\boldsymbol{z}} \wedge \boldsymbol{U} = -\eta \, \nabla_z (p_{\mathrm{a}} - p_{\mathrm{b}}) - (\eta - \eta_{\mathrm{b}}) \nabla_z p_{\mathrm{b}} - \nabla_z \mathscr{P} + \Delta \boldsymbol{\tau} \tag{41.108}$$

and then divide by the thickness of the ocean column to render

$$\rho_0 \, f \, \hat{\boldsymbol{z}} \wedge \overline{\boldsymbol{u}} = -(\eta/D) \, \nabla_z (p_{\mathrm{a}} - p_{\mathrm{b}}) - \nabla_z p_{\mathrm{b}} + D^{-1} \, (-\nabla_z \mathscr{P} + \Delta \boldsymbol{\tau}). \tag{41.109}$$

Taking the curl then leads to

$$\rho_0 \, \nabla_z \cdot (f \, \overline{\boldsymbol{u}}) = -\hat{\boldsymbol{z}} \cdot \nabla_z \wedge [(\eta/D) \, \nabla_z (p_{\mathrm{a}} - p_{\mathrm{b}})] + D^{-2} \, \hat{\boldsymbol{z}} \cdot [\nabla_z \wedge (D \, \nabla_z \mathscr{P})] + \hat{\boldsymbol{z}} \cdot [\nabla_z \wedge (D^{-1} \, \Delta \boldsymbol{\tau})] \tag{41.110a}$$

$$= \hat{\boldsymbol{z}} \cdot \nabla_z \wedge [(p_{\mathrm{a}} - p_{\mathrm{b}}) \, \nabla_z (\eta/D)] - D^{-2} \, \hat{\boldsymbol{z}} \cdot [\nabla_z \wedge (\mathscr{P} \, \nabla_z D)] + \hat{\boldsymbol{z}} \cdot [\nabla_z \wedge (D^{-1} \, \Delta \boldsymbol{\tau})]. \tag{41.110b}$$

The vorticity budget (41.110) is a bit less tidy than that for the depth integrated budget (41.106). Nonetheless, it offers some useful information as seen by the following. For that purpose, write the left hand side in the suggestive form

$$\nabla_z \cdot (f \, \overline{\boldsymbol{u}}) = \nabla_z \cdot [(f/D) \, \boldsymbol{U}], \tag{41.111}$$

with f/D reminiscient of the shallow water potential vorticity for the planetary geostrophic flow (Section 40.3). Motivated by this analog, we write the vorticity equation (41.112) in the form

$$\rho_0 \, \boldsymbol{U} \cdot \nabla_z (f/D) = -\rho_0 \, (f/D) \, \nabla_z \cdot \boldsymbol{U} + \hat{\boldsymbol{z}} \cdot \nabla_z \wedge [(p_{\mathrm{a}} - p_{\mathrm{b}}) \, \nabla_z (\eta/D)] \\ - D^{-2} \, \hat{\boldsymbol{z}} \cdot [\nabla_z \wedge (\mathscr{P} \, \nabla_z D)] + \hat{\boldsymbol{z}} \cdot [\nabla_z \wedge (D^{-1} \, \Delta \boldsymbol{\tau})]. \tag{41.112}$$

Contrary to the shallow water case, we here see that even for a perfect planetary geostrophic fluid, the 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.

41.6.2 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

$$\text{rigid lid approximation} \Longrightarrow \nabla_z \cdot \boldsymbol{U} = 0. \tag{41.113}$$

Furthermore, as part of the rigid lid approximation we assume the free surface undulations are much smaller than the resting ocean depth so that[3]

$$|\eta| \ll |\eta_{\mathrm{b}}| \Longrightarrow \eta/D \approx 0 \tag{41.114a}$$

$$1/(\eta - \eta_{\mathrm{b}}) \approx 1/(-\eta_{\mathrm{b}}) = 1/H. \tag{41.114b}$$

Hence, with the rigid lid approximation the vorticity equation (41.112) takes the simplified form

$$\rho_0 \, \boldsymbol{U} \cdot \nabla_z (f/H) = \hat{\boldsymbol{z}} \cdot [\nabla_z \mathscr{P} \wedge \nabla_z (1/H) + \nabla_z \wedge (H^{-1} \, \Delta \boldsymbol{\tau})]. \tag{41.115}$$

[3] We use the notation $\eta_{\mathrm{b}} = -H$ in this subsection to correspond to the literature for rigid lid models and the JEBAR term.

JEBAR drives deviations from f/H aligned flow

Equation (41.115) 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

$$\rho_0 \, H^2 \, \boldsymbol{U} \cdot \nabla_z (f/H) = \hat{\boldsymbol{z}} \cdot (\nabla_z H \wedge \nabla_z \mathscr{P}) = \hat{\boldsymbol{z}} \cdot (\nabla_z \wedge H \, \nabla_z \mathscr{P}) = -\hat{\boldsymbol{z}} \cdot (\nabla_z \wedge \mathscr{P} \, \nabla H). \tag{41.116}$$

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 (41.116) 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 horizontal average. For that purpose note that in the rigid lid approximation we have

$$\mathscr{P} = (1/2) \, g \, \overline{\rho} \, H^2 + g \int_{-H}^{0} \rho' \, \mathrm{d}z, \tag{41.117}$$

where $\overline{\rho}$ is the area mean density defined by equation (41.101). With $\nabla_z H \wedge \nabla_z (\overline{\rho} \, H^2) = 0$, we are left with just the baroclinic 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.

Relating JEBAR to pressure

JEBAR as given by the right hand side of equation (41.116) has the appearance of the curl of a form stress, and yet it is not. The reason is that $\mathscr{P}$ is the potential energy of the fluid column rather than bottom pressure. We make this point explicit by recalling the decomposition (41.99), here specialized to the rigid lid in which

$$\mathscr{P} = H \, (\overline{p} - p_{\mathrm{b}}) \qquad \text{with} \qquad \overline{p} = H^{-1} \int_{\eta_{\mathrm{b}}}^{0} p \, \mathrm{d}z. \tag{41.118}$$

Plugging into the vorticity equation (41.115) leads to

$$\rho_0 \, \boldsymbol{U} \cdot \nabla_z (f/H) = H^{-1} \, \hat{\boldsymbol{z}} \cdot [\nabla_z H \wedge \nabla_z (\overline{p} - p_{\mathrm{b}})] + \hat{\boldsymbol{z}} \cdot \nabla_z \wedge (H^{-1} \, \Delta \boldsymbol{\tau})]. \tag{41.119}$$

The JEBAR term thus arises from deviations of the depth averaged pressure from the bottom pressure.

Equation (41.106) provides the budget for the vorticity of the depth integrated flow, in which we find the curl of the topographic form stress that leads to vortex stretching. In contrast, the JEBAR term in equation (41.119) is not a pure vortex stretching term. Instead, it accounts for the fact that it is the horizontal bottom velocity flowing across isobaths, rather than depth averaged horizontal velocity, that leads to vortex stretching. When working with the vorticity of the depth averaged velocity, $\overline{\boldsymbol{u}}$, rather than the depth integrated velocity, $\boldsymbol{U}$, we must properly account for the role of vortex stretching, and to do so requires us to use JEBAR.

41.6.3 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.

42

Continuously stratified quasi-geostrophy

READER'S GUIDE FOR THIS CHAPTER

In this chapter we extend the shallow water discussions of quasi-geostrophy in Chapter 40 to continuously stratified fluids. We make use of stratified geophysical fluid dynamics from Chapters 18 and 26, as well as potential vorticity from Chapter 36 and the development of continuously stratified planetary geostrophy in Chapter 41. Material in this chapter is basic to the quasi-geostrophic theory of atmospheric and oceanic flows.

42.1 Open threads . . . **743**
42.2 Quasi-geostrophy . . . **744**
42.2.1 Zeroth order asymptotic equations . . . 745
42.2.2 First order asymptotic equations . . . 746
42.2.3 Dimensional QG-PV equation . . . 747
42.2.4 Properties of the steady state flow field . . . 748
42.2.5 Constant background buoyancy frequency . . . 748
42.2.6 Buoyancy advection at the boundaries . . . 749
42.3 Dimensions of various forms for potential vorticity . . . **750**
42.4 Connecting QG-PV to Ertel PV . . . **750**
42.4.1 Shallow water layer . . . 750
42.4.2 Continuously stratified hydrostatic Boussinesq fluid . . . 751
42.5 Energetics . . . **752**
42.5.1 Kinetic energy . . . 753
42.5.2 Available potential energy . . . 754
42.5.3 Energy conversion . . . 754
42.5.4 Scaling APE and KE . . . 755
42.6 Exercises . . . **755**

42.1 Open threads

- Rossby waves
- Impermeability theorem
- Synoptic development

- TEM

42.2 Quasi-geostrophy

We proceed much like for the single layer of shallow water fluid in Section 40.4. In particular, quasi-geostrophic scaling from Section 40.4.1 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. For this purpose, recall the non-dimensional momentum and continuity equations from Section 41.1.8

$$Ro\left[\frac{\partial \hat{\boldsymbol{u}}}{\partial \hat{t}} + (\hat{\boldsymbol{u}}\cdot\hat{\nabla}_{\hat{z}})\,\hat{\boldsymbol{u}} + \hat{w}\frac{\partial \hat{\boldsymbol{u}}}{\partial \hat{z}}\right] + \hat{\boldsymbol{f}}\wedge\hat{\boldsymbol{u}} = -\hat{\nabla}_z\,\hat{\varphi} \tag{42.1a}$$

$$\frac{\partial \hat{\varphi}}{\partial \hat{z}} = \hat{b} \tag{42.1b}$$

$$\hat{\nabla}\cdot\hat{\boldsymbol{v}} = 0 \tag{42.1c}$$

$$Ro\,F\,\frac{\mathrm{D}\hat{b}}{\mathrm{D}\hat{t}} + \hat{w} = 0. \tag{42.1d}$$

We expand the prognostic variables in an asymptotic series in Rossby number

$$\hat{u} = \hat{u}_0 + Ro\,\hat{u}_1 + Ro^2\,\hat{u}_2 + \ldots \tag{42.2a}$$

$$\hat{v} = \hat{v}_0 + Ro\,\hat{v}_1 + Ro^2\,\hat{v}_2 + \ldots \tag{42.2b}$$

$$\hat{w} = \hat{w}_0 + Ro\,\hat{w}_1 + Ro^2\,\hat{w}_2 + \ldots \tag{42.2c}$$

$$\hat{b} = \hat{b}_0 + Ro\,\hat{b}_1 + Ro^2\,\hat{b}_2 + \ldots \tag{42.2d}$$

along with the expansion (40.56) for the Coriolis parameter

$$\hat{f} = \hat{f}_0 + Ro\,\hat{\beta}\,\hat{y}, \tag{42.3}$$

and where (equation (40.57))

$$\hat{\beta}\,\hat{y} = \frac{\beta\,y}{Ro\,f_0} = T\,\beta\,y. \tag{42.4}$$

Furthermore, under quasi-geostrophic scaling it is important to retain the depth dependence of the Burger number through its dependence on the background stratification $N^2(z)$

$$F(z) = \left[\frac{L}{L_{\mathrm{d}}}\right]^2 = \frac{1}{[N(z)]^2}\left[\frac{L\,f_0}{H}\right]^2. \tag{42.5}$$

Additionally, 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

$$Bu = F^{-1} \sim 1 \Rightarrow L \sim L_{\mathrm{d}}. \tag{42.6}$$

42.2.1 Zeroth order asymptotic equations

The zeroth order asymptotic equations take the form

$$\hat{\boldsymbol{f}}_0 \wedge \hat{\boldsymbol{u}}_0 = -\hat{\nabla}_z \, \hat{\varphi}_0 \tag{42.7a}$$

$$\frac{\partial \hat{\varphi}_0}{\partial \hat{z}} = \hat{b}_0 \tag{42.7b}$$

$$\hat{\nabla}_z \cdot \hat{\boldsymbol{u}}_0 + \frac{\partial \hat{w}_0}{\partial \hat{z}} = 0 \tag{42.7c}$$

$$\hat{w}_0 = 0. \tag{42.7d}$$

The first equation represents f-plane geostrophy, which means that the horizontal velocity has zero divergence

$$\hat{\nabla}_z \cdot \hat{\boldsymbol{u}}_0 = 0. \tag{42.8}$$

Equation (42.7b) means the zeroth order buoyancy determines the zeroth order hydrostatic pressure. Since the horizontal velocity has zero divergence, the continuity equation (42.7c) means that the vertical velocity is depth independent

$$\frac{\partial \hat{w}_0}{\partial \hat{z}} = 0. \tag{42.9}$$

If it vanishes somewhere, such as a solid boundary, then it vanishes everywhere. This is a manifestation of the Taylor-Proudman theorem (see Section 26.4.3). Indeed, a vanishing $\hat{w}_0$ is required by the zeroth-order buoyancy equation (42.7d). Hence, the non-dimensional velocity has a nonzero contribution only at order Ro^1

$$\hat{w} = Ro \, \hat{w}_1 + Ro^2 \, \hat{w}_2 + \ldots, \tag{42.10}$$

thus manifesting the vertical stiffening of fluid columns found in rotating fluids. Hence, the dimensionful vertical velocity has the asymptotic expansion

$$w = W \hat{w} = W Ro \, (\hat{w}_1 + Ro \, \hat{w}_2 + \ldots), \tag{42.11}$$

so that to leading to order Ro^1

$$\hat{w}_1 = \frac{w}{W Ro}. \tag{42.12}$$

Since the zeroth-order velocity is non-divergent, we can introduce a geostrophic streamfunction

$$\hat{u}_0 = -\frac{\partial \hat{\psi}_0}{\partial \hat{y}} \qquad \hat{v}_0 = \frac{\partial \hat{\psi}_0}{\partial \hat{x}} \qquad \hat{\zeta}_0 = \hat{\nabla}^2 \hat{\psi}_0, \tag{42.13}$$

where the zeroth-order streamfunction is the ratio of the zeroth order pressure to zeroth order Coriolis parameter

$$\hat{\psi}_0 = \frac{\hat{\varphi}_0}{\hat{f}_0}. \tag{42.14}$$

Note also that the zeroth-order system satisfies the thermal wind balance

$$\hat{\boldsymbol{f}}_0 \wedge \frac{\partial \hat{\boldsymbol{u}}_0}{\partial \hat{z}} = -\hat{\nabla}_z \, \hat{b}_0. \tag{42.15}$$

Finally, note that the zeroth order buoyancy is related to the streamfunction through the hydrostatic balance

$$\hat{b}_0 = \frac{\partial \hat{\varphi}_0}{\partial \hat{z}} = \hat{f}_0 \, \frac{\partial \hat{\psi}_0}{\partial \hat{z}}. \tag{42.16}$$

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

$$\frac{\mathrm{D}_0 \hat{\boldsymbol{u}}_0}{\mathrm{D}\hat{t}} + \hat{\boldsymbol{f}}_0 \wedge \hat{\boldsymbol{u}}_1 + \hat{\beta}\, \hat{y}\, \hat{\boldsymbol{z}} \wedge \hat{\boldsymbol{u}}_0 = -\hat{\nabla}_z \hat{\varphi}_1 \tag{42.17a}$$

$$\hat{\nabla}_z \cdot \hat{\boldsymbol{u}}_1 + \frac{\partial \hat{w}_1}{\partial \hat{z}} = 0 \tag{42.17b}$$

$$F\, \frac{\mathrm{D}_0 \hat{b}_0}{\mathrm{D}\hat{t}} + \hat{w}_1 = 0. \tag{42.17c}$$

At this order, the material time derivative makes use *only* of the zeroth order geostrophic horizontal velocity

$$\frac{\mathrm{D}_0}{\mathrm{D}\hat{t}} = \frac{\partial}{\partial \hat{t}} + \hat{\boldsymbol{u}}_0 \cdot \hat{\nabla}. \tag{42.18}$$

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 (42.17a) eliminates the pressure gradient, $\hat{\nabla}\hat{\varphi}_1$, thus producing the vorticity equation

$$\frac{\partial \hat{\zeta}_0}{\partial \hat{t}} + (\hat{\boldsymbol{u}}_0 \cdot \hat{\nabla})\, (\hat{\zeta}_0 + \hat{\beta}\, \hat{y}) = -\hat{f}_0\, \hat{\nabla}_z \cdot \boldsymbol{u}_1. \tag{42.19}$$

We make use of the continuity equation (42.17b) to eliminate the horizontal convergence

$$\frac{\partial \hat{\zeta}_0}{\partial \hat{t}} + (\hat{\boldsymbol{u}}_0 \cdot \hat{\nabla})\, (\hat{\zeta}_0 + \hat{\beta}\, \hat{y}) = \hat{f}_0\, \frac{\partial \hat{w}_1}{\partial \hat{z}}. \tag{42.20}$$

The right hand side represents the contribution to vorticity evolution from stretching by planetary rotation. We can now eliminate the vertical velocity through the buoyancy equation (42.17c). When doing so, it is important to keep the $F(z)$ depth dependence according to equation (42.5), with this depth dependence arising from the background stratification. The resulting vorticity equation is

$$\frac{\partial \hat{\zeta}_0}{\partial \hat{t}} + (\hat{\boldsymbol{u}}_0 \cdot \hat{\nabla})\, (\hat{\zeta}_0 + \hat{\beta}\, \hat{y}) = -\hat{f}_0\, \frac{\partial}{\partial \hat{z}} \left[F\, \frac{\mathrm{D}_0 \hat{b}_0}{\mathrm{D}\hat{t}} \right]. \tag{42.21}$$

We now use the identity

$$\frac{\partial}{\partial \hat{z}} \left[F(\hat{z})\, \frac{\mathrm{D}_0 \hat{b}_0}{\mathrm{D}\hat{t}} \right] = \frac{\partial}{\partial \hat{z}} \left[F \left(\frac{\partial}{\partial \hat{t}} + \hat{\boldsymbol{u}}_0 \cdot \hat{\nabla} \right) \hat{b}_0 \right] \tag{42.22a}$$

$$= \frac{\mathrm{D}_0}{\mathrm{D}\hat{t}} \left[\frac{\partial}{\partial \hat{z}} \left(F\, \hat{b}_0 \right) \right] + F\, \frac{\partial \hat{\boldsymbol{u}}_0}{\partial \hat{z}} \cdot \hat{\nabla}_z \hat{b}_0 \tag{42.22b}$$

$$= \frac{\mathrm{D}_0}{\mathrm{D}\hat{t}} \left[\frac{\partial}{\partial \hat{z}} \left(F\, \hat{b}_0 \right) \right], \tag{42.22c}$$

where we set

$$\frac{\partial \hat{\boldsymbol{u}}_0}{\partial \hat{z}} \cdot \hat{\nabla}_z \hat{b}_0 = 0 \tag{42.23}$$

since the zeroth-order velocity maintains thermal wind balance (42.15). Bringing terms together then leads to the material conservation equation for quasi-geostrophic potential vorticity

$$\frac{\mathrm{D}_0}{\mathrm{D}\hat{t}} \left[\hat{\zeta}_0 + \hat{\beta}\, \hat{y} + \hat{f}_0\, \frac{\partial}{\partial \hat{z}} \left(F\, \hat{b}_0 \right) \right] = 0. \tag{42.24}$$

42.2.3 Dimensional QG-PV equation

From equation (42.24), we identify the non-dimensional QG-PV

$$\hat{q} = \hat{\zeta} + \hat{\beta}\,\hat{y} + \hat{f}_0\,\frac{\partial(F\,\hat{b}_0)}{\partial\hat{z}}, \tag{42.25}$$

where we dropped the 0 asymptotic subscript for brevity. Introducing dimensional quantities to the right hand side yields (recall $\hat{f}_0 = 1$ and $\hat{\beta}\,\hat{y} = T\,\beta\,y = (L/U)\,\beta\,y$)

$$\hat{q} = \frac{L}{U}\,[\zeta + \beta\,y] + \frac{\partial}{\partial z}\left[\frac{H\,F\,b'}{B}\right]. \tag{42.26}$$

The scale for the fluctuating buoyancy is given by equation (41.22)

$$B = \frac{f_0\,U\,L}{H}, \tag{42.27}$$

and the inverse Burger number is given by equation (42.5)

$$F(z) = \left[\frac{L}{L_{\mathrm{d}}}\right]^2 = \frac{1}{[N(z)]^2}\left[\frac{L\,f_0}{H}\right]^2. \tag{42.28}$$

These terms then yield for the non-dimensional PV

$$\hat{q} = \frac{L}{U}\,(\zeta + \beta\,y) + \frac{\partial}{\partial z}\left[\frac{H\,F\,b'}{B}\right] \tag{42.29a}$$

$$= \frac{L}{U}\,[\zeta + \beta\,y] + \frac{1}{Ro}\left[\frac{\partial}{\partial z}\left(\frac{b'}{N^2}\right)\right] \tag{42.29b}$$

$$= \frac{L}{U}\left[\nabla_z^2\psi + \beta\,y\right] + \frac{f_0}{Ro}\left[\frac{\partial}{\partial z}\left(\frac{1}{N^2}\frac{\partial\psi}{\partial z}\right)\right], \tag{42.29c}$$

where we introduced the QG streamfunction

$$\psi = \frac{\varphi'}{f_0} \tag{42.30}$$

for the final equality, and made use of the identities

$$u = -\frac{\partial\psi}{\partial y} \qquad v = \frac{\partial\psi}{\partial x} \qquad \zeta = \nabla_z^2\psi \qquad b' = f_0\left[\frac{\partial\psi}{\partial z}\right]. \tag{42.31}$$

We thus identify the dimensional QG-PV

$$q = Ro\,f_0\,\hat{q} \tag{42.32a}$$

$$= \zeta + \beta\,y + \frac{\partial}{\partial z}\left[\frac{f_0^2}{N^2}\frac{\partial\psi}{\partial z}\right] \tag{42.32b}$$

$$= \nabla_z^2\psi + \beta\,y + \frac{\partial}{\partial z}\left[\frac{f_0^2}{N^2}\frac{\partial\psi}{\partial z}\right]. \tag{42.32c}$$

We can add a constant to q without changing the dynamics. Consequently, some authors like to add f_0 in which case

$$q = \underbrace{\nabla_z^2\psi}_{\text{relative vorticity}} + \underbrace{f_0 + \beta\,y}_{\text{planetary vorticity}} + \underbrace{\frac{\partial}{\partial z}\left[\frac{f_0^2}{N^2}\frac{\partial\psi}{\partial z}\right]}_{\text{stretching by } f}. \tag{42.33}$$

There are three contributions to the QG-PV:

- relative vorticity of the geostrophic flow;
- planetary vorticity due to the rotation of the reference frame;
- stretching due to motion on the rotating planet (see equation (42.20) to be reminded why the third term represents stretching).

42.2.4 Properties of the steady state flow field

The geostrophic velocity takes the following form in terms of the QG streamfunction

$$\boldsymbol{u} = \hat{\boldsymbol{z}} \wedge \nabla\psi. \tag{42.34}$$

This equality then allows us to write the following equivalent forms for the material time derivative of QG-PV

$$\frac{\mathrm{D}q}{\mathrm{D}t} = \frac{\partial q}{\partial t} + \boldsymbol{u}\cdot\nabla q \tag{42.35a}$$

$$= \frac{\partial q}{\partial t} + (\hat{\boldsymbol{z}} \wedge \nabla\psi)\cdot\nabla q \tag{42.35b}$$

$$= \frac{\partial q}{\partial t} + (\nabla\psi \wedge \nabla q)\cdot\hat{\boldsymbol{z}} \tag{42.35c}$$

$$= \frac{\partial q}{\partial t} + J(\psi, q), \tag{42.35d}$$

where the final equality introduced the Jacobian operator

$$J(\psi, q) = \frac{\partial\psi}{\partial x}\frac{\partial q}{\partial y} - \frac{\partial\psi}{\partial y}\frac{\partial q}{\partial x}. \tag{42.36}$$

For a perfect fluid, in which $\mathrm{D}q/\mathrm{D}t = 0$, a steady state (zero Eulerian time derivative) is realized when

$$\boldsymbol{u}\cdot\nabla q = (\nabla\psi \wedge \nabla q)\cdot\hat{\boldsymbol{z}} = J(\psi, q) = 0. \tag{42.37}$$

The first expression says that the velocity field is aligned parallel to surfaces of constant q, with the second expression saying the same. We are ensured that these equalities hold if the streamfunction is a function only of the potential vorticity

$$\psi = \psi(q) \Rightarrow J(\psi, q) = 0. \tag{42.38}$$

42.2.5 Constant background buoyancy frequency

Consider the QG-PV for the special case of a constant background buoyancy frequency,

$$N^2 = \text{constant}. \tag{42.39}$$

For this case the QG-PV in equation (42.32c) can be written

$$q = \beta\, y + \nabla_z^2\psi + \frac{\partial}{\partial z}\left[\frac{f_0^2}{N^2}\frac{\partial\psi}{\partial z}\right] \tag{42.40a}$$

$$= \beta\, y + \frac{\partial^2\psi}{\partial x^2} + \frac{\partial^2\psi}{\partial y^2} + \frac{f_0^2}{N^2}\frac{\partial^2\psi}{\partial z^2} \tag{42.40b}$$

$$= \beta\, y + \frac{\partial^2\psi}{\partial x^2} + \frac{\partial^2\psi}{\partial y^2} + \frac{\partial^2\psi}{\partial \tilde{z}^2}. \tag{42.40c}$$

For the final equality we introduced the vertical coordinate

$$\tilde{z} = \frac{N}{f}\, z. \tag{42.41}$$

Since $N/f \gg 1$ for the stably stratified flows considered in QG, we term $\tilde{z}$ a *stretched* vertical coordinate.

42.2.6 Buoyancy advection at the boundaries

We need boundary conditions in order to invert the elliptic QG-PV equation (42.33) to solve for the streamfunction ψ. For lateral boundaries, one may choose periodicity, whereby the boundaries are in effect absent. Alternatively, we may choose to set the normal component of the flow to zero for the inviscid case. The top and bottom boundaries are less trivial and require some care.

Returning to the Boussinesq equations (41.1a)-(41.1c), we focus on the buoyancy equation, written as in equation (41.4)

$$\frac{\mathrm{D}b'}{\mathrm{D}t} + w\, N^2 = 0, \tag{42.42}$$

where advection is via the horizontal geostrophic currents. Inserting the geostrophic streamfunction

$$b' = f_0\, \frac{\partial \psi}{\partial z} \tag{42.43}$$

leads to

$$\frac{\partial}{\partial t}\left[\frac{\partial \psi}{\partial z}\right] + \boldsymbol{u}\cdot\nabla\left[\frac{\partial \psi}{\partial z}\right] + w\, N^2 = 0. \tag{42.44}$$

Consider two contributions to nonzero vertical velocity at a boundary. The first arises from slopes in the topography, $\nabla\eta_{\mathrm{b}} \neq 0$. The no-normal flow condition means that at the boundary, the velocity is constrained so that

$$\boldsymbol{v}\cdot\hat{\boldsymbol{n}} = 0 \Rightarrow w = \boldsymbol{u}\cdot\nabla\eta_{\mathrm{b}}, \tag{42.45}$$

where $\hat{\boldsymbol{n}}$ is the outward normal to the boundary (Section 14.4.1). The second contribution arises from Ekman pumping or suction (Section 28.1). In general, flow in an Ekman layer generates vorticity due to curl in the boundary stresses. We abstract this process by stating that at the boundary of the Ekman layer and the interior flow, the vertical velocity component takes on a value (see Section 14.1.1 of *Vallis* (2006))

$$w_{\mathrm{Ekman}} = r\, \nabla_z^2 \psi, \tag{42.46}$$

where r is a length scale proportional to the Ekman layer thickness, and $\nabla_z^2\psi$ is the quasi-geostrophic vorticity. Bringing these two effects together leads to the boundary condition

$$w = \boldsymbol{u}\cdot\nabla\eta_{\mathrm{b}} + r\, \nabla_z^2\psi. \tag{42.47}$$

Using this expression in the buoyancy equation (42.44) leads to the boundary evolution of buoyancy

$$\frac{\partial}{\partial t}\left[\frac{\partial \psi}{\partial z}\right] + \boldsymbol{u}\cdot\nabla\left[\frac{\partial \psi}{\partial z} + N^2\, \eta_{\mathrm{b}}\right] + N^2\, r\, \nabla_z^2\psi = 0. \tag{42.48}$$

We say that the QG fluid system is characterized by horizontal advection of PV in the interior, and advection of buoyancy on the boundaries.

42.3 Dimensions of various forms for potential vorticity

The following dimensions are taken by the various forms of potential vorticity seen thus far in these notes.

- Shallow water PV:
$$\frac{\zeta + f}{h} \; [\equiv] \; (\text{time} \times \text{length})^{-1}. \tag{42.49}$$

- Entropic Ertel PV based on potential temperature,
$$\frac{\boldsymbol{\omega}_a \cdot \nabla\theta}{\rho} \; [\equiv] \; \frac{\text{length}^2 \times \text{temperature}}{\text{time} \times \text{mass}}, \tag{42.50}$$
where θ is the potential temperature.

- Ertel PV with buoyancy in a Boussinesq fluid
$$\boldsymbol{\omega}_a \cdot \nabla b \; [\equiv] \; \text{time}^{-3}. \tag{42.51}$$

- Quasi-geostrophic PV
$$q = \nabla_z^2 \psi + \beta\, y + \frac{\partial}{\partial z}\left[\frac{f_0^2}{N^2}\frac{\partial \psi}{\partial z}\right] \; [\equiv] \; \text{time}^{-1}. \tag{42.52}$$

42.4 Connecting QG-PV to Ertel PV

Equations describing both the PG and QG systems can be encapsulated by the material conservation of PG or QG potential vorticity. We here determine how potential vorticity for PG and QG flows relates to the Ertel PV from Chapter 36.

42.4.1 Shallow water layer

The Ertel PV for a single layer of shallow water fluid is (Section 33.3)
$$Q = \frac{f + \zeta}{h}, \tag{42.53}$$
where
$$h = H + \Delta\eta - \eta_{\text{b}} \tag{42.54}$$
is the thickness of the layer (see Figure 29.1). We now consider the limit as the Rossby number is small, and one of the following two regimes for the thickness.

- PG: free surface height and bottom topography undulations are on the order of the resting depth H.

- QG: free surface height undulations ($\Delta\eta$) are small, and bottom topography deviations are small (η_{b} can be neglected);

The PG limit is simplest, in which case we merely drop the relative vorticity and keep the full form of the layer thickness

$$Q_{\rm PG} = \frac{f}{h}. \tag{42.55}$$

The QG limit requires a bit more algebra, whereby

$$Q = \frac{\zeta + f}{h} \tag{42.56a}$$

$$= \frac{\zeta + f}{H\,(1 + \Delta\eta/H)} \tag{42.56b}$$

$$\approx \frac{1}{H}\,(\zeta + f)\left[1 - \frac{\Delta\eta}{H}\right] \tag{42.56c}$$

$$\approx \frac{1}{H}\left[\zeta + f_0 + \beta\,y - f_0\,\frac{\Delta\eta}{H}\right]. \tag{42.56d}$$

42.4.2 Continuously stratified hydrostatic Boussinesq fluid

We now consider the connection between Ertel PV and QG-PV for the continuously stratified hydrostatic Boussinesq fluid. For this purpose, make use of the Ertel PV derived in Exercise 36.1, for which

$$Q = \boldsymbol{\omega}_a \cdot \nabla b = \frac{\partial u}{\partial z}\frac{\partial b}{\partial y} - \frac{\partial v}{\partial z}\frac{\partial b}{\partial x} + (\zeta + f)\,\frac{\partial b}{\partial z}. \tag{42.57}$$

Now split the buoyancy into a depth dependent background and a deviation from the background

$$b = \tilde{b}(z) + b'(x, y, z, t), \tag{42.58}$$

and write its vertical derivative as

$$\frac{\partial b}{\partial z} = N^2 + \frac{\partial b'}{\partial z}, \tag{42.59}$$

where

$$N^2 = \frac{\partial \tilde{b}}{\partial z} \tag{42.60}$$

is the squared buoyancy frequency for the background buoyancy field. Rather than introducing non-dimensional variables[1], we work more briefly by arranging the terms in the PV according to their QG scaling in terms of the Rossby number

$$Q = [f_0\,N^2] + \left[(\beta\,y + \zeta)\,N^2 + f_0\,\frac{\partial b'}{\partial z}\right] + \left[(\beta\,y + \zeta)\frac{\partial b'}{\partial z} + \frac{\partial u}{\partial z}\frac{\partial b'}{\partial y} - \frac{\partial v}{\partial z}\frac{\partial b'}{\partial x}\right]. \tag{42.61}$$

We drop the third bracket term, as it is order Ro^2, and write

$$Q = \tilde{Q} + N^2\,q_*, \tag{42.62}$$

where

$$\tilde{Q} = f_0\,N^2 \tag{42.63}$$

is the f-plane planetary geostrophic PV, and

$$q_* = \beta\,y + \zeta + \frac{f_0}{N^2}\,\frac{\partial b'}{\partial z}. \tag{42.64}$$

[1]SMG: Should introduce non-dimensional variables in future versions of the notes to enhance the analysis from *Vallis* (2006) Section 5.5.1.

The material conservation of PV now takes the form

$$N^2 \frac{\mathrm{D}q_*}{\mathrm{D}t} + w \left[1 + \frac{q_*}{f_0}\right] \frac{\partial \tilde{Q}}{\partial z} = 0 \tag{42.65}$$

where advection in the material time derivative operator is now assumed to occur from the horizontal geostrophic velocity. We drop the term q_*/f_0 as it is small, and divide by N^2, to have

$$\frac{\mathrm{D}q_*}{\mathrm{D}t} + \frac{w}{N^2} \frac{\partial \tilde{Q}}{\partial z} = 0. \tag{42.66}$$

To eliminate the vertical velocity component, we introduce the buoyancy equation

$$\frac{\mathrm{D}b'}{\mathrm{D}t} + w\, N^2 = 0, \tag{42.67}$$

so that

$$\frac{\mathrm{D}q_*}{\mathrm{D}t} - \frac{f_0}{N^4} \frac{\partial N^2}{\partial z} \frac{\mathrm{D}b'}{\mathrm{D}t} = 0. \tag{42.68}$$

Writing

$$\frac{\partial}{\partial z}\left[\frac{1}{N^2}\right] = -\frac{1}{N^4}\frac{\partial N^2}{\partial z} \tag{42.69}$$

leads to

$$\frac{\mathrm{D}q_*}{\mathrm{D}t} + f_0 \left[\frac{\partial N^{-2}}{\partial z}\right] \frac{\mathrm{D}b'}{\mathrm{D}t} = 0. \tag{42.70}$$

Since the material time derivative operator only involves horizontal advection, we can merge these two terms to render

$$\frac{\mathrm{D}q_*}{\mathrm{D}t} + f_0 \left(\frac{\partial N^{-2}}{\partial z}\right) \frac{\mathrm{D}b'}{\mathrm{D}t} = \frac{\mathrm{D}}{\mathrm{D}t}\left[q_* + f_0\, b' \left(\frac{\partial N^{-2}}{\partial z}\right)\right] \tag{42.71a}$$

$$= \frac{\mathrm{D}}{\mathrm{D}t}\left[\beta\, y + \zeta + \frac{f_0}{N^2}\left(\frac{\partial b'}{\partial z}\right) + f_0\, b' \left(\frac{\partial N^{-2}}{\partial z}\right)\right] \tag{42.71b}$$

$$= \frac{\mathrm{D}}{\mathrm{D}t}\left[\beta\, y + \zeta + f_0 \frac{\partial}{\partial z}\left(\frac{b'}{N^2}\right)\right] \tag{42.71c}$$

$$= 0. \tag{42.71d}$$

The term inside the bracket is the QG-PV given by equation (42.33).

42.5 Energetics

Consider the QG system with flat top and flat bottom boundaries, and assume for the lateral directions either periodicity or constant streamfunction on solid boundaries. The QG vorticity equation and buoyancy equations are given by

$$\frac{\partial \zeta}{\partial t} + \boldsymbol{u} \cdot \nabla \zeta = f_0 \frac{\partial w}{\partial z} \tag{42.72a}$$

$$\frac{\partial b}{\partial t} + \boldsymbol{u} \cdot \nabla b = -w\, N^2, \tag{42.72b}$$

where

$$\frac{\mathrm{D}}{\mathrm{D}t} = \frac{\partial}{\partial t} + \boldsymbol{u} \cdot \nabla \qquad b' = f_0 \frac{\partial \psi}{\partial z} \qquad u = -\frac{\partial \psi}{\partial y} \qquad v = \frac{\partial \psi}{\partial x} \qquad \zeta = \nabla_z^2 \psi. \tag{42.73}$$

42.5.1 Kinetic energy

The kinetic energy per mass for the total fluid domain (assumed to be a constant volume materia domain) is given by the integral

$$\mathcal{K} = \frac{1}{2}\int (\boldsymbol{u}\cdot\boldsymbol{u})\,\mathrm{d}V = \frac{1}{2}\int (\nabla\psi\cdot\nabla\psi)\,\mathrm{d}V, \tag{42.74}$$

and its time derivative is

$$\frac{\mathrm{d}\mathcal{K}}{\mathrm{d}t} = \int \nabla\psi\cdot\nabla\left[\frac{\partial\psi}{\partial t}\right]\mathrm{d}V. \tag{42.75}$$

For this result, we assumed the fluid domain has a constant volume to thus allow the time derivative to move inside the integral without introducing boundary terms. Manipulation renders

$$\frac{\mathrm{d}\mathcal{K}}{\mathrm{d}t} = \int \nabla\psi\cdot\nabla\left[\frac{\partial\psi}{\partial t}\right]\mathrm{d}V \tag{42.76a}$$

$$= \int \left[\nabla\cdot[\psi\,\nabla(\partial\psi/\partial t)] - \psi\,\partial(\nabla_z^2\psi)/\partial t\right]\mathrm{d}V \tag{42.76b}$$

$$= -\int \psi\,\frac{\partial\zeta}{\partial t}\,\mathrm{d}V, \tag{42.76c}$$

where we dropped the lateral boundary term and introduced relative vorticity. Use of the vorticity equation (42.72a) yields

$$\frac{\mathrm{d}\mathcal{K}}{\mathrm{d}t} = -\int \psi\,\frac{\partial\zeta}{\partial t}\,\mathrm{d}V \tag{42.77a}$$

$$= \int \psi\,[\boldsymbol{u}\cdot\nabla\zeta - f_0\,\partial_z\,w]\,\mathrm{d}V. \tag{42.77b}$$

The first term vanishes, since

$$\int \psi\,[\boldsymbol{u}\cdot\nabla\zeta]\,\mathrm{d}V = \int \psi\,\nabla\cdot(\boldsymbol{u}\,\zeta)\,\mathrm{d}V \tag{42.78a}$$

$$= \int [\nabla\cdot(\psi\,\boldsymbol{u}\,\zeta) - \nabla\psi\cdot\boldsymbol{u}\,\zeta]\,\mathrm{d}V \tag{42.78b}$$

$$= 0, \tag{42.78c}$$

where the boundary term vanishes and $\boldsymbol{u}\cdot\nabla\psi = 0$ since ψ is the streamfunction for the horizontal geostrophic flow. We are thus left with the expression for the kinetic energy evolution

$$\frac{\mathrm{d}\mathcal{K}}{\mathrm{d}t} = -\int \psi\,f_0\,\frac{\partial w}{\partial z}\,\mathrm{d}V. \tag{42.79}$$

Since the top and bottom are assumed flat, the vertical velocity vanishes on these boundaries (rigid lid top and solid bottom), in which case we can write

$$\frac{\mathrm{d}\mathcal{K}}{\mathrm{d}t} = -\int \psi\,f_0\,\frac{\partial w}{\partial z}\,\mathrm{d}V \tag{42.80a}$$

$$= -\int f_0\left[\frac{\partial(w\,\psi)}{\partial z} - w\,\frac{\partial\psi}{\partial z}\right]\mathrm{d}V \tag{42.80b}$$

$$= \int f_0\,w\,\frac{\partial\psi}{\partial z}\,\mathrm{d}V, \tag{42.80c}$$

where we dropped the boundary term given the rigid boundaries. Making use of

$$b' = f_0 \, \partial\psi/\partial z \tag{42.81}$$

leads to

$$\frac{\mathrm{d}\mathcal{K}}{\mathrm{d}t} = \int w \, b' \mathrm{d}V. \tag{42.82}$$

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, as discussed in Section 23.3. It is reassuring to see the same behavior in the quasi-geostrophic system.

42.5.2 Available potential energy

Available potential energy was introduced in Section 3.10 of *Vallis* (2006). Specialing that expression to the QG fluid leads to the available potential energy

$$\mathcal{A} = \frac{1}{2} \int \left[\frac{f_0}{N} \frac{\partial \psi}{\partial z} \right]^2 \mathrm{d}V. \tag{42.83}$$

Taking a time derivative leads to

$$\frac{\mathrm{d}\mathcal{A}}{\mathrm{d}t} = \int \left[\frac{f_0}{N} \right]^2 \frac{\partial \psi}{\partial z} \frac{\partial}{\partial t} \frac{\partial \psi}{\partial z} \, \mathrm{d}V \tag{42.84a}$$

$$= \int \frac{f_0}{N^2} \frac{\partial \psi}{\partial z} \left[-w \, N^2 - \nabla \cdot (\boldsymbol{u} \, b) \right] \mathrm{d}V, \tag{42.84b}$$

where we used the buoyancy equation (42.72b) for the second equality. Recall that the divergence operator acts just in the horizontal since $\boldsymbol{u}$ is the horizontal velocity. The second term vanishes, since

$$\int \frac{f_0}{N^2} \frac{\partial \psi}{\partial z} \left[\nabla \cdot (\boldsymbol{u} \, b) \right] \mathrm{d}V = \int \left[\frac{f_0^2}{N^2} \frac{\partial \psi}{\partial z} \right] \boldsymbol{u} \cdot \nabla \left[\frac{\partial \psi}{\partial z} \right] \mathrm{d}V \tag{42.85a}$$

$$= \frac{1}{2} \int \nabla \cdot \left[\boldsymbol{u} \left(\frac{f_0}{N} \frac{\partial \psi}{\partial z} \right)^2 \right] \mathrm{d}V \tag{42.85b}$$

$$= 0. \tag{42.85c}$$

Consequently, the APE changes according to

$$\frac{\mathrm{d}\mathcal{A}}{\mathrm{d}t} = - \int w \, f_0 \frac{\partial \psi}{\partial z} \, \mathrm{d}V = - \int w \, b' \, \mathrm{d}V. \tag{42.86}$$

42.5.3 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 kinetic energy increases through buoyancy work, available potential energy decreases

$$\frac{\mathrm{d}(\mathcal{K} + \mathcal{A})}{\mathrm{d}t} = 0. \tag{42.87}$$

The buoyancy work conversion term is given by

$$\text{buoyancy work} = \int w\, f_0 \frac{\partial \psi}{\partial z}\, \mathrm{d}V = \int w\, b'\, \mathrm{d}V, \tag{42.88}$$

which has the same form as that encountered for the conversion between potential and kinetic energy in the unapproximated equations (Section 23.3

42.5.4 Scaling APE and KE

The scale for the kinetic energy is given by

$$\mathcal{K} = \frac{1}{2} \int (\nabla \psi \cdot \nabla \psi)\, \mathrm{d}V \sim L^{-2}\, \Psi\, V \tag{42.89}$$

and the APE scale is

$$\mathcal{A} = \frac{1}{2} \int \left[\frac{f_0}{N} \frac{\partial \psi}{\partial z} \right]^2 \mathrm{d}V \sim L_{\mathrm{d}}^{-2}\, \Psi\, V, \tag{42.90}$$

where we wrote Ψ for the streamfunction scale, and V for the domain volume. Taking the ratio yields

$$\frac{\mathcal{K}}{\mathcal{A}} \sim \left[\frac{L_{\mathrm{d}}}{L} \right]^2 = \left[\frac{H}{L} \right]^2 \left[\frac{N}{f_0} \right]^2 = Bu. \tag{42.91}$$

The Burger number is hence the ratio of the kinetic energy scale to the 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.

42.6 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: most of 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) Continously stratified quasi-geostrophic flow on a beta plane

EXERCISE 42.2: QG-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 QG-PV in a continuously stratified Boussinesq fluid in the presence of friction, $\boldsymbol{F}$. 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

$$\boldsymbol{F} = \frac{\partial}{\partial z}\left[\nu\,\frac{\partial \boldsymbol{u}}{\partial z}\right], \tag{42.92}$$

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 QG-PV in which we drop relative vorticity (i.e., QG-PV is dominated by planetary vorticity and stretching). Determine the form for the vertical eddy viscosity so that the approximate form of QG-PV is laterally diffused via

$$\frac{\mathrm{D}q^{\mathrm{approx}}}{\mathrm{D}t} = A\,\nabla_z^2 q^{\mathrm{approx}}, \tag{42.93}$$

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: TRADITIONAL FORM OF THE QUASI-GEOSTROPHIC ω-EQUATION

As discussed in Section 42.2, 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 vertical velocity. However, the vertical velocity is non-zero and 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 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) From Section 42.2.2, the non-dimensional velocity, continuity, buoyancy and vorticity equations for β-plane flow valid at order Ro^1 are given by

$$\frac{\mathrm{D}_0\hat{\boldsymbol{u}}_0}{\mathrm{D}\hat{t}} + \hat{f}_0\,\hat{\boldsymbol{k}}\wedge\hat{\boldsymbol{u}}_1 + \hat{\beta}\,\hat{y}\,\hat{\boldsymbol{k}}\wedge\hat{\boldsymbol{u}}_0 = -\hat{\nabla}_z\hat{\varphi}_1 \tag{42.94a}$$

$$\hat{\nabla}_z\cdot\hat{\boldsymbol{u}}_1 + \frac{\partial\hat{w}_1}{\partial\hat{z}} = 0 \tag{42.94b}$$

$$F\,\frac{\mathrm{D}_0\hat{b}_0}{\mathrm{D}\hat{t}} + \hat{w}_1 = 0 \tag{42.94c}$$

$$\frac{\partial\hat{\zeta}_0}{\partial\hat{t}} + (\hat{\boldsymbol{u}}_0\cdot\hat{\nabla})\,(\hat{\zeta}_0 + \hat{\beta}\,\hat{y}) = \hat{f}_0\,\frac{\partial\hat{w}_1}{\partial\hat{z}}. \tag{42.94d}$$

Write the dimensional form of these four equations, showing all steps when moving from the dimensionless to dimensional equations. For the horizontal velocity, write

$$\boldsymbol{u}_g = U\hat{\boldsymbol{u}}_0 \tag{42.95}$$

for that portion of the geostrophic flow that is horizontally non-divergent, and

$$\boldsymbol{u}_{ag} = U(\hat{\boldsymbol{u}} - \hat{\boldsymbol{u}}_0) \approx Ro\, U\, \hat{\boldsymbol{u}}_1 \tag{42.96a}$$
$$\varphi_{ag} = \Phi(\hat{\varphi} - \hat{\varphi}_0) \approx Ro\, \Phi\, \hat{\varphi}_1. \tag{42.96b}$$

for an estimate of the ageostrophic portion of the velocity and pressure.

(b) Write the dimensional buoyancy and vorticity equations using the geostrophic streamfunction. Introduce the Jacobian operator for the advection.

(c) Cross-multiply the dimensional buoyancy and vorticity equations to eliminate the time derivative, thus revealing a diagnostic equation for the vertical velocity that is valid to order Ro^1.

(d) The equation for the vertical velocity takes the form

$$\mathcal{L} w = \sigma, \tag{42.97}$$

where

$$\mathcal{L} = N^2 \nabla_z^2 + f_0^2 \frac{\partial^2}{\partial z^2} \tag{42.98}$$

is a linear partial differential operator and

$$\sigma = f_0 J(\psi, \zeta + \beta\, y) - \nabla^2 J(\psi, b) \tag{42.99}$$

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 3.3. 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)

$$\frac{\mathrm{D}u}{\mathrm{D}t} - fv = -\frac{\partial \varphi}{\partial x} \tag{42.100a}$$
$$\frac{\mathrm{D}v}{\mathrm{D}t} + fu = -\frac{\partial \varphi}{\partial y} \tag{42.100b}$$
$$\frac{\partial \varphi}{\partial z} = b \tag{42.100c}$$
$$\frac{\mathrm{D}b}{\mathrm{D}t} = 0 \tag{42.100d}$$
$$\nabla \cdot \boldsymbol{v} = \nabla_z \cdot \boldsymbol{u} + \frac{\partial w}{\partial z} = 0 \tag{42.100e}$$
$$\frac{\mathrm{D}}{\mathrm{D}t} = \frac{\partial}{\partial t} + \boldsymbol{v} \cdot \nabla. \tag{42.100f}$$

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

$$\boldsymbol{u} = \boldsymbol{u}_g + \boldsymbol{u}_{ag}, \tag{42.101}$$

with the geostrophic velocity balancing the horizontal gradient of the <u>full</u> pressure field

$$f\boldsymbol{u}_g = \hat{\boldsymbol{z}} \wedge \nabla\varphi. \tag{42.102}$$

This definition for $\boldsymbol{u}_g$ 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

$$\frac{\mathrm{D}\boldsymbol{u}}{\mathrm{D}t} + f\hat{\boldsymbol{z}} \wedge \boldsymbol{u}_{ag} = 0, \tag{42.103}$$

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

$$\frac{\partial \boldsymbol{u}_g}{\partial t} + (\boldsymbol{u}_g \cdot \nabla_z)\boldsymbol{u}_g + f\hat{\boldsymbol{z}} \wedge \boldsymbol{u}_{ag} = 0 \tag{42.104}$$

and the buoyancy equation equal to

$$\frac{\partial b}{\partial t} + \boldsymbol{u}_g \cdot \nabla_z b + N^2 w = 0, \tag{42.105}$$

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 (42.104) 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

$$\left[\frac{\partial}{\partial t} + \boldsymbol{u}_g \cdot \nabla_z\right] |\nabla_z b|^2 = \boldsymbol{Q} \cdot \nabla_z b. \tag{42.106}$$

Hence, horizontal buoyancy gradients grow in magnitude in regions where the horizontal buoyancy gradient projects positively onto the $\boldsymbol{Q}$-vector. Write the expression for the vector $\boldsymbol{Q}$. Hint: A general version of this result was derived in Exercise 12.1.

(b) Show that the quasi-geostrophic ω-equation on an f-plane can be written

$$N^2\nabla_z^2 w + f_0^2\frac{\partial^2 w}{\partial z^2} = 2\nabla_z\cdot\boldsymbol{Q}. \tag{42.107}$$

We see that the source for vertical motion is the divergence of the $\boldsymbol{Q}$-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 (42.107) and show each of the relevant steps.

(c) The equation (42.107) for the vertical velocity takes the form

$$\mathcal{L}w = \sigma, \tag{42.108}$$

where

$$\mathcal{L} = N^2\nabla_z^2 + f_0^2\frac{\partial^2}{\partial z^2} \tag{42.109}$$

is a linear partial differential operator and

$$\sigma = 2\nabla_z\cdot\boldsymbol{Q} \tag{42.110}$$

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 3.3. What about when $N^2 < 0$? What new physical phenomena do you expect when $N^2 < 0$?

Part VIII

Topics in ocean fluid mechanics

In this part of the book presents a suite of topics arising in studies of ocean dynamics. We build on much of the earlier material in this book, while specializing to the needs and interests of physical oceanography. After examining a variety of special topics, we develop the physics and mathematics of scalar tracer fields, including thermodynamic tracers such as potential temperature and material tracers such as salinity and humidity. This *tracer mechanics* offers a distinct view on the fluid flows and thus provides a useful complement to the mechanics of momentum, vorticity, and energy.

We devote Chapter 47 to exploring the physical and mathematical properties of advection and diffusion when acting on scalar fields. On the large-scales, the advection and diffusion felt by a tracer generally arises from more than just the mean flow and molecular diffusion. In Chapter 48 we introduce the notions of wave-mean flow interactions that give rise to eddy-induced advection (or skew diffusion) and diffusion. Throughout this part of the book, in particular in Chapter 47, 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.

43

Space-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 brief 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 increasing interest to climate science concerns the study of how the 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 (*Farrell and Clark* (1976) and *Mitrovica et al.* (2001)) 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 9.12. 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.

43.1 Gravitational potential **764**
43.1.1 Simple geopotential 764
43.1.2 General geopotential 764
43.2 Momentum equation **764**
43.3 Primitive equations **765**
43.4 Depth independent perturbed geopotential **765**
43.5 Forces contributing to ocean tides **766**
43.5.1 Tidal acceleration in a spherically symmetric gravity field 766
43.5.2 Heuristics of tidal acceleration on the surface of a sphere 767
43.5.3 Gravitational potential for an idealized earth-moon system 769
43.5.4 Concerning realistic tides 773
43.5.5 Comments 773

43.1 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.

43.1.1 Simple geopotential

As detailed in Section 9.12, 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 9.12.2),

$$\boldsymbol{g} = -\nabla \Phi, \tag{43.1}$$

with Φ the geopotential. In most applications of this book, the local vertical direction is denoted by

$$z = r - R_{\mathrm{e}}, \tag{43.2}$$

with $z = 0$ the geopotential surface corresponding to a resting ocean and $R_{\mathrm{e}} = 6.367 \times 10^6\,\mathrm{m}$ the average radius of the earth (Section 9.2). The geopotential in this case is given by

$$\Phi = \Phi_0 = g\, z, \tag{43.3}$$

with $g \approx 9.8\,\mathrm{m\,s^{-2}}$ the typical value used for the gravitational acceleration at the earth's surface.

43.1.2 General geopotential

Consider a generalized geopotential written in the form

$$\Phi = \Phi_0(r) + \Phi_1(r, \lambda, \phi, t), \tag{43.4}$$

where $\Phi_0(r)$ is the geopotential given by equation (43.3), 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 (43.4), 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.

43.2 Momentum equation

As detailed in Section 18.1.3, the inviscid momentum equation for a rotating fluid in a gravitational field is given by

$$\rho \frac{\mathrm{D}\boldsymbol{v}}{\mathrm{D}t} + 2\,\boldsymbol{\Omega} \wedge \rho\,\boldsymbol{v} = -\nabla p - \rho\,\nabla\Phi. \tag{43.5}$$

In writing the momentum equation in the form (43.5), 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.

43.3 Primitive equations

As detailed in Section 24.1, the primitive equations reduce the vertical momentum equation to its static inviscid form, which is the hydrostatic balance

$$\frac{\partial p}{\partial z} = -\rho\,\frac{\partial \Phi}{\partial z} = -\rho\,(g + \partial_z \Phi_1). \tag{43.6}$$

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 24.1) takes the form

$$\rho\,\frac{\mathrm{D}\boldsymbol{u}}{\mathrm{D}t} + \hat{\boldsymbol{z}}\,f \wedge \rho\,\boldsymbol{u} = -(\rho\,\nabla_z \Phi_1 + \nabla_z p) \tag{43.7}$$

where ∇_z is the horizontal gradient taken on surfaces of constant z. In their oceanic Boussinesq form (Chapter 25), the inviscid horizontal momentum equation becomes

$$\frac{\mathrm{D}\boldsymbol{u}}{\mathrm{D}t} + \hat{\boldsymbol{z}}\,f \wedge \boldsymbol{u} = -(1/\rho_0)\,(\rho_0\,\nabla_z \Phi_1 + \nabla_z p) \tag{43.8}$$

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.

43.4 Depth independent perturbed geopotential

A particularly simple form of Φ_1 occurs when it is depth independent,

$$\Phi_1 = \Phi_1(\lambda, \phi, t), \tag{43.9}$$

in which case the hydrostatic balance (43.6) returns to its traditional form $\partial_z p = -\rho\,g$. This perturbed geopotential is generally sufficient for the study of ocean tides. In this case it is convenient to write the geopotential as

$$\Phi_1 = -g\,h, \tag{43.10}$$

with $h = h(\lambda, \phi, t)$ the perturbed geopotential height field. The full geopotential is thus written

$$\Phi = g\,(z - h), \tag{43.11}$$

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

$$-\int_{\eta_b}^{\eta} \nabla_z \Phi_1 \, dz = g \int_{\eta_b}^{\eta} \nabla_z h \, dz = g \left(-\eta_b + \eta\right) \nabla_z h. \tag{43.12}$$

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.

43.5 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.

43.5.1 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 43.1 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 9.12.1)

$$\boldsymbol{g} = -\frac{GM}{r^2} \, \hat{\boldsymbol{r}}, \tag{43.13}$$

where r is the distance from the sphere's center, G is Newton's gravitational constant, M is the mass of the sphere, and $\hat{\boldsymbol{r}}$ 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 43.1, 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

$$\boldsymbol{g}(r_B) - \boldsymbol{g}(r_A) = \boldsymbol{g}(r_0 - L/2) - \boldsymbol{g}(r_0 + L/2), \tag{43.14}$$

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

$$\boldsymbol{g}(r_B) - \boldsymbol{g}(r_A) \approx -L \, \frac{\partial \boldsymbol{g}}{\partial r} = -2L \, \frac{GM}{r_0^3} \, \hat{\boldsymbol{r}} = (2\, L/r_0) \, \boldsymbol{g}(r_0). \tag{43.15}$$

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 43.5.3 the gravitational acceleration generated from a remote body (e.g., the moon) acting on the surface of a sphere (e.g., the earth).

Figure 43.1: *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.*

43.5.2 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 43.2 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 43.5.1, 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.

Figure 43.2: *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 43.1. The tidal acceleration acting at point B, relative to the earth's center, points toward the moon (equation (43.16)). In contrast, the tidal acceleration at point A, relative to the earth's center, points in the opposite direction (equation (43.17)). 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 43.5.3, the bulge shown in the right panel is greatly exaggerated.*

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 43.1, the tidal acceleration at point B relative to the center of the earth is given by

$$\boldsymbol{g}(r_B) - \boldsymbol{g}(R_{\rm em}) = (2R_{\rm e}/R_{\rm em})\,\boldsymbol{g}(R_{\rm em}). \tag{43.16}$$

This acceleration points towards the moon. In contrast, the tidal acceleration at point A relative to the center of the earth is given by

$$\boldsymbol{g}(r_A) - \boldsymbol{g}(R_{\rm em}) = -(2R_{\rm e}/R_{\rm em})\,\boldsymbol{g}(R_{\rm em}), \tag{43.17}$$

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 43.5.3 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 43.2. 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 43.5.3 by explicitly computing the gravitational potential for this idealized earth-moon system and then taking the gradient to compute the gravitational acceleration (see Figure 43.4).

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 $\boldsymbol{g}(R_{\rm em})$, which has magnitude $GM_{\rm m}/R_{\rm em}^2$ 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

$$\boldsymbol{a}_{\text{orbital centrifugal}} = -\boldsymbol{g}(R_{\rm em}). \tag{43.18}$$

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 9.12). 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 (43.16) and (43.17). We are thus led to the same result as before.

43.5.3 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 9.12.1, whereby the gravitational potential for a point at distance r from the center of a spherical earth is given by

$$\Phi_{\rm e}(r) = -\frac{GM_{\rm e}}{r}, \tag{43.19}$$

where $M_{\rm e}$ is the mass of the earth. The corresponding radial gravitational acceleration is given by

$$\boldsymbol{g}_{\rm e} = -\nabla\Phi_{\rm e} = -\frac{GM_{\rm e}\,\hat{\boldsymbol{r}}}{r^2}. \tag{43.20}$$

The same considerations hold for the moon's gravitational potential. Hence, referring to Figure 43.3, the moon's gravitational potential evaluated at a distance L from the moon's center is given by

$$\Phi_{\rm m}(L) = -\frac{GM_{\rm m}}{L}. \tag{43.21}$$

Trigonometry leads to the law of cosines relation

$$L^2 = (R_{\rm em} - r\cos\psi)^2 + (r\sin\psi)^2 = R_{\rm em}^2 + r^2 - 2\,r\,R_{\rm em}\cos\psi, \tag{43.22}$$

where again r is the distance to the earth's center and ψ is the polar angle relative to the $\hat{\boldsymbol{x}}$ axis pointing between the earth and moon centers (see Figure 43.3).

Figure 43.3: *Geometry of an idealized earth-moon system. The center of the earth is a distance $R_{\rm em}$ from the center of the moon; the moon has a mass $M_{\rm m}$; and the earth has a radius $R_{\rm e}$. 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 $\hat{\boldsymbol{x}}$ axis, where the $\hat{\boldsymbol{x}}$ 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 7.3 for details on relating polar and Cartesian coordinates.*

Identifying the leading order contributions

Assuming the test point in Figure 43.3 is closer to the earth than to the moon, we can perform a Taylor series expansion in the small parameter $r/R_{\rm em}$ to render

$$\Phi_{\rm m}(L) = -\frac{GM_{\rm m}}{L} = -\frac{GM_{\rm m}}{R_{\rm em}}\left[1 + \frac{r\cos\psi}{R_{\rm em}} + \frac{r^2}{2R_{\rm em}^2}\,(3\cos^2\psi - 1) + \mathcal{O}(r/R_{\rm em})^3\right]. \tag{43.23}$$

We thus identify the leading three terms to the geopotential

$$\Phi_{\rm m}^{(0)} = -\frac{GM_{\rm m}}{R_{\rm em}} \tag{43.24}$$

$$\Phi_{\rm m}^{(1)} = -\frac{GM_{\rm m}}{R_{\rm em}^2}\,r\cos\psi \tag{43.25}$$

$$\Phi_{\rm m}^{(2)} = -\frac{GM_{\rm m}}{2R_{\rm em}^3}\,r^2\,(3\cos^2\psi - 1). \tag{43.26}$$

Assuming the distance between the earth and moon remains fixed, the zeroth order term $\Phi_{\rm m}^{(0)}$ 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, $\Phi_{\rm m}^{(1)}$, we introduce the Cartesian coordinate as in Figure 43.3 to write

$$\Phi_{\rm m}^{(1)} = -\frac{GM_{\rm m}\,x}{R_{\rm em}^2}, \tag{43.27}$$

where $x = r\cos\psi$ is the distance along $\hat{\boldsymbol{x}}$. Hence, the gradient of $\Phi_{\rm m}^{(1)}$ leads to the gravitational acceleration

$$\boldsymbol{g}_{\rm m}^{(1)} = -\nabla\Phi_{\rm m}^{(1)} = \hat{\boldsymbol{x}}\,\frac{GM_{\rm m}}{R_{\rm em}^2}. \tag{43.28}$$

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 $\boldsymbol{g}_{\rm m}^{(1)}$ equals to that of the moon's gravitational acceleration, $\boldsymbol{g}_{\rm m}$, when evaluated at the earth's center. As seen in Section 43.5.2, the acceleration $\boldsymbol{g}_{\rm m}^{(1)}$ 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 $\boldsymbol{g}_{\rm m}^{(1)}$ is tiny relative to the gravitational acceleration from the earth itself, with their ratios given by

$$\frac{M_{\rm m}/R_{\rm em}^2}{M_{\rm e}/R_{\rm e}^2} \approx 3.4\times10^{-6}, \tag{43.29}$$

where we set

$$M_{\rm e} = 5.97\times10^{24}\ {\rm kg} \qquad M_{\rm m} = 7.35\times10^{22}\ {\rm kg} = (1/81.2)\,M_{\rm e} \tag{43.30a}$$

$$R_{\rm e} = 6.367\times10^{6}\ {\rm m} \qquad R_{\rm em} = 3.84\times10^{8}\ {\rm m} = 60.3\,R_{\rm e}. \tag{43.30b}$$

Figure 43.4: *The tide producing gravitational acceleration $\boldsymbol{g}_{\rm m}^{(2)}$ given by equation (43.34). The moon is assumed to be positioned in the equatorial plane of the earth.*

Tide producing geopotential

The main tide producing acceleration results from $\Phi_{\rm m}^{(2)}$. Introducing the second Cartesian coordinate, $y = r\sin\psi$, leads to

$$\Phi_{\rm m}^{(2)} = -\frac{GM_{\rm m}}{2R_{\rm em}^3}\,r^2\,(3\cos^2\psi - 1) = -\frac{GM_{\rm m}}{2R_{\rm em}^3}\,(2\,x^2 - y^2). \tag{43.31}$$

The corresponding perturbed geopotential height field (see equation (43.11)) is given by

$$h = -\frac{\Phi_{\rm m}^{(2)}}{g} = \frac{R_{\rm e}^2}{2\,R_{\rm em}^3}\,\frac{M_{\rm m}}{M_{\rm e}}\,r^2\,(3\cos^2\psi - 1). \tag{43.32}$$

Placing the test point on the earth surface, $r = R_{\rm e}$, renders

$$h = \frac{R_{\rm e}^4}{2\,R_{\rm em}^3}\,\frac{M_{\rm m}}{M_{\rm e}}\,(3\cos^2\psi - 1) \approx 2.8\times 10^{-8}\,R_{\rm e}\,(3\cos^2\psi - 1). \tag{43.33}$$

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 43.2 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

$$\boldsymbol{g}_{\rm m}^{(2)} = -\nabla\Phi_{\rm m}^{(2)} = \frac{G M_{\rm m}}{R_{\rm em}^3}\,(2\,x\,\hat{\boldsymbol{x}} - y\,\hat{\boldsymbol{y}}). \tag{43.34}$$

We illustrate the vector field $\boldsymbol{g}_{\rm m}^{(2)}$ in Figure 43.4. 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 $\hat{\boldsymbol{r}}$ and $\hat{\boldsymbol{\psi}}$ according to Section 7.3.2

$$\hat{\boldsymbol{r}} = \hat{\boldsymbol{x}}\cos\psi + \hat{\boldsymbol{y}}\sin\psi \tag{43.35a}$$

$$\hat{\boldsymbol{\psi}} = -\hat{\boldsymbol{x}}\sin\psi + \hat{\boldsymbol{y}}\cos\psi \tag{43.35b}$$

thus rendering

$$\boldsymbol{g}_{\rm m}^{(2)} = \frac{G M_{\rm m} R_{\rm e}}{R_{\rm em}^3}\left[\hat{\boldsymbol{r}}\,(3\cos^2\psi - 1) - (3/2)\,\hat{\boldsymbol{\psi}}\,\sin 2\psi\right], \tag{43.36}$$

where we evaluated the acceleration at the earth surface so that $r = R_{\rm e}$. Evaluating the acceleration at $\psi = 0, \pi$ verifies the heuristic calculation performed in Section 43.5.2 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

$$\boldsymbol{g}_{\rm m}^{(2)} = g_{\rm e}\frac{M_{\rm m}}{M_{\rm e}}\,\frac{R_{\rm e}^3}{R_{\rm em}^3}\left[\hat{\boldsymbol{r}}\,(3\cos^2\psi - 1) - (3/2)\,\hat{\boldsymbol{\psi}}\,\sin 2\psi\right], \tag{43.37}$$

where $g_{\rm e} = G\,M_{\rm e}/R_{\rm e}^2$ is the acceleration at the earth's surface from the earth's gravity field. The dimensional prefactor has magnitude $\approx 5.6\times 10^{-8}\,g_{\rm e}$, 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 43.4, thus leading to tidal motion.

43.5.4 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 $28.5°N$ 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

$$\frac{\text{moon tidal acceleration}}{\text{sun tidal acceleration}} = \frac{M_{\mathrm{m}}/R_{\mathrm{em}}^3}{M_{\mathrm{s}}/R_{\mathrm{es}}^3} \approx 2.2 \tag{43.38}$$

  where we set

$$M_{\mathrm{s}} = 1.99 \times 10^{30} \text{ kg} \qquad R_{\mathrm{es}} = 23460\, R_{\mathrm{e}}. \tag{43.39}$$

  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 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. Since we have incomplete information about that geometry, the best predictions for tides are generally 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.

43.5.5 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 43.2), 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 43.5.2, 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.

44

Surface tension

Surface tension is present on surfaces that separate two immiscible liquids or between a liquid and gas. It has many consequences familiar from nature, 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 well outside of our scope. Instead, we develop some heuristics sufficient to determine when one needs to be concerned with surface tension in the study of geophysical fluid mechanics.

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 dynamics, where most of the focus concerns much 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

This chapter makes use of the notions of stress detailed in Chapter 19, as well as curvature described in Chapter 4. We make use of Cartesian tensors as discussed in Chapters 1 and 2. Results can be readily generalized for arbitrary coordinates through the rules of general covariance detailed in Chapter 5. The material in this chapter has no direct impact on other chapters.

44.1 A container of water **775**
44.2 Force balance on an air-water interface **776**
44.3 Young-Laplace formula **778**
44.4 Some oceanographic examples **780**
44.4.1 Soluble gas bubbles inside water 780
44.4.2 Length scale for capillary waves 780
44.5 Further study **780**

44.1 A container of water

Atmospheric pressure at the earth's surface is roughly $p_{\text{atm}} = 10^5$ N m^{-2}. As we saw in Section 19.6, 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, $\rho\, g\, h < p_{\rm atm}$ and turn the container upside-down as in Figure 44.1. 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.

Figure 44.1: *A container of water with density ρ and height h is placed upside-down. Atmospheric pressure, $p_{\rm atm}$, will support water with thickness $h < p_{\rm atm}/(\rho\, g) \approx 10$ 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 31.2) 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 onto the floor.*

44.2 Force balance on an air-water interface

Consider two immiscible 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.[1] 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 44.2. 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.

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

[1] As discussed in Section 11.1, the mean free path is a statistical measure of the distance a molecule moves before hitting another molecule.

Figure 44.2: *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.*

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

$$\boldsymbol{f}_{\text{interface}} = -\gamma\, \hat{\boldsymbol{n}} \wedge \delta \boldsymbol{x}, \tag{44.1}$$

where $\hat{\boldsymbol{n}}$ is a normal vector pointing towards the center of the curved interface, and $\delta\boldsymbol{x}$ is a line element oriented so that the normal $\hat{\boldsymbol{n}}$ points to the left facing in the direction of the line increment. Figure 44.3 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 $\gamma\, \mathrm{d}\mathcal{S}$ as the work (units of N m^{-1} = Joule) required to create an area, $\mathrm{d}\mathcal{S}$, on the interface. We make use of this energetic perspective in Section 44.3.

To develop an expression for the pressure jump across the liquid-gas interface, consider a spherical droplet of radius R shown in Figure 44.3 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

$$\boldsymbol{F}_{\text{circle}} = \oint_{\text{circle}} \boldsymbol{f}_{\text{interface}} = -\oint_{\text{circle}} \gamma\, \hat{\boldsymbol{n}} \wedge \delta\boldsymbol{x} = -2\pi R\, \gamma\, \hat{\boldsymbol{z}}. \tag{44.2}$$

Equilibrium of the spherical droplet is realized by a pressure jump across the circular cross-sectional area

$$\pi R^2 \left(p_{\text{in}} - p_{\text{out}}\right) = 2\pi R\, \gamma \Longrightarrow \left(p_{\text{in}} - p_{\text{out}}\right) = 2\gamma/R. \tag{44.3}$$

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 (44.3) is a special case of the Young-Laplace formula, specialized here to a sphere.

Figure 44.3: *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 $2\pi R\gamma = \pi R^2 (p_{in} - p_{out})$, leading to a pressure jump across the droplet interface $p_{in} - p_{out} = 2\gamma/R$.*

44.3 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 44.4 that represents the interface seperating fluid-A from fluid-B, with $\hat{\boldsymbol{n}}$ 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 $\hat{\boldsymbol{n}}\,\delta h$ 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 $\mathrm{d}A$ is given by $\delta h\,\mathrm{d}A$, with this volume realized by applying the pressure work to the surface

$$W_{\text{volume}} = (p_{\text{B}} - p_{\text{A}})\,\delta h\,\mathrm{d}A. \tag{44.4}$$

For example, if $p_{\text{B}} > p_{\text{A}}$ 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.

In the presence of surface tension, work must overcome the surface area energy in order to change to the interface area

$$W_{\text{area}} = \gamma\,\delta A, \tag{44.5}$$

where δA is the change in area of an infinitesimal element on the interface

$$\delta A = \mathrm{d}\mathcal{S} - \mathrm{d}A \tag{44.6a}$$

$$= \mathrm{d}A\left[\sqrt{1 + (\nabla\delta h)^2} - 1\right] \tag{44.6b}$$

$$\approx \mathrm{d}A\,(\nabla\delta h)^2/2. \tag{44.6c}$$

To reach this result we made use of equation (4.29) that relates the area of an infinitesimal element on a curved surface to the area of its horizontal projection (see Section 4.3.1). We next make use

Figure 44.4: *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, $\hat{\boldsymbol{n}}$, 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.*

of the surface curvature detailed in Section 4.3.2, where equation (4.34) shows that the vertical displacement is given, for small displacements, by

$$\delta h \approx \frac{1}{2} R_1^{-1} (\boldsymbol{x} \cdot \boldsymbol{e}_1)^2 + \frac{1}{2} R_2^{-1} (\boldsymbol{x} \cdot \boldsymbol{e}_2)^2. \tag{44.7}$$

R_1^{-1}, R_2^{-1} are the eigenvalues and $\boldsymbol{e}_1, \boldsymbol{e}_2$ 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

$$(\nabla \delta h)^2 \approx (x/R_1)^2 + (y/R_2)^2 = \delta h \left[\frac{1}{R_1} + \frac{1}{R_2} \right], \tag{44.8}$$

where we set

$$\delta h / R_1 = (x/R_1)^2 \qquad \text{and} \qquad \delta h / R_2 = (y/R_2)^2. \tag{44.9}$$

We are thus led to the area difference

$$\delta A \approx \mathrm{d}A \, \delta h \left[\frac{1}{R_1} + \frac{1}{R_2} \right]. \tag{44.10}$$

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

$$W_{\text{area}} + W_{\text{volume}} = \mathrm{d}A \, \delta h \, \left[\gamma \, (R_1^{-1} + R_2^{-2}) + p_{\text{B}} - p_{\text{A}} \right] \tag{44.11}$$

and equilibrium results if the work vanishes

$$p_{\text{A}} - p_{\text{B}} = \gamma \, (R_1^{-1} + R_2^{-2}). \tag{44.12}$$

This equation is the Young-Laplace formula, which reduces to equation (44.3) 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 44.4 or the inside of a bubble/droplet.

44.4 Some oceanographic examples

We close this chapter with examples relevant to the ocean.

44.4.1 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}

$$p_{\text{in}} - p_{\text{out}} = 2\,\gamma/R \approx 144 \times 10^3 \text{ N m}^{-2} = 1.42\,p_{\text{atm}}, \tag{44.13}$$

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.

44.4.2 Length scale for capillary waves

Capillary waves arise 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

$$\omega^2 = k\,g\left[1 + \frac{k^2\,\gamma}{g\,\rho}\right], \tag{44.14}$$

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

$$\lambda_{\text{cap-grav}} = 2\pi\sqrt{\frac{\gamma}{\rho\,g}}, \tag{44.15}$$

with $\lambda < \lambda_{\text{cap-grav}}$ the capillary wave regime and $\lambda > \lambda_{\text{cap-grav}}$ the gravity wave regime. Using $\gamma = 0.072$ N m^{-1} and $\rho = 1000$ kg m^{-3} leads to

$$\lambda_{\text{cap-grav}} = 2\pi\sqrt{\frac{\gamma}{\rho\,g}} \approx 0.017 \text{ m} = 17 \text{ cm}. \tag{44.16}$$

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.

44.5 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 Section 44.1 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 44.4.1 is taken from Section 1.3 of *Kundu et al.* (2012). Section 4.10 of *Kundu et al.* (2012) 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.

45

Surface gravity waves and Stokes drift

In this chapter we consider the ocean surface to be a material interface and examine its small amplitude fluctuations. These fluctuations manifest as linear surface gravity waves whose existence relies on the restoring force from gravity. They provide the prototypical wave-like response that offer an ideal physical system to introduce linear wave behavior. In formulating the equations for surface gravity waves, we bring together a number of fluid mechanical principles considered earlier in this book. In so doing, this analysis supports our understanding of how basic principles are used to describe motion. Notably, we here ignore rotation, although many properties illustrated here well approximate surface waves found in the ocean.

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 24. 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 12. Generalizations of Stokes drift appear in Chapter 48 in our study of wave-mean flow interactions, isopycnal averaging, and the corresponding eddy-induced tracer transport. Finally, we ignore surface tension in this chapter given that our interests are for wave motions with lengths far larger than the capillary wavelengths (a few centimeters) discussed in Section 44.4.2.

45.1 Surface gravity waves **784**
45.1.1 Harmonic scalar potential 784
45.1.2 Equation of motion and Bernoulli's Principle 784
45.1.3 Non-hydrostatic pressure and the shallow water limit 786

45.1.4 Dynamic boundary condition at the free surface 787
45.1.5 Kinematic boundary conditions 787
45.1.6 Summary of the linear equation set 788
45.2 Surface gravity waves in a flat bottom channel **788**
45.2.1 Monochromatic wave solution 789
45.2.2 Dispersion relation for surface gravity waves 789
45.2.3 Heuristics of waves in the shortwave/deep water limit 790
45.2.4 Heuristics of waves in the longwave/shallow water limit 790
45.2.5 Further study 791
45.3 Stokes drift **791**
45.3.1 General formulation of Stokes drift 791
45.3.2 Stokes drift in surface gravity waves 793
45.3.3 Comments and further study 796
45.4 Exercises **797**

45.1 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 45.1). 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.

45.1.1 Harmonic scalar potential

We are interested in fluctuations that have zero vorticity so that the velocity field has zero curl[1]

$$\nabla \wedge \boldsymbol{v} = 0. \tag{45.1}$$

Recall from Section 2.3.2 that vector fields with zero curl can be determined from a scalar potential

$$\boldsymbol{v} = -\nabla\psi. \tag{45.2}$$

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 45.1.2.

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)

$$\nabla \cdot \boldsymbol{v} = -\nabla \cdot \nabla\psi = \nabla^2\psi = 0. \tag{45.3}$$

ψ is termed a *harmonic function*, with salient properties noted in Section 3.4. To fully specify the scalar potential requires boundary conditions, which enter our development via the equation of motion.

45.1.2 Equation of motion and Bernoulli's Principle

The vector-invariant equation of motion (18.27) for a non-rotating, irrotational, inviscid, uniform density fluid is given by

$$\frac{\partial \boldsymbol{v}}{\partial t} = -\nabla(\Phi + \mathcal{K} + p/\rho), \tag{45.4}$$

[1] We study fluids with nonzero vorticity in Part VI of this book.

Figure 45.1: *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, $\boldsymbol{g} = -g\,\hat{\boldsymbol{z}}$. 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 31.1.*

with

$$\mathcal{K} = \frac{\boldsymbol{v} \cdot \boldsymbol{v}}{2} \tag{45.5}$$

the kinetic energy per mass of a fluid element. In the following we assume the simple form of the geopotential (Section 9.12.2)

$$\Phi = g\,z \tag{45.6}$$

with g the constant gravitational acceleration. Inserting the scalar potential, $\boldsymbol{v} = -\nabla\psi$, brings the equation of motion (45.4) to the form

$$\nabla\left(\Phi + \mathcal{K} + p/\rho - \partial_t\psi\right) = 0. \tag{45.7}$$

This equation means that everywhere in the fluid the dynamical fields satisfy

$$\Phi + \mathcal{K} + p/\rho - \partial_t\psi = C(t), \tag{45.8}$$

for some arbitrary time dependent function $C(t)$. This equation is a particular expression of Bernoulli's theorem encountered in Section 23.4.4.

Removing the time dependent function via a gauge transformation

We ascribe no physical meaning to the arbitrary function $C(t)$ appearing in equation (45.8). In fact, it can be completely removed by exploiting the gauge degree of freedom in the scalar potential as noted following equation (45.2). We do so by introducing a modified scalar potential

$$\Psi(\boldsymbol{x}, t) = \psi(\boldsymbol{x}, t) + \int^t C(t')\,\mathrm{d}t'. \tag{45.9}$$

Both ψ and Ψ lead to the same velocity vector

$$\boldsymbol{v} = -\nabla\psi = -\nabla\Psi \tag{45.10}$$

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

$$\Phi + \mathcal{K} + p/\rho - \partial_t \Psi = 0. \tag{45.11}$$

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 (45.11) leads to

$$\Phi + \mathcal{K} + p/\rho = 0. \tag{45.12}$$

Furthermore, assuming motion occurs on a constant geopotential then leads to

$$p + \rho \mathcal{K} = \text{constant}. \tag{45.13}$$

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 (45.13). 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.

45.1.3 Non-hydrostatic pressure and the shallow water limit

The fluid is non-hydrostatic since

$$\frac{\partial p}{\partial z} = -\rho \frac{\partial \Phi}{\partial z} + \frac{\partial(\partial_t \Psi - \mathcal{K})}{\partial z} = \frac{\partial p_{\text{hydrostatic}}}{\partial z} + \frac{\partial p_{\text{non-hydrostatic}}}{\partial z}. \tag{45.14}$$

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 26.4.3), 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 29, the hydrostatic balance is fundamental to the dynamics. In particular, in Section 29.1 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 29.1.5) and its magnitude is far smaller than horizontal motions. Therfore, we can drop all contributions to $\partial_z \mathcal{K}$ for the shallow water layer, in which case equation (45.14) reduces to the hydrostatic limit

$$\frac{\partial p}{\partial z} = -\rho \frac{\partial \Phi}{\partial z} = -\rho\, g \qquad \text{hydrostatic limit.} \tag{45.15}$$

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

$$\partial_z \left(\nabla_z p_{\text{hydrostatic}} \right) = -g \, \nabla_z \rho = 0. \tag{45.16}$$

45.1.4 Dynamic boundary condition at the free surface

Return now to the time dependent expression of the equation of motion (45.11). 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

$$g \, \eta + \mathcal{K} - \partial_t \Psi = -p_{\text{a}}/\rho = \text{constant}. \tag{45.17}$$

Without loss of generality we can set this constant to zero.[2] We are thus left with the boundary condition

$$g \, \eta + \mathcal{K} - \partial_t \Psi = 0 \qquad \text{at } z = \eta. \tag{45.18}$$

At this point we linearize relative to a state of rest with $\eta = 0$, $\boldsymbol{v} = 0$, 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

$$g \, \eta = \frac{\partial \Psi}{\partial t} \qquad \text{linearized dynamic b.c. at } z = \eta. \tag{45.19}$$

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.

45.1.5 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 14.4.2, we have

$$\frac{\partial \eta}{\partial t} + \boldsymbol{u} \cdot \nabla \eta = w \qquad \text{at } z = \eta. \tag{45.20}$$

Linearizing this boundary condition about the state of rest, and introducing the scalar potential, leads to

$$\frac{\partial \eta}{\partial t} = -\frac{\partial \Psi}{\partial z} \qquad \text{linearized kinematic b.c. at } z = \eta. \tag{45.21}$$

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 14.4.1 means that

$$\hat{\boldsymbol{n}} \cdot \nabla \Psi = 0 \qquad \boldsymbol{x} \text{ on rigid solid boundaries}, \tag{45.22}$$

where $\hat{\boldsymbol{n}}$ is the outward normal on the solid boundaries.

[2] Alternatively, make use of a further gauge transformation $\Psi' = \Psi - t \, (p_{\text{a}}/\rho)$ to eliminate the constant.

45.1.6 Summary of the linear equation set

The boundary value problem for the velocity potential and free surface is given by

$$\nabla^2 \Psi = 0 \qquad \text{irrotational and non-divergent velocity throughout domain} \tag{45.23a}$$

$$\frac{\partial \Psi}{\partial t} = g\,\eta \qquad \text{linearized dynamic b.c. at } z = \eta \tag{45.23b}$$

$$\frac{\partial \Psi}{\partial z} = -\frac{\partial \eta}{\partial t} \qquad \text{linearized kinematic b.c. at } z = \eta \tag{45.23c}$$

$$\hat{\boldsymbol{n}} \cdot \nabla \Psi = 0 \qquad \text{no-normal flow kinematic b.c. on rigid boundaries.} \tag{45.23d}$$

The first equation holds throughout the fluid whereas the remaining equations hold only at the boundaries.

Although the equations (45.23a)-(45.23d) 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 (45.23b) according to

$$g\,\frac{\partial \eta}{\partial t} = \left[\frac{\partial}{\partial t} + \frac{\partial \eta}{\partial t}\,\frac{\partial}{\partial z}\right]\frac{\partial \Psi}{\partial t} \tag{45.24}$$

and then combine with equation (45.23c) to render

$$\left[\frac{\partial}{\partial t} + \frac{\partial \eta}{\partial t}\,\frac{\partial}{\partial z}\right]\frac{\partial \Psi}{\partial t} = -g\,\frac{\partial \Psi}{\partial z}. \tag{45.25}$$

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

$$|\eta|/\lambda \ll 1. \tag{45.26}$$

In summary, the fully linearized equation set takes the form

$$\nabla^2 \Psi = 0 \qquad \text{irrotational and non-divergent velocity throughout domain} \tag{45.27a}$$

$$g\,\eta = \frac{\partial \Psi}{\partial t} \qquad \text{linearized dynamic b.c. at } z = 0 \tag{45.27b}$$

$$\frac{\partial \eta}{\partial t} = -\frac{\partial \Psi}{\partial z} \qquad \text{linearized kinematic b.c. at } z = 0 \tag{45.27c}$$

$$\hat{\boldsymbol{n}} \cdot \nabla \Psi = 0 \qquad \text{no-normal flow kinematic b.c. on rigid boundaries.} \tag{45.27d}$$

45.2 Surface gravity waves in a flat bottom channel

We now derive a wave-like solution to the equations (45.27a)-(45.27d) posed in a flat bottom channel domain as illustrated in Figure 45.1. 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.

45.2.1 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

$$\Psi(x, y, z, t) = \Gamma(z) \cos(k\, x - \omega\, t). \tag{45.28}$$

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 ansätz (45.28) 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

$$\frac{\mathrm{d}^2\Gamma}{\mathrm{d}z^2} = k^2\,\Gamma \qquad -H \le z \le 0 \tag{45.29a}$$

$$\frac{\mathrm{d}\Gamma}{\mathrm{d}z} = 0 \qquad \text{at } z = -H, \tag{45.29b}$$

where the bottom boundary condition is required to satisfy the no-normal flow condition (45.27d). We write the solution in the form

$$\Psi = \Psi_o \cosh[k\,(z + H)]\, \cos(k\, x - \omega\, t), \tag{45.30a}$$

$$\Psi_o = \frac{g\,\eta_o/\omega}{\cosh(k\,H)} \tag{45.30b}$$

so that the boundary condition (45.27b) renders the free surface height

$$\eta(x, t) = -\eta_o \sin(k\, x - \omega\, t). \tag{45.31}$$

The corresponding velocity field, $\boldsymbol{v} = -\nabla\Psi$, is given by

$$u = -\Psi_o\, k \cosh[k\,(z + H)]\, \sin(k\, x - \omega\, t) \tag{45.32a}$$

$$w = -\Psi_o\, k \sinh[k\,(z + H)]\, \cos(k\, x - \omega\, t). \tag{45.32b}$$

45.2.2 Dispersion relation for surface gravity waves

Combining the two $z = 0$ boundary conditions (45.27b) and (45.27c) yields

$$\frac{\partial^2\Psi}{\partial t^2} + g\,\frac{\partial\Psi}{\partial z} = 0 \qquad \mathrm{z} = 0. \tag{45.33}$$

Substituting the wave solution (45.30a) into this relation leads to the constraint

$$\omega^2 = g\, k \tanh(k\, H). \tag{45.34}$$

Figure 45.2: *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)).*

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 (45.34). The corresponding phase speed for the wave is given by

$$c = \frac{\omega}{k} = \sqrt{(g/k)\,\tanh(k\,H)}. \tag{45.35}$$

We emphasize the two limits: $k\,H \gg 1$ (shortwaves/deep water) and $k\,H \ll 1$ (longwave/shallow water), in which the phase speed satisfies

$$\omega \approx \sqrt{g\,k} \qquad c \approx \sqrt{g/k} \qquad k\,H \gg 1 \qquad \text{shortwave/deep water limit} \tag{45.36a}$$

$$\omega \approx k\,\sqrt{g\,H} \qquad c \approx \sqrt{g\,H} \qquad k\,H \ll 1 \qquad \text{longwave/shallow water limit.} \tag{45.36b}$$

45.2.3 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.

45.2.4 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 $\sqrt{g\,H}$. Tsunamis are the prototypical shallow water waves that travel at speeds well approximated by $\sqrt{g\,H}$.

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

Figure 45.3: *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.*

trough and, at some point, the assumptions of linearity breakdown and the shallow water waves break on the shore as depicted in Figure 45.3.

45.2.5 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 a decades-long problem in fluid mechanics that stretches the limits of theory, experiment, and simulation. *Bühler* (2014) 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.

45.3 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 48.

45.3.1 General formulation of Stokes drift

Consider a three-dimensional particle trajectory written in Cartesian coordinates,

$$\mathbf{X}(\mathbf{a}, t) = X(\mathbf{a}, t)\,\hat{\mathbf{x}} + Y(\mathbf{a}, t)\,\hat{\mathbf{y}} + Z(\mathbf{a}, t)\,\hat{\mathbf{z}}. \tag{45.37}$$

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 12.8.1, the particle trajectory is determined by time integrating the particle velocity (also known as the "flow map")

$$\frac{\partial \boldsymbol{X}(\boldsymbol{a},t)}{\partial t} = \mathbf{v}[\mathbf{X}(\mathbf{a},t),t] \tag{45.38}$$

so that

$$\mathbf{X}(\mathbf{a},t) = \mathbf{X}(\boldsymbol{a},0) + \int_0^t \mathbf{v}[\mathbf{X}(\mathbf{a},t'),t']\,\mathrm{d}t'. \tag{45.39}$$

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 $\boldsymbol{a}$ is given by the initial position of the particle, $\mathbf{X}(\boldsymbol{a},0)$, 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

$$v^n[\mathbf{X}(\mathbf{a},t),t] \approx v^n[\mathbf{X}(\mathbf{a},0),t] + \nabla v^n[\mathbf{X}(\mathbf{a},0),t]\cdot[\mathbf{X}(\mathbf{a},t)-\mathbf{X}(\mathbf{a},0)] \tag{45.40a}$$

$$= v^n[\mathbf{X}(\mathbf{a},0),t] + \nabla v^n[\mathbf{X}(\mathbf{a},0),t]\cdot\int_0^t \frac{\mathrm{d}\mathbf{X}(\mathbf{a},t')}{\mathrm{d}t'}\,\mathrm{d}t' \tag{45.40b}$$

$$= v^n[\mathbf{X}(\mathbf{a},0),t] + \nabla v^n[\mathbf{X}(\mathbf{a},0),t]\cdot\int_0^t \boldsymbol{v}[\mathbf{X}(\mathbf{a},t'),t']\,\mathrm{d}t', \tag{45.40c}$$

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 (45.40c).

- 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 accuracty 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

$$\epsilon = \frac{|\mathbf{X}(\mathbf{a},t)-\mathbf{X}(\mathbf{a},0)|}{\Lambda}. \tag{45.41}$$

  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 (45.40c) 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

$$v^n[\mathbf{X}(\mathbf{a},t),t] \approx v^n[\mathbf{X}(\mathbf{a},0),t] + \nabla v^n[\mathbf{X}(\mathbf{a},0),t]\cdot\int_0^t \boldsymbol{v}[\mathbf{X}(\mathbf{a},0),t']\,\mathrm{d}t', \tag{45.42}$$

with rearrangement leading to

$$v^n[\mathbf{X}(\mathbf{a},t),t] - v^n[\mathbf{X}(\mathbf{a},0),t] \approx \nabla v^n[\mathbf{X}(\mathbf{a},0),t]\cdot\int_0^t \boldsymbol{v}[\mathbf{X}(\mathbf{a},0),t']\,\mathrm{d}t'. \tag{45.43}$$

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 (45.43) 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 (45.43) when time averaged over a wave period, which we write as

$$v^n_{(S)}[\mathbf{X}(\mathbf{a}, 0), t] = \overline{v^n[\mathbf{X}(\mathbf{a}, t), t] - v^n[\mathbf{X}(\mathbf{a}, 0), t]} \tag{45.44a}$$

$$\approx \overline{\nabla v^n[\mathbf{X}(\mathbf{a}, 0), t] \cdot \int_0^t \boldsymbol{v}[\mathbf{X}(\mathbf{a}, 0), t'] \, \mathrm{d}t'}. \tag{45.44b}$$

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

$$v^n_{(S)}(\boldsymbol{x}, t) \approx \overline{\nabla v^n(\boldsymbol{x}, t) \cdot \int_0^t \boldsymbol{v}(\boldsymbol{x}, t') \, \mathrm{d}t'}. \tag{45.45}$$

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.

45.3.2 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 45.1 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

$$\frac{\mathrm{d}X}{\mathrm{d}t} = U \sin(\omega t) \tag{45.46a}$$

$$\frac{\mathrm{d}Z}{\mathrm{d}t} = U \cos(\omega t), \tag{45.46b}$$

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 $k\,x - \omega\,t$ phase from Section 45.2. Figure 45.4 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

$$X(t) - X_o = -\frac{U_o}{\omega} \left[\cos(\omega t) - 1\right] \tag{45.47a}$$

$$Z(t) - Z_o = \frac{U_o}{\omega} \sin(\omega t), \tag{45.47b}$$

Figure 45.4: *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.*

where the initial position at time $t = 0$ is

$$\mathbf{X}(t = t_o) = \mathbf{X}_o, \tag{45.48}$$

and the center of the circle is

$$\mathbf{X}_{\text{center}} = \left[X_o + \frac{U_o}{\omega}\right] \hat{\mathbf{x}} + Z_o \, \hat{\mathbf{z}}. \tag{45.49}$$

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 45.2. 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

$$U \approx U_o + R\,(Z - Z_o) \tag{45.50}$$

where R has units of inverse time and is given by

$$R = \left[\frac{\mathrm{d}U}{\mathrm{d}Z}\right]_{Z=Z_o}. \tag{45.51}$$

For the surface gravity waves with velocity components (45.32a) and (45.32b), we have

$$U = \Psi_o \, k \, \sqrt{\cosh[2\,k\,(z + H)]} \tag{45.52}$$

so that with $Z_o = 0$,

$$R = \Psi_o \, k^2 \, \frac{\sinh(2\,k\,H)}{\sqrt{\cosh(2\,k\,H)}}. \tag{45.53}$$

In general, the Taylor series is valid so long as the vertical trajectories maintain the inequality

$$|R|\,|Z - Z_o| \ll U_o, \tag{45.54}$$

which says that the vertical shear is small

$$|R| \ll \frac{U_o}{|Z - Z_o|}. \tag{45.55}$$

We use the Taylor series expansion (45.50) to solve for the vertical trajectory as determined by

$$\frac{\mathrm{d}(Z - Z_o)}{\mathrm{d}t} = [U_o + R\,(Z - Z_o)]\cos(\omega t). \tag{45.56}$$

Rearrangement leads to

$$\int_{Z_o}^{Z} \frac{\mathrm{d}(Z - Z_o)}{U_o + R\,(Z - Z_o)} = \int_0^t \cos(\omega t)\,\mathrm{d}t. \tag{45.57}$$

The left hand side integral can be computed by changing variables

$$\Sigma = U_o + R\,(Z - Z_o) \tag{45.58a}$$

$$\mathrm{d}\Sigma = R\,\mathrm{d}(Z - Z_o), \tag{45.58b}$$

so that equation (45.57) becomes

$$\int_{U_o}^{\Sigma} \frac{\mathrm{d}\Sigma}{\Sigma} = R \int_0^t \cos(\omega t)\,\mathrm{d}t. \tag{45.59}$$

Performing the integrals and evaluating the end points renders

$$\ln\left[1 + \frac{R}{U_o}(Z - Z_o)\right] = \frac{R\sin(\omega t)}{\omega}, \tag{45.60}$$

which yields the exponential solution

$$1 + \frac{R}{U_o}(Z - Z_o) = e^{(R/\omega)\sin(\omega t)} \Rightarrow Z - Z_o = \frac{U_o}{R}\left(-1 + e^{(R/\omega)\sin(\omega t)}\right). \tag{45.61}$$

The vertical particle position is seen to oscillate around its initial position Z_o.

We next consider the zonal particle position, in which case

$$\frac{\mathrm{d}(X - X_o)}{\mathrm{d}t} = U_o\left[1 + \frac{R}{U_o}(Z - Z_o)\right]\sin(\omega t) \tag{45.62a}$$

$$= U_o\, e^{(R/\omega)\sin(\omega t)}\sin(\omega t) \tag{45.62b}$$

where we used equation (45.61) for the vertical trajectory. To make progress, we expand the exponential assuming the ratio of inverse time scales, R/ω, is small

$$|R/\omega| \ll 1. \tag{45.63}$$

In this limit, the vertical trajectory retains its unperturbed form (45.47b), and the zonal trajectory satisfies

$$\frac{\mathrm{d}(X - X_o)}{\mathrm{d}t} \approx U_o\,\sin(\omega t)\left[1 + \frac{R}{\omega}\sin(\omega t)\right], \tag{45.64}$$

where we dropped terms of order $(R/\omega)^2$. We can understand the scaling in equation (45.63) by noting that the period for the circular motion is given by

$$\tau_{\text{circle}} = \frac{2\,\pi}{\omega}. \tag{45.65}$$

The inverse time R introduces a time scale for the drift, defined according to

$$\tau_{\text{drift}} = \frac{2\,\pi}{|R|}. \tag{45.66}$$

A small ratio $|R/\omega|$ thus implies

$$|R/\omega| = \tau_{\text{circle}}/\tau_{\text{drift}} \ll 1. \tag{45.67}$$

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 (45.64) yields

$$\frac{\mathrm{d}(X - X_o)}{\mathrm{d}t} = U_o \sin(\omega t)\left[1 + \frac{R}{\omega}\sin(\omega t)\right] \tag{45.68a}$$

$$= U_o \sin(\omega t) + \frac{U_o\,R}{2\,\omega}\,[1 - \cos(2\,\omega t)], \tag{45.68b}$$

which integrates to

$$X - X_o = \left(\frac{U_o}{\omega}\right)\left[1 - \cos(\omega\,t) - \frac{R\sin(2\omega t)}{4\omega} + \frac{R\,t}{2}\right] \tag{45.69a}$$

$$= \underbrace{\left(\frac{U_o}{\omega}\right)[1 - \cos(\omega\,t)]}_{\text{homogeneous}} + \underbrace{\frac{U_o\,R\,t}{2\,\omega}}_{\text{Stokes drift}} - \underbrace{\frac{U_o\,R\sin(2\omega t)}{4\,\omega^2}}_{\text{higher harmonic}} + \mathcal{O}(R/\omega)^2. \tag{45.69b}$$

The leading order term is the homogeneous motion given by equation (45.47a). 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

$$\left[\frac{\mathbf{X} - \mathbf{X_o}}{t}\right]^{\text{drift}} = \frac{R\,U_o}{2\omega}\,\hat{\mathbf{x}}. \tag{45.70}$$

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 45.5 for an illustration based on a particular choice for the dimensional parameters.

45.3.3 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 48 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 (45.45)

$$v^n_{(S)}(\boldsymbol{x}, t) \approx \overline{\nabla v^n(\boldsymbol{x}, t) \cdot \int_0^t \boldsymbol{v}(\boldsymbol{x}, t')\,\mathrm{d}t'}. \tag{45.71}$$

Figure 45.5: *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 (45.47a) and (45.47b), 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 (45.61) for the vertical component and equation (45.69b) 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.*

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 nonlinearites.

45.4 Exercises

EXERCISE 45.1: STOKES DRIFT FOR ONE-DIMENSIONAL MONOCHROMATIC WAVE

Consider a one-dimensional monochromatic longitudinal wave with velocity

$$u = u_o \sin(k\,x - \omega\,t), \tag{45.72}$$

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 $\hat{\boldsymbol{x}}$ direction. Determine the wave period averaged Stokes velocity to first order accuracy in the small parameter

$$\epsilon = \frac{u_o}{c} = \frac{u_o\,k}{\omega} = \frac{u_o\,T}{\lambda} \tag{45.73}$$

with this parameter the ratio of the wave amplitude to wave speed, or equivelently the ratio of the length scale of particle displacements to the wavelength. Hint: make use of the general result given by equation (45.45).

46

Isopycnal ocean 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 8, kinematics in Chapter 17, and dynamics in Chapter 38. 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 38.

46.1 Layered isopycnal primitive equations **800**
46.1.1 Montgomery potential and the pressure force 801
46.1.2 Material time derivative 802
46.1.3 Layer thickness and specific thickness 802
46.1.4 Ocean equations 803
46.1.5 Vector-invariant horizontal momentum equation 804
46.1.6 Connection to the shallow water equations 804
46.1.7 Diapycnal transfer 805
46.1.8 Momentum transfer 805
46.1.9 Allowing for layers to vanish and reappear 805
46.2 Perfect fluid PV using isopycnal coordinates **805**

46.2.1 Derivation of the vorticity equation 806
46.2.2 Derivation of the potential vorticity equation 806
46.2.3 Coordinate transforming vorticity and potential vorticity 807
46.3 Isopycnal coordinate PV with irreversible processes **807**
46.3.1 Derivation method I 807
46.3.2 Derivation method II 809
46.3.3 Comments 810

46.1 Layered isopycnal primitive equations

Rather than specializing the generalized vertical coordinate equations provided in Section 38.1, we find it pedagogical to start from the equations written using the geopotential vertical coordinate (see Section 25.2.6)

$$\frac{\mathrm{D}\boldsymbol{u}}{\mathrm{D}t} + f\,\hat{\boldsymbol{z}} \wedge \boldsymbol{u} = -\nabla_z \varphi + \boldsymbol{F} \qquad \text{horizontal momentum} \tag{46.1a}$$

$$\frac{\partial \varphi}{\partial z} = b \qquad \text{hydrostatic} \tag{46.1b}$$

$$\nabla_z \cdot \boldsymbol{u} + \frac{\partial w}{\partial z} = 0 \qquad \text{continuity} \tag{46.1c}$$

$$\frac{\mathrm{D}b}{\mathrm{D}t} = \dot{b} \qquad \text{thermodynamics} \tag{46.1d}$$

$$\frac{\mathrm{D}C}{\mathrm{D}t} = \dot{C} \qquad \text{tracers,} \tag{46.1e}$$

where $\boldsymbol{v} = (\boldsymbol{u}, w)$ is the velocity field, $\boldsymbol{u}$ 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 8.11), written here in the form

$$\nabla_b = \hat{\boldsymbol{x}} \left[\frac{\partial}{\partial x}\right]_b + \hat{\boldsymbol{y}} \left[\frac{\partial}{\partial y}\right]_b . \tag{46.2}$$

Following the discussion in Section 38.1.3, the horizontal pressure thus gradient transforms as

$$\nabla_z \varphi = \nabla_b \varphi - \frac{\partial \varphi}{\partial z} \nabla_b z \tag{46.3a}$$
$$= \nabla_b \varphi - b \nabla_b z \tag{46.3b}$$
$$= \nabla_b (\varphi - b\, z) \tag{46.3c}$$
$$= \nabla_b M, \tag{46.3d}$$

where

$$M = \varphi - b\, z \tag{46.4}$$

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 38.1).

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

$$\frac{\partial M}{\partial b} = \frac{\partial \varphi}{\partial b} - b\,\frac{\partial z}{\partial b} - z = \frac{\partial \varphi}{\partial z}\,\frac{\partial z}{\partial b} - b\,\frac{\partial z}{\partial b} - z = -z, \tag{46.5}$$

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 17.4, there are two equivalent forms for the material time derivative

$$\frac{\mathrm{D}}{\mathrm{D}t} = \left[\frac{\partial}{\partial t}\right]_z + \boldsymbol{u}\cdot\nabla_z + w\,\frac{\partial}{\partial z} \qquad \text{geopotential form} \tag{46.6a}$$

$$= \left[\frac{\partial}{\partial t}\right]_b + \boldsymbol{u}\cdot\nabla_b + w^{(\dot{b})}\,\frac{\partial}{\partial z} \qquad \text{isopycnal form,} \tag{46.6b}$$

where

$$w^{(\dot{b})} = \frac{\partial z}{\partial b}\,\frac{\mathrm{D}b}{\mathrm{D}t} \tag{46.7}$$

is the diapycnal velocity component that measures the rate of flow crossing buoyancy surfaces (Section 17.3). 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

$$\boldsymbol{u} = \hat{\boldsymbol{x}}\,u + \hat{\boldsymbol{y}}\,v = (\mathrm{D}/\mathrm{D}t)\,(\hat{\boldsymbol{x}}\,x + \hat{\boldsymbol{y}}\,y). \tag{46.8}$$

46.1.3 Layer thickness and specific thickness

The continuity equation, $\nabla_z\cdot\boldsymbol{u} + \partial_z w = 0$, is an expression of volume conservation. We already derived the generalized vertical coordinate version of this equation in Section 17.9.3, and thus quote the isopycnal layer thickness result here

$$\left[\frac{\partial h}{\partial t}\right]_b + \nabla_b\cdot(h\,\boldsymbol{u}) + \Delta_b\,w^{(\dot{b})} = 0. \tag{46.9}$$

The field h measures the isopycnal layer thickness (dimensions of length) and is given by the vertical integral over a layer

$$h = \int_{z(b-\delta b/2)}^{z(b+\delta b/2)} \mathrm{d}z = \int_{b-\delta b/2}^{b+\delta b/2} \frac{\partial z}{\partial b}\,\mathrm{d}b = \int_{b-\delta b/2}^{b+\delta b/2} \hbar\,\mathrm{d}b = \int_{b-\delta b/2}^{b+\delta b/2} N^{-2}\,\mathrm{d}b = \overline{\hbar}\,\mathrm{d}b. \tag{46.10}$$

The specific thickness, $\hbar$, equals to the inverse squared buoyancy frequency

$$\hbar = \frac{\partial z}{\partial b} = N^{-2}, \tag{46.11}$$

and its layer averaged value is $\overline{\hbar} = h/\delta b$. Furthermore, the dia-surface transport velocity is given by

$$w^{(\dot{b})} = \hbar\,\dot{b}. \tag{46.12}$$

Its difference across layer interfaces

$$\Delta_b\,w^{(\dot{b})} = \int_{b-\delta b/2}^{b+\delta b/2} \frac{\partial w^{(\dot{b})}}{\partial b}\,\mathrm{d}b = w^{(\dot{b})}(b+\delta b/2) - w^{(\dot{b})}(b-\delta b/2) \tag{46.13}$$

measures the amount of fluid that diverges from the layer through cross-layer transport.

Specific thickness equation

Inserting $h = \overline{\hbar}\,\delta b$ into the thickness equation (46.9) leads to

$$\left[\frac{\partial \overline{\hbar}}{\partial t}\right]_b + \nabla_b \cdot (\overline{\hbar}\,\boldsymbol{u}) + \partial_b\, w^{(\dot{b})} = 0, \tag{46.14}$$

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

$$\delta b\, \partial_b = \Delta_b. \tag{46.15}$$

For a vertically continuous treatment, equation (46.14) can be written with $\hbar$ rather than the discrete layer averaged value

$$\left[\frac{\partial \hbar}{\partial t}\right]_b + \nabla_b \cdot (\hbar\,\boldsymbol{u}) + \partial_b\, w^{(\dot{b})} = 0. \tag{46.16}$$

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 $w^{(\dot{b})} \neq 0$, the three terms in the thickness equation (46.9) or the specific thickness equation (46.14) are coupled. We discussed this coupling in Section 17.5 as part of our broader study of the vertical velocity and the dia-surface velocity. When considering perfect fluids, we can set $w^{(\dot{b})} = \hbar\,\dot{b} = 0$ 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

$$\left[\frac{\partial h}{\partial t}\right]_b + \nabla_b \cdot (h\,\boldsymbol{u}) = 0. \tag{46.17}$$

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

$$\left[\frac{\partial \boldsymbol{u}}{\partial t}\right]_b + (\boldsymbol{u}\cdot\nabla_b)\,\boldsymbol{u} + (w^{(\dot{b})}\,\partial_z)\,\boldsymbol{u} + f\,\hat{\boldsymbol{z}}\wedge\boldsymbol{u} = -\nabla_b M + \boldsymbol{F}^h \tag{46.18a}$$

$$\frac{\partial M}{\partial b} = -z \tag{46.18b}$$

$$\left[\frac{\partial h}{\partial t}\right]_b + \nabla_b \cdot (h\,\boldsymbol{u}) + \Delta_b w^{(\dot{b})} = 0 \tag{46.18c}$$

$$\left[\frac{\partial (h\,C)}{\partial t}\right]_b + \nabla_b \cdot (h\,C\,\boldsymbol{u} + h\,\boldsymbol{J}^h) + \Delta_b(C\,w^{(\dot{b})} + J^{(b)}) = 0, \tag{46.18d}$$

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

$$f\,\hat{\boldsymbol{z}} \wedge \boldsymbol{u}_{\mathrm{g}} = -\nabla_b\, M \Longrightarrow f u_{\mathrm{g}} = -\left[\frac{\partial M}{\partial y}\right]_b \qquad \text{and} \qquad f v_{\mathrm{g}} = \left[\frac{\partial M}{\partial x}\right]_b . \tag{46.19}$$

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 38.1.5. Introducing the isopycnal version of the relative vorticity

$$\hat{\boldsymbol{z}}\,\widetilde{\zeta} \equiv \nabla_b \wedge \boldsymbol{u} = \left[\frac{\partial v}{\partial x}\right]_b - \left[\frac{\partial u}{\partial y}\right]_b \tag{46.20}$$

renders the vector-invariant horizontal momentum equation

$$\left[\frac{\partial \boldsymbol{u}}{\partial t}\right]_b + w^{(\dot{b})}\,\partial_z\,\boldsymbol{u} + \widetilde{\zeta}_{\mathrm{a}}\,\hat{\boldsymbol{z}} \wedge \boldsymbol{u} = -\nabla_b \mathcal{B} + \boldsymbol{F}^h, \tag{46.21}$$

where

$$\mathcal{B} = M + \boldsymbol{u}\cdot\boldsymbol{u}/2 = \varphi - b\,z + \boldsymbol{u}\cdot\boldsymbol{u}/2 \tag{46.22}$$

is the Bernoulli potential for a hydrostatic Boussinesq fluid (see Section 23.4.4), and

$$\widetilde{\zeta}_{\mathrm{a}} = \widetilde{\zeta} + f \tag{46.23}$$

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)

$$h\,Q = \widetilde{\zeta}_{\mathrm{a}} \tag{46.24}$$

to bring the momentum equation to the form

$$\left[\frac{\partial \boldsymbol{u}}{\partial t}\right]_b + w^{(\dot{b})}\,\partial_z\,\boldsymbol{u} + Q\,\hat{\boldsymbol{z}} \wedge (h\,\boldsymbol{u}) = -\nabla_b \mathcal{B} + \boldsymbol{F}^h. \tag{46.25}$$

This form is commonly used as the starting point for certain theoretical analyses, particularly when considering the adiabatic limit in which $w^{(\dot{b})} = 0$.

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)

$$\frac{\mathrm{D}\boldsymbol{u}}{\mathrm{D}t} + f\,\hat{\boldsymbol{z}} \wedge \boldsymbol{u} = -\nabla_b\, M \tag{46.26a}$$

$$\frac{\partial M}{\partial b} = -z \tag{46.26b}$$

$$\frac{\mathrm{D}h}{\mathrm{D}t} + h\,\nabla_b \cdot \boldsymbol{u} = 0. \tag{46.26c}$$

These isopycnal momentum and thickness equations are isomorphic to those for a single layer of adiabatic shallow water fluid (see Section 29.1). 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 33.3.

46.1.7 Diapycnal transfer

At ocean boundaries, the diapycnal term, $w^{(\dot{b})}$, acccounts 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, $w^{(\dot{b})}$ 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 30.2 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 formulational 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 25.6. 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 36.5, we showed that the absolute vorticity in a Boussinesq hydrostatic fluid with a simplified seawater equation of state (Section 37.2), when projected into the direction normal to constant buoyancy surface, $\boldsymbol{\omega}_{\mathrm{a}} \cdot \nabla b$, is not affected by baroclinicity; i.e., that projection annihilates the baroclinicity vector. From that property we conclude that $\boldsymbol{\omega}_{\mathrm{a}} \cdot \nabla b$ 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 38.3.3, 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 38.3, 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, $\hat{\boldsymbol{z}} \cdot (\nabla_b \wedge)$, onto the adiabatic and inviscid form of the vector-invariant velocity equation (46.21) leads to the isopycnal vorticity equation

$$\left[\frac{\partial \tilde{\zeta}_{\mathrm{a}}}{\partial t}\right]_b + (\boldsymbol{u} \cdot \nabla_b)\, \tilde{\zeta}_{\mathrm{a}} = -\tilde{\zeta}_{\mathrm{a}}\, \nabla_b \cdot \boldsymbol{u} \tag{46.27}$$

where

$$\tilde{\zeta}_{\mathrm{a}} = f + \hat{\boldsymbol{z}} \cdot (\nabla_b \wedge \boldsymbol{u}) = f + \tilde{\zeta} \tag{46.28}$$

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

$$\frac{\mathrm{D}\tilde{\zeta}_{\mathrm{a}}}{\mathrm{D}t} = -\tilde{\zeta}_{\mathrm{a}}\, \nabla_b \cdot \boldsymbol{u}. \tag{46.29}$$

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 $\nabla_b \cdot \boldsymbol{u}$ on the right hand side of equation (46.29), thus leading to

$$\frac{\mathrm{D}\tilde{\zeta}_{\mathrm{a}}}{\mathrm{D}t} - \frac{\tilde{\zeta}_{\mathrm{a}}}{h}\frac{\mathrm{D}h}{\mathrm{D}t} = 0. \tag{46.30}$$

Introducing the isopycnal potential vorticity

$$Q = \frac{\tilde{\zeta}_{\mathrm{a}}}{h} = \frac{f + \tilde{\zeta}}{h} \tag{46.31}$$

leads to

$$\frac{\mathrm{D}Q}{\mathrm{D}t} = 0. \tag{46.32}$$

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

$$\left[\frac{\partial\,(h\,Q)}{\partial t}\right]_b + \nabla_b \cdot (h\,Q\,\boldsymbol{u}) = 0. \tag{46.33}$$

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

$$Q = \frac{\tilde{\zeta}_{\mathrm{a}}}{\hbar} = \frac{f + \tilde{\zeta}}{\hbar}. \tag{46.34}$$

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 isopyncal coordinates that also reveals this form

$$(\boldsymbol{\omega}^{\text{hy}} + f\,\hat{\boldsymbol{z}}) \cdot \nabla b = -\frac{\partial v}{\partial z}\frac{\partial b}{\partial x} + \frac{\partial u}{\partial z}\frac{\partial b}{\partial y} + \left(\frac{\partial v}{\partial x} - \frac{\partial u}{\partial y} + f\right)\frac{\partial b}{\partial z} \tag{46.35a}$$

$$= \frac{\partial b}{\partial z}\left[f + \left(\frac{\partial v}{\partial x} - \frac{\partial v}{\partial z}\frac{\partial b/\partial x}{\partial b/\partial z}\right) - \left(\frac{\partial u}{\partial y} - \frac{\partial u}{\partial z}\frac{\partial b/\partial y}{\partial b/\partial z}\right)\right] \tag{46.35b}$$

$$= \frac{\partial b}{\partial z}\left[f + \left(\frac{\partial v}{\partial x}\right)_b - \left(\frac{\partial u}{\partial y}\right)_b\right] \tag{46.35c}$$

$$= \frac{f + (\partial v/\partial x)_b - (\partial u/\partial y)_b}{\partial z/\partial b} \tag{46.35d}$$

$$= \frac{f + \tilde{\zeta}}{\hbar} \tag{46.35e}$$

$$= Q. \tag{46.35f}$$

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 36.2 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, $\hbar$, 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 38.3. 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

$$\left[\frac{\partial \boldsymbol{u}}{\partial t}\right]_b + (\boldsymbol{u}\cdot\nabla_b)\,\boldsymbol{u} + \dot{b}\,\frac{\partial \boldsymbol{u}}{\partial b} + \boldsymbol{f}\wedge\boldsymbol{u} = -\nabla_b M + \boldsymbol{F} \tag{46.36a}$$

$$\frac{\partial M}{\partial b} = -z \tag{46.36b}$$

$$\left[\frac{\partial \hbar}{\partial t}\right]_b + \nabla_b\cdot(\hbar\,\boldsymbol{u}) = -\frac{\partial(\hbar\,\dot{b})}{\partial b} \tag{46.36c}$$

$$\frac{\mathrm{D}\,b}{\mathrm{D}t} = \dot{b}. \tag{46.36d}$$

Note that in this section choose to write the dia-surface transport operator in the form

$$w^{(\dot{b})}\,\partial_z = \dot{b}\,\partial_b. \tag{46.37}$$

We can make use of the material time derivative operator (46.6b) to write the material form of the equations

$$\frac{\mathrm{D}\boldsymbol{u}}{\mathrm{D}t} + \boldsymbol{f} \wedge \boldsymbol{u} = -\nabla_b M + \boldsymbol{F} \tag{46.38a}$$

$$\frac{\partial M}{\partial b} = -z \tag{46.38b}$$

$$\frac{\mathrm{D}\hbar}{\mathrm{D}t} + \hbar\, \nabla_b \cdot \boldsymbol{u} = -\hbar\, \frac{\partial \dot{b}}{\partial b}. \tag{46.38c}$$

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

$$\left[\frac{\partial \tilde{\zeta}_{\mathrm{a}}}{\partial t}\right]_b + (\boldsymbol{u} \cdot \nabla_b)\, \tilde{\zeta}_{\mathrm{a}} + \dot{b} \left[\frac{\partial \tilde{\zeta}_{\mathrm{a}}}{\partial b}\right] = -\tilde{\zeta}_{\mathrm{a}}\, \nabla_b \cdot \boldsymbol{u} + \hat{\boldsymbol{z}} \cdot \left[\frac{\partial \boldsymbol{u}}{\partial b} \wedge \nabla_b \dot{b} + \nabla_b \wedge \boldsymbol{F}\right]. \tag{46.39}$$

The left hand side of equation (46.39) is the material time derivative of absolute vorticity (see equation (46.6b)), so that

$$\frac{\mathrm{D}\tilde{\zeta}_{\mathrm{a}}}{\mathrm{D}t} = -\tilde{\zeta}_{\mathrm{a}}\, \nabla_b \cdot \boldsymbol{u} + \hat{\boldsymbol{z}} \cdot \left[\frac{\partial \boldsymbol{u}}{\partial b} \wedge \nabla_b \dot{b} + \nabla_b \wedge \boldsymbol{F}\right]. \tag{46.40}$$

Now make use of the thickness equation in the material form (46.38c) to eliminate the convergence $\nabla_b \cdot \boldsymbol{u}$ on the right hand side, thus leading to

$$\frac{\mathrm{D}\tilde{\zeta}_{\mathrm{a}}}{\mathrm{D}t} - \frac{\tilde{\zeta}_{\mathrm{a}}}{\hbar} \left[\frac{\mathrm{D}\hbar}{\mathrm{D}t} - \hbar \frac{\partial \dot{b}}{\partial b}\right] = \hat{\boldsymbol{z}} \cdot \left[\frac{\partial \boldsymbol{u}}{\partial b} \wedge \nabla_b \dot{b} + \nabla_b \wedge \boldsymbol{F}\right]. \tag{46.41}$$

Introducing the isopycnal potential vorticity

$$Q = \frac{\tilde{\zeta}_{\mathrm{a}}}{\hbar} = \frac{\tilde{\zeta} + f}{\hbar} \tag{46.42}$$

leads to

$$\hbar\, \frac{\mathrm{D}Q}{\mathrm{D}t} = \zeta_a\, \frac{\partial \dot{b}}{\partial b} + \hat{\boldsymbol{z}} \cdot \left[\frac{\partial \boldsymbol{u}}{\partial b} \wedge \nabla_b \dot{b} + \nabla_b \wedge \boldsymbol{F}\right]. \tag{46.43}$$

Massaging the diabatic terms

The diabatic terms can be written

$$\zeta_a\, \frac{\partial \dot{b}}{\partial b} + \hat{\boldsymbol{z}} \cdot \left[\frac{\partial \boldsymbol{u}}{\partial b} \wedge \nabla_b \dot{b}\right] = \zeta_a\, \frac{\partial \dot{b}}{\partial b} + \dot{b}\, \frac{\partial \tilde{\zeta}}{\partial b} - \hat{\boldsymbol{z}} \cdot \left[\nabla_b \wedge \dot{b} \frac{\partial \boldsymbol{u}}{\partial b}\right] \tag{46.44a}$$

$$= \zeta_a\, \frac{\partial \dot{b}}{\partial b} + \dot{b}\, \frac{\partial \tilde{\zeta}_{\mathrm{a}}}{\partial b} - \hat{\boldsymbol{z}} \cdot \left[\nabla_b \wedge \dot{b} \frac{\partial \boldsymbol{u}}{\partial b}\right] \tag{46.44b}$$

$$= \frac{\partial (\tilde{\zeta}_{\mathrm{a}}\, \dot{b})}{\partial b} - \hat{\boldsymbol{z}} \cdot \left[\nabla_b \wedge \dot{b} \frac{\partial \boldsymbol{u}}{\partial b}\right] \tag{46.44c}$$

$$= \frac{\partial (\tilde{\zeta}_{\mathrm{a}}\, \dot{b})}{\partial b} + \nabla_b \cdot \left[\hat{\boldsymbol{z}} \wedge \dot{b} \frac{\partial \boldsymbol{u}}{\partial b}\right], \tag{46.44d}$$

where the second equality follows since the Coriolis parameter is independent of the buoyancy.

The PV equation

The PV equation takes the material form

$$\hbar \left[\frac{\mathrm{D}Q}{\mathrm{D}t}\right] = \frac{\partial(\tilde{\zeta}_{\mathrm{a}}\,\dot{b})}{\partial b} + \nabla_b \cdot \left[\hat{\boldsymbol{z}} \wedge \dot{b}\frac{\partial \boldsymbol{u}}{\partial b} - \hat{\boldsymbol{z}} \wedge \boldsymbol{F}\right]. \tag{46.45}$$

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

$$\left[\frac{\partial\,(\hbar\,Q)}{\partial t}\right]_b + \nabla_b \cdot (\hbar\,Q\,\boldsymbol{u}) + \frac{\partial\,(\hbar\,Q\,\dot{b})}{\partial b} = \frac{\partial(\tilde{\zeta}_{\mathrm{a}}\,\dot{b})}{\partial b} + \nabla_b \cdot \left[\hat{\boldsymbol{z}} \wedge \dot{b}\frac{\partial \boldsymbol{u}}{\partial b} - \hat{\boldsymbol{z}} \wedge \boldsymbol{F}\right]. \tag{46.46}$$

Since $\hbar\,Q = \tilde{\zeta}_{\mathrm{a}}$, the ∂_b terms cancel, thus leaving the flux-form PV equation

$$\left[\frac{\partial\,(\hbar\,Q)}{\partial t}\right]_b = -\nabla_b \cdot \left[\hbar\,Q\,\boldsymbol{u} - \hat{\boldsymbol{z}} \wedge \dot{b}\frac{\partial \boldsymbol{u}}{\partial b} + \hat{\boldsymbol{z}} \wedge \boldsymbol{F}\right]. \tag{46.47}$$

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 36.5.2 to write the material evolution of PV for a hydrostatic and Boussinesq fluid

$$\frac{\mathrm{D}Q}{\mathrm{D}t} = \nabla \cdot \left[(f\,\hat{\boldsymbol{z}} + \boldsymbol{\omega}_{\mathrm{hy}})\,\dot{b} + b\,\nabla \wedge \boldsymbol{F}\right], \tag{46.48}$$

where

$$Q = \boldsymbol{\omega}_{\mathrm{a}}^{\mathrm{hy}} \cdot \nabla b = \boldsymbol{\omega}^{\mathrm{hy}} \cdot \nabla b + f\,\frac{\partial b}{\partial z} \qquad \text{and} \qquad \boldsymbol{\omega}_{\mathrm{hy}} = -\hat{\boldsymbol{x}}\frac{\partial v}{\partial z} + \hat{\boldsymbol{y}}\frac{\partial u}{\partial z} + \hat{\boldsymbol{z}}\left[\frac{\partial v}{\partial x} - \frac{\partial u}{\partial y}\right]. \tag{46.49}$$

The simplest term in equation (46.48) to transform to isopycnal coordinates is the curl of the horizontal friction vector

$$\hat{\boldsymbol{z}} \cdot (\nabla \wedge \boldsymbol{F}) = \hat{\boldsymbol{z}} \cdot \nabla \wedge (F^x, F^y, 0), \tag{46.50}$$

which takes the form

$$\hat{\boldsymbol{z}} \cdot (\nabla \wedge \boldsymbol{F}) = \hbar^{-1}\,\hat{\boldsymbol{z}} \cdot (\nabla_b \wedge \boldsymbol{F}) = -\hbar^{-1}\,\nabla_b \cdot (\hat{\boldsymbol{z}} \wedge \boldsymbol{F}). \tag{46.51}$$

The diabatic term requires some more work. Since the vorticity has zero divergence, the diabatic term can be written as

$$\nabla \cdot \left[(f\,\hat{\boldsymbol{z}} + \boldsymbol{\omega}_{\mathrm{hy}})\,\dot{b}\right] = (f\,\hat{\boldsymbol{z}} + \boldsymbol{\omega}_{\mathrm{hy}}) \cdot \nabla\dot{b} \tag{46.52a}$$

$$= f\,\frac{\partial \dot{b}}{\partial z} - \frac{\partial v}{\partial z}\frac{\partial \dot{b}}{\partial x} + \frac{\partial u}{\partial z}\frac{\partial \dot{b}}{\partial y} + \left[\frac{\partial v}{\partial x} - \frac{\partial u}{\partial y}\right]\frac{\partial \dot{b}}{\partial z}. \tag{46.52b}$$

We now introduce horizontal derivatives on isopycnal surfaces according to the following relation (see Section 8.11.2)

$$\nabla_z = \nabla_b + \nabla_z b\,\left[\frac{\partial z}{\partial b}\right]\frac{\partial}{\partial z} \tag{46.53}$$

Doing so leads to

$$
\begin{aligned}
\nabla \cdot \left[(f\,\hat{\boldsymbol{z}} + \boldsymbol{\omega}_{\text{hy}})\,\dot{b} \right] &= 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) \\
&\qquad + \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}_{\text{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) \\
&= \hbar^{-1}\left(\tilde{\zeta}_{\text{a}}\left[\frac{\partial \dot{b}}{\partial b} \right] + \hat{\boldsymbol{z}} \cdot \left[\frac{\partial \boldsymbol{u}}{\partial b} \wedge \nabla_b \dot{b} \right] \right) \\
&= \hbar^{-1}\left(\tilde{\zeta}_{\text{a}}\left[\frac{\partial \dot{b}}{\partial b} \right] + \dot{b}\left[\frac{\partial \tilde{\zeta}_{\text{a}}}{\partial b} \right] + \nabla_b \cdot \left[\hat{\boldsymbol{z}} \wedge \dot{b}\,\frac{\partial \boldsymbol{u}}{\partial b} \right] \right) \\
&= \hbar^{-1}\left(\frac{\partial(\dot{b}\,\tilde{\zeta}_{\text{a}})}{\partial b} + \nabla_b \cdot \left[\hat{\boldsymbol{z}} \wedge \dot{b}\,\frac{\partial \boldsymbol{u}}{\partial b} \right] \right).
\end{aligned}
\tag{46.54}
$$

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 37 since the right hand side is the isopycnal convergence of a flux.

47

Advection and diffusion

In this chapter we discuss physical and mathematical aspects of advection and diffusion, which are the two primary processes that affect tracer concentration at a point in the fluid. Advection and diffusion have complementary physical and mathematical properties. 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 supported by the discussion in this chapter.

For most of this chapter we consider a compressible/non-Boussinesq fluid. For a Boussinesq fluid (Chapter 25), the density factor, ρ, appearing in the compressible formulation is set to a constant and trivially cancels from all formula. As we will see, the compressible/non-Boussinesq formulation adds no formal complexity to the derivations.

47.1 Introduction **812**
47.2 Diffusion physics **812**
47.2.1 Diffusion of matter by random molecular motions 813
47.2.2 Diffusion of matter by random turbulent motions 813
47.2.3 Fick's law for matter diffusion 813
47.2.4 Fourier's law for heat diffusion 815
47.2.5 Newtonian frictional stress and momentum diffusion 815
47.2.6 Further study 816
47.3 Diffusion maths **816**
47.3.1 Sample diffusion tensors 817
47.3.2 Evolution of tracer concentration powers 817
47.3.3 Moments of tracer concentration 817
47.3.4 Connecting tracer variance to the diffusion operator 819
47.4 Advection physics **821**
47.4.1 The advection equation 821
47.4.2 Eulerian time tendencies from advection 821
47.4.3 Impermeability property of tracer isosurfaces 822
47.5 Advection maths **822**
47.5.1 Material constancy of C^Γ 823
47.5.2 Mass transport 823
47.5.3 Advective tracer fluxes and skew tracer fluxes 824
47.5.4 Skew diffusion 824

47.5.5 A comment about skew fluxes and Lagrangian kinematics 825
47.5.6 Further reading 826
47.6 Advection and skewsion **826**
47.6.1 Choosing a gauge 826
47.6.2 Vertical gauge 827
47.6.3 Boundary conditions 828
47.7 Finite volume budgets with eddy velocities **828**
47.7.1 Advective formulation 829
47.7.2 Skew flux formulation 830
47.7.3 Domain with a tracer boundary 830
47.7.4 Budget for a region with interior sides 833
47.7.5 Budget for a stirred fluid in a region with interior sides 835
47.8 Active tracers and dia-surface flow **837**
47.8.1 Adiabatic flow 838
47.8.2 Diabatic processes generating dia-surface transport 838
47.9 Exercises **839**

47.1 Introduction

As derived in Section 15.1, the tracer equation takes on the general form

$$\rho \frac{\mathrm{D}C}{\mathrm{D}t} = -\nabla \cdot \boldsymbol{J}(C), \tag{47.1}$$

where $\boldsymbol{J}$ is a flux that embodies molecular diffusion as well as subgrid scale advection and subgrid scale diffusion (Chapter 49). Advective transport appears when transforming to an Eulerian or laboratory reference frame, in which case

$$\rho \frac{\mathrm{D}C}{\mathrm{D}t} = \frac{\partial(\rho\, C)}{\partial t} + \nabla \cdot (\boldsymbol{v}\, \rho\, C), \tag{47.2}$$

with $\boldsymbol{v}\, \rho\, C$ 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 propertes between fluid elements. Correspondingly, diffusion reduces the magnitude of property gradients as it irreversibly exchanges properties between fluid elements.

47.2 Diffusion physics

The continuum hypothesis from Chapter 11 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.

[1] Recall from Chapter 22 that specific entropy remains materially constant on fluid elements in the absence of mixing.

[2] For our purposes, turbulence is characterized by a quasi-random motion of fluid elements.

47.2.1 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 12.2, 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 15.2) 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 15.1. 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 hypothesis 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.

47.2.2 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.

47.2.3 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 47.1. 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

$$\boldsymbol{J} = -\kappa\,\rho\,\nabla C. \tag{47.3}$$

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*

$$\kappa \qquad \text{kinematic diffusivity with SI units m s}^{-2} \tag{47.4}$$

$$\rho\,\kappa \qquad \text{dynamic diffusivity with SI units kg m}^{-2}\text{ s}^{-2}. \tag{47.5}$$

The kinematic diffusivity has dimensions of squared length per time and it sets the efficiency or strength of the diffusion. The diffusive flux (47.3) 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.

Figure 47.1: *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 particlar example, $\partial C/\partial z > 0$, so that random fluid motions (either molecular or turbulent) lead to a diffusive flux directed in the $-\hat{\boldsymbol{z}}$ 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.*

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 11.2.3), and the root-mean-square molecular speed, v_{rms} (see Section 11.2.4). 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 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

$$\rho \frac{\partial C}{\partial t} = -\nabla \cdot \boldsymbol{J} = \nabla \cdot (\kappa \, \rho \, \nabla C). \tag{47.6}$$

That is, the concentration increases in regions where the diffusive flux, $\boldsymbol{J}$, converges, and decreases where the flux diverges. Expanding the divergence operator leads to

$$\frac{\partial C}{\partial t} = \rho^{-1} \, \nabla(\kappa \, \rho) \cdot \nabla C + \kappa \, \nabla^2 C. \tag{47.7}$$

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.

47.2.4 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

$$\boldsymbol{J} = -\gamma \, \rho \, \nabla T, \tag{47.8}$$

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.

47.2.5 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 47.2, with shear (i.e., nonzero velocity gradient) in a single direction, Newtonian frictional stress takes the form

$$\tau = \rho \, \mu \, \frac{\partial u}{\partial z}, \tag{47.9}$$

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

Figure 47.2: *Shown here is a line graph illustrating the 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.*

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

$$\mathrm{Pr} = \frac{\mu}{\kappa}. \tag{47.10}$$

Theories for the turbulent Prandtl number are largely empirical in nature, with first principles arguments elusive.

47.2.6 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.* (2012). A lucid treatment of Brownian motion in the context of turbulent diffusion is given by *Vallis* (2017).

47.3 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 21.4). 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

$$J_m = -\rho\, \mathbb{K}_{mn} \frac{\partial C}{\partial x^n}, \tag{47.11}$$

and the corresponding diffusion equation is

$$\rho \frac{\partial C}{\partial t} = -\nabla \cdot \boldsymbol{J} = \frac{\partial(\rho\, \mathbb{K}_{mn}\, \partial_n C)}{\partial x_m}. \tag{47.12}$$

47.3.1 Sample diffusion tensors

For the isotropic case of molecular diffusion considered in equation (47.3), the diffusion tensor takes on the form

$$\mathbb{K}_{mn} = \kappa\, \delta_{mn} \qquad \text{isotropic diffusion.} \tag{47.13}$$

If we rotate the diffusive fluxes to be along surfaces of constant γ, then

$$\mathbb{K}_{mn} = \kappa\, (\delta_{mn} - \hat{\gamma}_m\, \hat{\gamma}_n) \qquad \text{rotated diffusion,} \tag{47.14}$$

where

$$\hat{\boldsymbol{\gamma}} = \frac{\nabla \gamma}{|\nabla \gamma|} \tag{47.15}$$

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 49.3).

47.3.2 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 $\Gamma \geq 1$

$$\rho\, \frac{\mathrm{D}C^\Gamma}{\mathrm{D}t} = \Gamma\, C^{\Gamma-1}\, \rho\, \frac{\mathrm{D}C}{\mathrm{D}t} \tag{47.16a}$$

$$= -\Gamma\, C^{\Gamma-1}\, \nabla \cdot \boldsymbol{J} \tag{47.16b}$$

$$= -\Gamma\, \nabla \cdot (C^{\Gamma-1}\, \boldsymbol{J}) + \Gamma\, (\Gamma - 1)\, C^{\Gamma-2}\, \nabla C \cdot \boldsymbol{J} \tag{47.16c}$$

$$= \partial_m \left[\rho\, \mathbb{K}_{mn}\, \frac{\partial C^\Gamma}{\partial x^n} \right] - \rho\, \Gamma\, (\Gamma - 1)\, C^{\Gamma-2}\, \frac{\partial C}{\partial x^n}\, \mathbb{K}_{mn}\, \frac{\partial C}{\partial x^m} \tag{47.16d}$$

$$= -\nabla \cdot \boldsymbol{J}(C^\Gamma) + \Gamma\, (\Gamma - 1)\, C^{\Gamma-2}\, \boldsymbol{J} \cdot \nabla\, C. \tag{47.16e}$$

The first term in equation (47.16e) 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 (47.16e) is negative since the diffusion tensor is symmetric and positive-definite

$$\boldsymbol{J} \cdot \nabla\, C = -\rho\, \mathbb{K}_{mn}\, \frac{\partial C}{\partial x^m}\, \frac{\partial C}{\partial x^n} < 0. \tag{47.17}$$

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 (47.16e) always acts to reduce the magnitude of C^Γ towards zero. We make use of this result in Section 47.3.3 when considering moments of tracer concentration.

47.3.3 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 $\boldsymbol{J} \cdot \hat{\boldsymbol{n}} = 0$ with $\hat{\boldsymbol{n}}$ the outward normal at the boundary. Also assume the total fluid mass in the domain remains fixed

$$\mathcal{M} = \int \rho\, \mathrm{d}V \iff \frac{\mathrm{d}\mathcal{M}}{\mathrm{d}t} = 0. \tag{47.18}$$

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

$$\overline{C} = \frac{\int C\,\rho\,\mathrm{d}V}{\mathcal{M}}, \tag{47.19}$$

and it follows that its time derivative vanishes since

$$\mathcal{M}\,\frac{\mathrm{d}\overline{C}}{\mathrm{d}t} = \frac{\mathrm{d}}{\mathrm{d}t}\int C\,\rho\,\mathrm{d}V = \int \frac{\mathrm{D}C}{\mathrm{D}t}\,\rho\,\mathrm{d}V = -\int \nabla\cdot\boldsymbol{J}\,\mathrm{d}V = -\oint \boldsymbol{J}\cdot\hat{\boldsymbol{n}}\,\mathrm{d}\mathcal{S} = 0, \tag{47.20}$$

where we set $\boldsymbol{J}\cdot\hat{\boldsymbol{n}} = 0$ 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 15.3.5. The result (47.20) 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

$$\mathrm{var}(C) \equiv \frac{\int (C-\overline{C})^2\,\rho\,\mathrm{d}V}{\mathcal{M}} = \overline{C^2} - \overline{C}^2 \geq 0. \tag{47.21}$$

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

$$\frac{\mathrm{d}[\mathrm{var}(C)]}{\mathrm{d}t} = \frac{\mathrm{d}\overline{C^2}}{\mathrm{d}t}. \tag{47.22}$$

Thus, it is common to refer to $\overline{C^2}$ as the tracer variance, though strictly speaking only their time derivatives are equal as per equation (47.22). Performing the time derivative, and again noting that the domain is material thus allowing us to use Reynolds transport theorem, renders

$$\frac{\mathrm{d}\overline{C^2}}{\mathrm{d}t} = \frac{\mathrm{d}}{\mathrm{d}t}\int C^2\,\rho\,\mathrm{d}V = 2\int C\,\frac{\mathrm{D}C}{\mathrm{D}t}\,\rho\,\mathrm{d}V = -2\int C\,\nabla\cdot\boldsymbol{J}\,\mathrm{d}V = 2\int \nabla C\cdot\boldsymbol{J}\,\mathrm{d}V. \tag{47.23}$$

The final equality again made use of the assumed boundary condition $\boldsymbol{J}\cdot\hat{\boldsymbol{n}} = 0$. 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 (47.17) that diffusive fluxes are oriented down the tracer gradient. Consequently, diffusion of the tracer concentration results in a reduction in tracer variance

$$\frac{\mathrm{d}[\mathrm{var}(C)]}{\mathrm{d}t} = \frac{\mathrm{d}\overline{C^2}}{\mathrm{d}t} \leq 0. \tag{47.24}$$

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

$$\frac{\mathrm{d}\overline{C^\Gamma}}{\mathrm{d}t} = \Gamma\,(\Gamma-1)\int C^{\Gamma-2}\,\nabla C\cdot\boldsymbol{J}\,\mathrm{d}V \leq 0. \tag{47.25}$$

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 (47.24). 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.

47.3.4 Connecting tracer variance to the diffusion operator

We take a slight foray here into notions from linear operator theory, in which we note that with natural boundary conditions on the diffusive tracer flux (defined below), the diffusion operator is a linear self-adjoint operator. 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

$$\mathcal{F} \equiv -(1/2) \int |\nabla C|^2 \, \rho \, \mathrm{d}^3 x \tag{47.26}$$

is the associated functional. In this subsection we prove this result for a general diffusion tensor $\mathbb{K}_{mn}$ 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

$$\mathrm{d}^3 x = \mathrm{d}V \tag{47.27}$$

for the volume element. The reason will become apparent when reaching equation (47.35).

Derivative of the diffusion dissipation functional

We introduce the *diffusion dissipation functional*

$$\mathcal{F} = \int \mathcal{L} \, \mathrm{d}^3 x \tag{47.28}$$

where the integrand is the negative semi-definite quadratic form

$$2\,\mathcal{L} = \boldsymbol{J} \cdot \nabla C = -\rho \, \partial_m C \, \mathbb{K}_{mn} \, \partial_n C \leq 0. \tag{47.29}$$

Our goal is to relate the diffusion operator, given by the convergence of the diffusion flux, $\mathcal{R} = -\nabla \cdot \boldsymbol{J}$, 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

$$\delta\mathcal{F} = \int \left[\delta C \, \frac{\delta \mathcal{L}}{\delta C} + \delta(\partial_m C) \, \frac{\delta \mathcal{L}}{\delta(\partial_m C)} \right] \mathrm{d}^3 x. \tag{47.30}$$

Note that functional variations are small perturbations to the form of the function, in which case

$$C \to C + \delta C \qquad \text{with} \qquad |\delta C| \ll |C|. \tag{47.31}$$

Integration by parts on the second term leads to

$$\delta\mathcal{F} = \int \left[\delta C \, \frac{\delta \mathcal{L}}{\delta C} + \partial_m \left(\delta C \, \frac{\delta \mathcal{L}}{\delta(\partial_m C)} \right) - \delta C \, \partial_m \left(\frac{\delta \mathcal{L}}{\delta(\partial_m C)} \right) \right] \mathrm{d}^3 x. \tag{47.32}$$

The middle term is a total derivative that integrates to a boundary contribution and the associated *natural boundary condition*

$$\hat{\boldsymbol{n}} \cdot \frac{\delta \mathcal{L}}{\delta \nabla C} = \hat{\boldsymbol{n}} \cdot \boldsymbol{J} = \text{boundary flux}, \tag{47.33}$$

with $\hat{\boldsymbol{n}}$ the boundary outward normal. 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

$$\delta \mathcal{F} = \int \delta C \left[\frac{\delta \mathcal{L}}{\delta C} - \partial_m \left(\frac{\delta \mathcal{L}}{\delta(\partial_m C)} \right) \right] \mathrm{d}^3 x. \tag{47.34}$$

Consequently, the functional derivative is given by

$$(\mathrm{d}^3 y)^{-1} \frac{\delta \mathcal{F}}{\delta C(\boldsymbol{y})} = \frac{\delta \mathcal{L}}{\delta C} - \partial_m \left[\frac{\delta \mathcal{L}}{\delta(\partial_m C)} \right], \tag{47.35}$$

where $\mathrm{d}^3 y$ is the volume element at the field point $\boldsymbol{y}$. To reach the last step required the identity

$$\frac{\delta C(\boldsymbol{x})}{\delta C(\boldsymbol{y})} = \mathrm{d}^3 y \, \delta(\boldsymbol{x} - \boldsymbol{y}), \tag{47.36}$$

where $\delta(\boldsymbol{x} - \boldsymbol{y})$ is the Dirac delta function satisfying

$$\int \delta(\boldsymbol{x} - \boldsymbol{y}) \, \mathrm{d}^3 y \, = 1, \tag{47.37}$$

so long as the integration range includes the singular point $\boldsymbol{x} = \boldsymbol{y}$. Note that the delta function has dimensions of inverse volume, which necessitates the appearance of the volume factor, $\mathrm{d}^3 y$, on the right hand side of equation (47.36). 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

$$(\mathrm{d}^3 y)^{-1} \frac{\delta \mathcal{F}}{\delta C(\boldsymbol{y})} = -\partial_m \left[\frac{\delta \mathcal{L}}{\delta(\partial_m C)} \right] = \partial_m \left(\rho\, \mathbb{K}_{mn}\, \partial_n C \right). \tag{47.38}$$

The second equality identifies the diffusion operator, thus revealing the connection between the dissipation functional, the diffusion fluxes, and the diffusion operator

$$(\mathrm{d}^3 y)^{-1} \frac{\delta \mathcal{F}}{\delta C(\boldsymbol{y})} = -\nabla \cdot \boldsymbol{J}(\boldsymbol{y}) = \mathcal{R}. \tag{47.39}$$

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 49.3 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 in equation (47.38) becomes

$$2\, \frac{\delta \mathcal{L}}{\delta(\partial_m C)} = -2\, \rho\, \mathbb{K}_{mn}\, \partial_n C - \rho\, \partial_m C\, \partial_n C\, \frac{\delta \mathbb{K}_{mn}}{\delta(\partial_m C)}. \tag{47.40}$$

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 47.8.

47.4 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 47.2, 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.

47.4.1 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*

$$\frac{\mathrm{D}C}{\mathrm{D}t} = \frac{\partial C}{\partial t} + \boldsymbol{v} \cdot \nabla C = 0. \tag{47.41}$$

The first equality relates the material time derivative to the Eulerian time derivative plus advective transport (see Section 12.5), with $\boldsymbol{v}$ the velocity of a fluid element. We can convert the *material* form of the advection equation (47.41) into its flux-form by combining with the mass continuity equation (14.9)

$$\frac{\partial \rho}{\partial t} + \nabla \cdot (\rho\, \boldsymbol{v}) = 0, \tag{47.42}$$

which yields

$$\frac{\partial (\rho\, C)}{\partial t} + \nabla \cdot (\rho\, C\, \boldsymbol{v}) = 0. \tag{47.43}$$

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

$$C(\boldsymbol{x}, t) = C[\boldsymbol{X}(0)], \tag{47.44}$$

where $\boldsymbol{X}(0)$ is the initial position of a fluid element that is at the position $\boldsymbol{x}$ 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.

47.4.2 Eulerian time tendencies from advection

At a point in the fluid, the advection equation (47.41) leads to the Eulerian time tendency for tracer concentration

$$\frac{\partial C}{\partial t} = -\boldsymbol{v} \cdot \nabla C. \tag{47.45}$$

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 (47.43), the

density-weighted tracer concentration (the tracer mass per volume) has an Eulerian time tendency given by the convergence of the advective flux

$$\frac{\partial\,(\rho\,C)}{\partial t} = -\nabla\cdot(\rho\,C\,\boldsymbol{v}). \tag{47.46}$$

The tendency vanishes at a point if there is no convergence of tracer mass towards the point.

47.4.3 Impermeability property of tracer isosurfaces

We now offer a geometric interpretation of the advection equation

$$\frac{\partial C}{\partial t} + \boldsymbol{v}\cdot\nabla C = 0, \tag{47.47}$$

following the discussion of dia-surface transport in Section 17.3. For this purpose, introduce the unit normal on a tracer isosurface

$$\hat{\boldsymbol{n}} = \frac{\nabla C}{|\nabla C|} \tag{47.48}$$

and the normal projection for the velocity of a point on that surface

$$\boldsymbol{v}^{(b)}\cdot\hat{\boldsymbol{n}} = -\frac{\partial_t C}{|\nabla C|}. \tag{47.49}$$

The advection equation (47.47) thus can be written as an impermeability condition for a tracer isosurface

$$\rho\,(\boldsymbol{v}-\boldsymbol{v}^{(b)})\cdot\hat{\boldsymbol{n}} = 0 \qquad \text{on } C \text{ isosurfaces.} \tag{47.50}$$

We encountered this condition in Section 14.4.2 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 $\boldsymbol{v}$, to cross them. This is an important kinematic result that will be extended in Section 47.7.5 to include effects from an eddy induced velocity.

47.5 Advection maths

We now explore various mathematical properties of the advection equation. For that purpose, recall the mass continuity equation (47.42) and flux-form tracer advection equation (47.46)

$$\frac{\partial\rho}{\partial t} + \nabla\cdot(\rho\,\boldsymbol{v}) = 0 \tag{47.51a}$$

$$\frac{\partial(\rho\,C)}{\partial t} + \nabla\cdot(\rho\,C\,\boldsymbol{v}) = 0. \tag{47.51b}$$

These equations are manifestly compatible in that the tracer equation (47.51b) reduces to the continuity equation (47.51a) if the tracer concentration is spatially uniform (see Section 15.1.5 for more discussion of compatibility).

47.5.1 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

$$\frac{\mathrm{D}C^\Gamma}{\mathrm{D}t} = \Gamma\, C^{\Gamma-1}\frac{\mathrm{D}C}{\mathrm{D}t} = 0. \tag{47.52}$$

Likewise, making use of the Eulerian form yields

$$\frac{\partial C^\Gamma}{\partial t} + \boldsymbol{v}\cdot\nabla C^\Gamma = \Gamma\, C^{\Gamma-1}\left[\frac{\partial C}{\partial t} + \boldsymbol{v}\cdot\nabla C\right] = 0. \tag{47.53}$$

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 47.3.3.

47.5.2 Mass transport

The mass density time tendency

$$\frac{\partial \rho}{\partial t} = -\nabla\cdot(\boldsymbol{v}\,\rho) \tag{47.54}$$

remains unchanged if the advective mass flux, $\rho\,\boldsymbol{v}$ (dimensions of mass per time per area), is modified by the addition of a total curl

$$\rho\,\boldsymbol{v} \to \rho\,\boldsymbol{v}^\dagger = \rho\,\boldsymbol{v} + \nabla\wedge(\rho\,\boldsymbol{\Psi}^*). \tag{47.55}$$

As in Section 16.5.2, the arbitrariness manifest in equation (47.55) is known as a *gauge symmetry*. The additional mass flux, $\nabla\wedge(\rho\,\boldsymbol{\Psi}^*)$, 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 $\nabla\wedge\boldsymbol{\Psi}^*$ leads to zero accumulation of volume at a point.

The non-divergent mass flux

$$\rho\,\boldsymbol{v}^* \equiv \nabla\wedge(\rho\,\boldsymbol{\Psi}^*) \tag{47.56}$$

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:

$$\boldsymbol{v} = \text{Eulerian mean velocity} \tag{47.57a}$$
$$\rho\,\boldsymbol{v} = \text{Eulerian mean mass flux} \tag{47.57b}$$
$$\boldsymbol{v}^* = \text{eddy-induced velocity} \tag{47.57c}$$
$$\rho\,\boldsymbol{\Psi}^* = \text{eddy-induced mass streamfunction} \tag{47.57d}$$
$$\rho\,\boldsymbol{v}^* = \nabla\wedge(\rho\,\boldsymbol{\Psi}^*) = \text{eddy-induced mass flux} \tag{47.57e}$$
$$\boldsymbol{v}^\dagger = \boldsymbol{v} + \boldsymbol{v}^* = \text{residual mean velocity} \tag{47.57f}$$
$$\rho\,\boldsymbol{v}^\dagger = \rho\,(\boldsymbol{v} + \boldsymbol{v}^*) = \text{residual mean mass flux.} \tag{47.57g}$$

The name "residual mean" is motivated since the sum $\boldsymbol{v} + \boldsymbol{v}^*$ 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 49.

47.5.3 Advective 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

$$\frac{\partial(\rho\, C)}{\partial t} + \nabla\cdot(\rho\, C\, \boldsymbol{v}^\dagger) = 0. \tag{47.58}$$

Given the form (47.56) for the eddy mass flux $\rho\,\boldsymbol{v}^*$, we can write the advective tracer flux as

$$\begin{aligned}
\rho\, C\, \boldsymbol{v}^\dagger &= C\,(\rho\,\boldsymbol{v} + \rho\,\boldsymbol{v}^*) && \text{(47.59a)}\\
&= C\,\rho\,\boldsymbol{v} + C\,\nabla\wedge(\rho\,\boldsymbol{\Psi}^*) && \text{(47.59b)}\\
&= C\,\rho\,\boldsymbol{v} + \nabla\wedge(C\,\rho\,\boldsymbol{\Psi}^*) - \nabla C\wedge\rho\,\boldsymbol{\Psi}^*. && \text{(47.59c)}
\end{aligned}$$

It is the divergence of the tracer flux that determines the time tendency, in which case the total curl plays no role

$$\begin{aligned}
\nabla\cdot(\rho\, C\, \boldsymbol{v}^\dagger) &= \nabla\cdot(\rho\, C\, \boldsymbol{v} + \rho\, C\, \boldsymbol{v}^*) && \text{(47.60a)}\\
&= \nabla\cdot(\rho\, C\, \boldsymbol{v} - \nabla C\wedge\rho\,\boldsymbol{\Psi}^*). && \text{(47.60b)}
\end{aligned}$$

That is, the divergence of the advective mass flux equals to the divergence of the skew tracer flux

$$\underbrace{\nabla\cdot(\rho\, C\, \boldsymbol{v}^*)}_{\text{advective flux divergence}} = \underbrace{\nabla\cdot(-\nabla C\wedge\rho\,\boldsymbol{\Psi}^*)}_{\text{skew flux divergence}} \tag{47.61}$$

since the advective flux and skew flux differ by a rotational flux

$$\boldsymbol{J}^{\text{adv}} = \boldsymbol{J}^{\text{skew}} + \boldsymbol{J}^{\text{rot}} \tag{47.62}$$

where

$$\boldsymbol{J}^{\text{adv}} = C\,\rho\,\boldsymbol{v}^* \qquad \text{and} \qquad \boldsymbol{J}^{\text{skew}} = -\nabla C\wedge\rho\,\boldsymbol{\Psi}^* \qquad \text{and} \qquad \boldsymbol{J}^{\text{rot}} = \nabla\wedge(\rho\, C\,\boldsymbol{\Psi}^*). \tag{47.63}$$

Notably, the skew tracer flux is neither upgradient nor downgradient. Rather, it is oriented parallel to iso-surfaces of tracer concentration

$$\nabla C\cdot\boldsymbol{J}^{\text{skew}} = \nabla C\cdot(-\nabla C\wedge\rho\,\boldsymbol{\Psi}^*) = 0. \tag{47.64}$$

This orientation serves as motivation for the name "skew." Figure 47.3 provides a schematic of the skew tracer fluxes.

47.5.4 Skew diffusion

Introducing tensor labels brings the skew tracer flux into the form

$$\begin{aligned}
J^{\text{skew}}_m &= -(\nabla C\wedge\rho\,\boldsymbol{\Psi}^*)_m && \text{(47.65a)}\\
&= -\epsilon_{mnp}\,\frac{\partial C}{\partial x^n}\,\rho\,\Psi^*_p && \text{(47.65b)}\\
&= -\rho\,\mathbb{A}_{mn}\,\frac{\partial C}{\partial x^n}, && \text{(47.65c)}
\end{aligned}$$

Figure 47.3: *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).*

where we defined the anti-symmetric *skew diffusion* tensor

$$\mathbb{A}_{mn} = \epsilon_{mnp}\,\Psi^*_p = \begin{bmatrix} 0 & \Psi^*_3 & -\Psi^*_2 \\ -\Psi^*_3 & 0 & \Psi^*_1 \\ \Psi^*_2 & -\Psi^*_1 & 0 \end{bmatrix}. \tag{47.66}$$

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 48.3 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 47.6.3.

In Section 47.6 we pursue the above points to further detail the dual relation between advection and skewsion.

47.5.5 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 47.6 and 48.3. 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.

47.5.6 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.

47.6 Advection and skewsion

We introduced skew diffusion in Section 47.5.4 and will again encounter it in Chapterd 48 and 49. Following the terminology of Section 9.2 of *Griffies* (2004), we refer to *skewsion* 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 49.1. We here consider facets of advection and skewsion for those interested in diving deeper into the details.

47.6.1 Choosing a gauge

Consider an arbitrary divergent-free mass transport

$$\nabla \cdot (\rho \, \boldsymbol{v}^*) = 0, \tag{47.67}$$

where the divergent-free constraint is satisfied by introducing a vector streamfunction

$$\rho \, \boldsymbol{v}^* = \nabla \wedge (\rho \, \boldsymbol{\Psi}^*). \tag{47.68}$$

The streamfunction is arbitrary up to a gauge transformation

$$\rho \, \boldsymbol{\Psi}' = \rho \, \boldsymbol{\Psi}^* + \nabla(\rho \, \Lambda), \tag{47.69}$$

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, $\boldsymbol{J}^{\text{skew}} = -\nabla C \wedge \rho \, \boldsymbol{\Psi}^*$, changes. Nonetheless, the divergence of the skew flux is invariant, as we see by noting that

$$\nabla C \wedge [\rho \, \boldsymbol{\Psi}^* + \nabla(\rho \Lambda)] = \nabla C \wedge (\rho \, \boldsymbol{\Psi}^*) + \nabla \wedge [C \, \nabla(\rho \Lambda)]. \tag{47.70}$$

and since $\nabla \cdot \nabla \wedge [C \, \nabla(\rho \Lambda)] = 0$, the flux divergence, $\nabla \cdot \boldsymbol{J}^{\text{skew}}$, 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

$$\nabla \cdot (\rho\, \boldsymbol{\Psi}^*) = 0 \qquad \text{Coulomb gauge.} \tag{47.71}$$

Making use of the curl identity (2.36c) leads to the Poisson equation for the vector potential

$$\nabla^2(\rho \boldsymbol{\Psi}^*) = -\nabla \wedge (\rho\, \boldsymbol{v}^*). \tag{47.72}$$

This equation has a free-space Green's function given by the Coulomb-Ampere expression

$$\rho(\boldsymbol{x}, t)\, \boldsymbol{\Psi}^*(\boldsymbol{x}, t) = \nabla \wedge \int \frac{\rho(\boldsymbol{x}', t)\, \boldsymbol{v}^*(\boldsymbol{x}', t)}{4\,\pi\, |\boldsymbol{x} - \boldsymbol{x}'|}\, \mathrm{d}V', \tag{47.73}$$

where $\mathrm{d}V'$ is the volume element for integration over the test points, $\boldsymbol{x}'$. Although common in electrodynamics, we know of no geophysical fluid application making use of the Coulomb gauge.

47.6.2 Vertical gauge

As introduced in Section 16.5.3, a gauge commonly used for eddy parameterizations (Section 49.1) 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

$$\Psi_3^* = 0 \qquad \text{vertical gauge.} \tag{47.74}$$

Let us see how we can generally make this gauge choice. Consider a vector streamfunction $\boldsymbol{\Phi}$ that has all three components nonzero. Now consider the alternative streamfunction $\rho\, \boldsymbol{\Psi}^* = \rho\, \boldsymbol{\Phi} + \nabla(\rho\Lambda)$, with $\nabla(\rho\Lambda) = -\hat{\boldsymbol{z}}\, \rho\, \Phi_3$. This choice in turn means that the third component of $\boldsymbol{\Psi}^*$ is zero.

To further specify the vertical gauge we invert the relations

$$\rho\, u^* = -\partial_z(\rho\Psi_2^*) \qquad \text{and} \qquad \rho\, v^* = \partial_z\, (\rho\Psi_1^*) \qquad \text{and} \qquad \rho\, w^* = \partial_x\, (\rho\Psi_2^*) - \partial_y(\rho\Psi_1^*) \tag{47.75}$$

to render the vector streamfunction

$$\rho\, \boldsymbol{\Psi}^* = \hat{\boldsymbol{z}} \wedge \int_{-H}^{z} \rho\, \boldsymbol{u}^*\, \mathrm{d}z' = \hat{\boldsymbol{z}} \wedge \underline{\boldsymbol{U}}^{(*\rho)} \tag{47.76}$$

where

$$\underline{\boldsymbol{U}}^{(*\rho)}(z) = \int_{-H}^{z} \rho\, \boldsymbol{u}^*\, \mathrm{d}z' \tag{47.77}$$

is the horizontal mass transport associated with $\boldsymbol{u}^*$ passing between the bottom and a depth $z \geq -H$. The anti-symmetric stirring tensor for the vertical gauge is given by

$$\rho\, \mathbb{A}_{mn} = \begin{pmatrix} 0 & 0 & \underline{U}^{(*\rho)} \\ 0 & 0 & \underline{V}^{(*\rho)} \\ -\underline{U}^{(*\rho)} & -\underline{V}^{(*\rho)} & 0 \end{pmatrix}, \tag{47.78}$$

and the corresponding skew, rotational, and advective fluxes are

$$\boldsymbol{J}^{\text{skew}} = -\underline{\boldsymbol{U}}^{(*\rho)}\,\partial_z C + \hat{\boldsymbol{z}}\,\underline{\boldsymbol{U}}^{(*\rho)} \cdot \nabla_z C \tag{47.79a}$$

$$\boldsymbol{J}^{\text{rot}} = \partial_z(C\,\underline{\boldsymbol{U}}^{(*\rho)}) - \hat{\boldsymbol{z}}\,\nabla_z \cdot (C\,\underline{\boldsymbol{U}}^{(*\rho)}) \tag{47.79b}$$

$$\boldsymbol{J}^{\text{adv}} = C\,(\partial_z \underline{\boldsymbol{U}}^{(*\rho)}) - \hat{\boldsymbol{z}}\,C\,\nabla_z \cdot \underline{\boldsymbol{U}}^{(*\rho)}. \tag{47.79c}$$

Note that the identity $\boldsymbol{J}^{\text{adv}} = \boldsymbol{J}^{\text{skew}} + \boldsymbol{J}^{\text{rot}}$ 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.

47.6.3 Boundary conditions

We assume that all external domain boundaries are material in regards to the velocity $\boldsymbol{v}^*$. Furthermore, even for moving domain boundaries, we assume that the suite of kinematic boundary conditions is based on the barycentric velocity $\boldsymbol{v}$ (see Section 14.4), so that $\boldsymbol{v}^*$ satisfies the no-normal flow condition even on moving boundaries

$$\hat{\boldsymbol{n}} \cdot \boldsymbol{v}^* = 0 \qquad \text{external domain boundaries.} \tag{47.80}$$

As we discuss in Section 47.7.1, 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

$$\hat{\boldsymbol{n}} \cdot \boldsymbol{J}^{\text{adv}} = \hat{\boldsymbol{n}} \cdot \boldsymbol{v}^* \rho\, C = 0. \tag{47.81}$$

The corresponding boundary condition for the skew flux is found by inserting the relation (47.62) into the advective flux boundary condition (47.81) to render

$$\hat{\boldsymbol{n}} \cdot \boldsymbol{J}^{\text{adv}} = \hat{\boldsymbol{n}} \cdot [\boldsymbol{J}^{\text{skew}} + \boldsymbol{J}^{\text{rot}}] = 0. \tag{47.82}$$

Hence, the skew flux generally has a non-zero normal component at the solid boundaries as determined by the rotational flux

$$\hat{\boldsymbol{n}} \cdot \boldsymbol{J}^{\text{skew}} = -\hat{\boldsymbol{n}} \cdot \boldsymbol{J}^{\text{rot}}. \tag{47.83}$$

Even so, there might be occasions in which $\hat{\boldsymbol{n}} \cdot \boldsymbol{J}^{\text{skew}} = 0$, which is ensured so long as

$$(-\nabla C \wedge \rho \boldsymbol{\Psi}^*) \cdot \hat{\boldsymbol{n}} = -(\rho\, \boldsymbol{\Psi}^* \wedge \hat{\boldsymbol{n}}) \cdot \nabla C = 0. \tag{47.84}$$

A sufficient condition is to have $\boldsymbol{\Psi}^* \wedge \hat{\boldsymbol{n}} = 0$, 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 $\boldsymbol{v}^*$. We discuss one example in Section 49.1 as prescribed by the *Gent et al.* (1995) mesoscale eddy parameterization.

47.7 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

$$\frac{\partial \rho}{\partial t} + \nabla \cdot (\rho \, \boldsymbol{v}^{\dagger}) = 0 \tag{47.85a}$$

$$\frac{\partial (\rho \, C)}{\partial t} + \nabla \cdot (\rho \, \boldsymbol{v}^{\dagger} \, C + \boldsymbol{J}^{\text{diff}}) = 0, \tag{47.85b}$$

where (see Section 47.6)

$$\boldsymbol{v}^{\dagger} = \boldsymbol{v} + \boldsymbol{v}^{*} \qquad \text{and} \qquad \nabla \cdot (\rho \, \boldsymbol{v}^{*}) = 0, \tag{47.86}$$

and where $\boldsymbol{J}^{\text{diff}}$ 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 $\nabla \cdot (\rho \, \boldsymbol{v}^{*}) = 0$, the mass budget (47.85a) can we written in the equivalent manners

$$\frac{\partial \rho}{\partial t} + \nabla \cdot (\rho \, \boldsymbol{v}^{\dagger}) = \frac{\partial \rho}{\partial t} + \nabla \cdot (\rho \, \boldsymbol{v}) = 0. \tag{47.87}$$

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, $\boldsymbol{v}^{\dagger}$. That result is not surprising, since the finite volume budgets are consistent with the differential budgets (47.85a) and (47.85b). 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 51. 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 47.6.3.

47.7.1 Advective formulation

Making use of the tracer equation (47.85b) in the Leibniz-Reynolds transport theorem (15.35) renders the finite volume tracer mass budget for an arbitrary domain, $\mathcal{R}$

$$\frac{\mathrm{d}}{\mathrm{d}t} \left[\int_{\mathcal{R}} \rho \, C \, \mathrm{d}V \right] = - \oint_{\partial \mathcal{R}} \left[\rho \, C \, (\boldsymbol{v}^{\dagger} - \boldsymbol{v}^{(b)}) + \boldsymbol{J}^{\text{diff}} \right] \cdot \hat{\boldsymbol{n}} \, \mathrm{d}\mathcal{S}, \tag{47.88}$$

where $\boldsymbol{v}^{(b)}$ is the velocity of a point on the domain boundary. Appearance of the residual mean velocity, $\boldsymbol{v}^{\dagger}$, in the finite volume budget (47.88) follows from its appearance in the local tracer budget (47.85b). 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

$$\boldsymbol{v}^{*} \cdot \hat{\boldsymbol{n}} = 0 \qquad \text{on external fluid boundaries.} \tag{47.89}$$

We already encountered this boundary condition in Section 47.6.3. 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 49.2).

Setting the tracer concentration to a constant in equation (47.88) leads to the fluid mass budget

$$\frac{\mathrm{d}}{\mathrm{d}t} \left[\int_{\mathcal{R}} \rho \, \mathrm{d}V \right] = - \oint_{\partial \mathcal{R}} \rho \, (\boldsymbol{v}^{\dagger} - \boldsymbol{v}^{(b)}) \cdot \hat{\boldsymbol{n}} \, \mathrm{d}\mathcal{S}, \tag{47.90}$$

where we set the diffusive tracer flux, $\boldsymbol{J}^{\text{diff}}$, to zero since there is no diffusion of fluid mass between fluid elements (Section 15.1). As for the differential form (47.85a), the mass budget for any domain is not changed by the eddy-induced velocity since

$$\nabla \cdot (\rho \, \boldsymbol{v}^*) = 0 \Longrightarrow \oint_{\partial \mathcal{R}} \rho \, \boldsymbol{v}^* \cdot \hat{\boldsymbol{n}} \, \mathrm{d}\mathcal{S} = 0, \tag{47.91}$$

so that the mass budget is given by

$$\frac{\mathrm{d}}{\mathrm{d}t} \left[\int_{\mathcal{R}} \rho \, \mathrm{d}V \right] = - \oint_{\partial \mathcal{R}} \left[\rho \, (\boldsymbol{v}^{\dagger} - \boldsymbol{v}^{(b)}) \right] \cdot \hat{\boldsymbol{n}} \, \mathrm{d}\mathcal{S} = - \oint_{\partial \mathcal{R}} \left[\rho \, (\boldsymbol{v} - \boldsymbol{v}^{(b)}) \right] \cdot \hat{\boldsymbol{n}} \, \mathrm{d}\mathcal{S} \tag{47.92}$$

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 $\boldsymbol{v}^* \neq 0$, the accumulation of mass within the region are identical for the two formulations.

47.7.2 Skew flux formulation

Now consider the complement perspective afforded by the skew flux formulation from Section 47.6. Here we decompose the advective tracer flux according to

$$C \, \rho \, \boldsymbol{v}^{\dagger} = C \, \rho \, \boldsymbol{v} - \nabla C \wedge \rho \, \boldsymbol{\Psi}^* + \nabla \wedge (C \, \rho \, \boldsymbol{\Psi}^*) = C \, \rho \, \boldsymbol{v} + \boldsymbol{J}^{\text{skew}} + \nabla \wedge (C \, \rho \, \boldsymbol{\Psi}^*), \tag{47.93}$$

where we introduced the skew tracer flux arising from the eddy-induced streamfunction

$$\boldsymbol{J}^{\text{skew}} = -\nabla C \wedge (\rho \, \boldsymbol{\Psi}^*). \tag{47.94}$$

The differential budget for tracer is thus given by

$$\frac{\partial (\rho \, C)}{\partial t} + \nabla \cdot [\rho \, C \, \boldsymbol{v} + \boldsymbol{J}^{\text{skew}} + \boldsymbol{J}^{\text{diff}}] = 0, \tag{47.95}$$

where the rotational term, $\nabla \wedge (C \, \rho \, \boldsymbol{\Psi}^*)$, has zero divergence and thus does not affect the tracer budget. The corresponding finite volume tracer mass budget is

$$\frac{\mathrm{d}}{\mathrm{d}t} \left[\int_{\mathcal{R}} \rho \, C \, \mathrm{d}V \right] = - \oint_{\partial \mathcal{R}} \left[\rho \, C \, (\boldsymbol{v} - \boldsymbol{v}^{(b)}) - \nabla C \wedge (\rho \, \boldsymbol{\Psi}^*) + \boldsymbol{J}^{\text{diff}} \right] \cdot \hat{\boldsymbol{n}} \, \mathrm{d}\mathcal{S}. \tag{47.96}$$

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 (47.92).

47.7.3 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 51.5. 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 47.4.

Figure 47.4: *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 $\partial\Omega_{\rm surf}(\tilde{C})$; at the bottom by the $C = \tilde{C}$ isosurface as well as the solid-earth bottom $\partial\Omega_{\rm bot}(\tilde{C})$. 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 47.7.4, whereas the budget in the full region $C \geq \tilde{C}$ is considered in Section 47.7.3.*

Advective formulation

The tracer mass budget written using the advective formulation (47.88) takes the form

$$\frac{\mathrm{d}}{\mathrm{d}t}\left[\int_{\mathcal{R}} \rho\, C\, \mathrm{d}V\right] = \int_{\partial\Omega_{\rm surf}(\tilde{C})} Q_{\rm m}\, C\, \mathrm{d}A - \tilde{C} \int_{C=\tilde{C}} \rho\, (\boldsymbol{v}^{\dagger} - \boldsymbol{v}^{(b)}) \cdot \hat{\boldsymbol{n}}\, \mathrm{d}\mathcal{S} - \oint_{\partial\mathcal{R}} \boldsymbol{J}^{\rm diff} \cdot \hat{\boldsymbol{n}}\, \mathrm{d}\mathcal{S}. \tag{47.97}$$

For the first right hand side term we made use of the surface kinematic boundary condition (14.70c), with $Q_{\rm m}\, \mathrm{d}A$ the mass per time crossing the surface interface and where $\mathrm{d}A$ is the horizontal projection of the interface area element $\mathrm{d}\mathcal{S}$. We also made use of the exterior boundary condition (47.89) 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

$$\frac{\mathrm{d}}{\mathrm{d}t}\left[\int_{\mathcal{R}} \rho\, \mathrm{d}V\right] = \int_{\partial\Omega_{\rm surf}(\tilde{C})} Q_{\rm m}\, \mathrm{d}A - \int_{C=\tilde{C}} \rho\, (\boldsymbol{v}^{\dagger} - \boldsymbol{v}^{(b)}) \cdot \hat{\boldsymbol{n}}\, \mathrm{d}\mathcal{S}. \tag{47.98}$$

Combining this budget with the tracer mass budget allows us to write

$$\frac{\mathrm{d}}{\mathrm{d}t}\left[M_C - \tilde{C}\, M\right] = \int_{\partial\Omega_{\rm surf}(\tilde{C})} Q_{\rm m}\, (C - \tilde{C})\, \mathrm{d}A - \oint_{\partial\mathcal{R}} \boldsymbol{J}^{\rm diff} \cdot \hat{\boldsymbol{n}}\, \mathrm{d}\mathcal{S}, \tag{47.99}$$

where we introduced the shorthand for the tracer mass and fluid mass in the region

$$M_C = \int_{\mathcal{R}} C\, \rho\, \mathrm{d}V \qquad \text{and} \qquad M = \int_{\mathcal{R}} \rho\, \mathrm{d}V. \tag{47.100}$$

In Section 51.5.2 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 (47.96). It takes on the following form for the domain in Figure 47.4

$$\frac{\mathrm{d}}{\mathrm{d}t}\left[\int_{\mathcal{R}} \rho\, C\, \mathrm{d}V\right] = \int_{\partial\Omega_{\mathrm{surf}}(\tilde{C})} Q_{\mathrm{m}}\, C\, \mathrm{d}A - \tilde{C}\int_{C=\tilde{C}} \rho\,(\boldsymbol{v}-\boldsymbol{v}^{(b)})\cdot\hat{\boldsymbol{n}}\,\mathrm{d}\mathcal{S} - \oint_{\partial\mathcal{R}} [-\nabla C\wedge(\rho\,\boldsymbol{\Psi}^*) + \boldsymbol{J}^{\mathrm{diff}}]\cdot\hat{\boldsymbol{n}}\,\mathrm{d}\mathcal{S}, \tag{47.101}$$

and the corresponding budget for the fluid mass is

$$\frac{\mathrm{d}}{\mathrm{d}t}\left[\int_{\mathcal{R}} \rho\,\mathrm{d}V\right] = \int_{\partial\Omega_{\mathrm{surf}}(\tilde{C})} Q_{\mathrm{m}}\,\mathrm{d}A - \int_{C=\tilde{C}} \rho\,(\boldsymbol{v}-\boldsymbol{v}^{(b)})\cdot\hat{\boldsymbol{n}}\,\mathrm{d}\mathcal{S}. \tag{47.102}$$

As for the advective formulation, we combine the fluid mass budget equation (47.102) with the tracer mass equation (47.101) to render a budget equation for the internal mass content

$$\frac{\mathrm{d}}{\mathrm{d}t}\left[M_C - \tilde{C}\,M\right] = \int_{\partial\Omega_{\mathrm{surf}}(\tilde{C})} Q_{\mathrm{m}}\,(C-\tilde{C})\,\mathrm{d}A - \oint_{\partial\mathcal{R}} \boldsymbol{J}^{\mathrm{diff}}\cdot\hat{\boldsymbol{n}}\,\mathrm{d}\mathcal{S}, \tag{47.103}$$

which is the same as for the advective formulation given by equation (47.99).

Proving the budgets based on the two formulations are equivalent

The two tracer budgets, (47.97) and (47.101), must be the same since they measure changes to the tracer mass within the same region. Likewise, the two mass budgets, (47.98) and (47.102), must be the same, as are the two internal tracer mass budgets (47.99) and (47.103). We here expose the manipulations required to verify these equalities.

To prove the $C=\tilde{C}$ terms in the tracer budget equations (47.97) and (47.101) are the same, consider the identity (47.91) applied to the region under consideration

$$0 = \oint_{\partial\mathcal{R}} \rho\,\boldsymbol{v}^*\cdot\hat{\boldsymbol{n}}\,\mathrm{d}\mathcal{S} = \int_{\partial\Omega_{\mathrm{surf}}(\tilde{C})} \rho\,\boldsymbol{v}^*\cdot\hat{\boldsymbol{n}}\,\mathrm{d}\mathcal{S} + \int_{\partial\Omega_{\mathrm{bot}}(\tilde{C})} \rho\,\boldsymbol{v}^*\cdot\hat{\boldsymbol{n}}\,\mathrm{d}\mathcal{S} + \int_{C=\tilde{C}} \rho\,\boldsymbol{v}^*\cdot\hat{\boldsymbol{n}}\,\mathrm{d}\mathcal{S}. \tag{47.104}$$

The surface and bottom boundary terms vanish due to the external boundary condition (47.89); namely, $\boldsymbol{v}^*\cdot\hat{\boldsymbol{n}}=0$ for each point along an external fluid boundary. We are thus led to conclude that

$$\int_{C=\tilde{C}} \rho\,\boldsymbol{v}^*\cdot\hat{\boldsymbol{n}}\,\mathrm{d}\mathcal{S} = 0. \tag{47.105}$$

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 $\boldsymbol{v}^*\cdot\hat{\boldsymbol{n}}\neq 0$ at any particular point on the $C=\tilde{C}$ surface. Given the boundary integral (47.105) we are thus led to conclude

$$\int_{C=\tilde{C}} \rho\,(\boldsymbol{v}^\dagger - \boldsymbol{v}^{(b)})\cdot\hat{\boldsymbol{n}}\,\mathrm{d}\mathcal{S} = \int_{C=\tilde{C}} \rho\,(\boldsymbol{v}-\boldsymbol{v}^{(b)})\cdot\hat{\boldsymbol{n}}\,\mathrm{d}\mathcal{S}. \tag{47.106}$$

This identity proves that the two mass budgets (47.98) and (47.102) are indeed measuring changes to the same fluid mass, even though one computes the domain boundary fluxes based on the residual mean velocity, $\boldsymbol{v}^\dagger$, whereas the other uses the Eulerian mean, $\boldsymbol{v}$.

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

$$\int_{C=\tilde{C}} [\nabla C\wedge(\rho\,\boldsymbol{\Psi}^*)]\cdot\hat{\boldsymbol{n}}\,\mathrm{d}\mathcal{S} = 0, \tag{47.107}$$

which follows since $\hat{\boldsymbol{n}}$ is parallel to ∇C along this boundary. For the external boundaries, equality of the tracer mass budgets (47.97) and (47.101) is satisfied for an arbitrary tracer concentration if one of the boundary conditions discussed in Section (47.6.3) is maintained; i.e., if $\boldsymbol{\Psi}^*$ vanishes on an external boundary or if it is parallel to the boundary normal direction ($\hat{\boldsymbol{n}} \wedge \boldsymbol{\Psi}^* = 0$). Maintenance of either of these two boundary conditions allows us to conclude that the two budgets (47.97) and (47.101) are indeed identical.

47.7.4 Budget for a region with interior sides

Consider the subregion $\mathcal{R}_{\text{sub}}$ shown in Figure 47.4. 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 47.7.3. Additionally, region $\mathcal{R}_{\text{sub}}$ 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 $\mathcal{R}_{\text{sub}}$ can be formulated using either the residual mean velocity or the Eulerian mean velocity

$$\frac{\mathrm{d}}{\mathrm{d}t}\left[\int_{\mathcal{R}_{\text{sub}}} \rho \,\mathrm{d}V\right] = \int_{\partial\Omega_{\text{surf}}(\tilde{C})} Q_{\text{m}} \,\mathrm{d}A - \int_{C=\tilde{C}} \rho\,(\boldsymbol{v}^{\dagger} - \boldsymbol{v}^{(b)}) \cdot \hat{\boldsymbol{n}} \,\mathrm{d}\mathcal{S} - \int_{\text{sides}} \rho\,(\boldsymbol{v}^{\dagger} - \boldsymbol{v}^{(b)}) \cdot \hat{\boldsymbol{n}} \,\mathrm{d}\mathcal{S} \tag{47.108a}$$

$$\frac{\mathrm{d}}{\mathrm{d}t}\left[\int_{\mathcal{R}_{\text{sub}}} \rho \,\mathrm{d}V\right] = \int_{\partial\Omega_{\text{surf}}(\tilde{C})} Q_{\text{m}} \,\mathrm{d}A - \int_{C=\tilde{C}} \rho\,(\boldsymbol{v} - \boldsymbol{v}^{(b)}) \cdot \hat{\boldsymbol{n}} \,\mathrm{d}\mathcal{S} - \int_{\text{sides}} \rho\,(\boldsymbol{v} - \boldsymbol{v}^{(b)}) \cdot \hat{\boldsymbol{n}} \,\mathrm{d}\mathcal{S}. \tag{47.108b}$$

The two budgets are identical since the eddy velocity satisfies $\oint \rho\, \boldsymbol{v}^* \cdot \hat{\boldsymbol{n}} \,\mathrm{d}\mathcal{S} = 0$ for any domain, as well as $\boldsymbol{v}^* \cdot \hat{\boldsymbol{n}} = 0$ along any external domain boundary. Hence, as already noted in Section 47.7.1, the eddy velocity contribution to the right hand side of equation (47.108a) 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

$$\frac{\mathrm{d}}{\mathrm{d}t}\left[\int_{\mathcal{R}_{\text{sub}}} \rho\, C \,\mathrm{d}V\right] = \int_{\partial\Omega_{\text{surf}}(\tilde{C})} Q_{\text{m}}\, C \,\mathrm{d}A - \tilde{C} \int_{C=\tilde{C}} \rho\,(\boldsymbol{v}^{\dagger} - \boldsymbol{v}^{(b)}) \cdot \hat{\boldsymbol{n}} \,\mathrm{d}\mathcal{S} - \oint_{\partial\mathcal{R}_{\text{sub}}} \boldsymbol{J}^{\text{diff}} \cdot \hat{\boldsymbol{n}} \,\mathrm{d}\mathcal{S} - \int_{\text{sides}} C\,\rho\,(\boldsymbol{v}^{\dagger} - \boldsymbol{v}^{(b)}) \cdot \hat{\boldsymbol{n}} \,\mathrm{d}\mathcal{S}, \tag{47.109}$$

and the corresponding skew flux formulation is

$$\frac{\mathrm{d}}{\mathrm{d}t}\left[\int_{\mathcal{R}_{\text{sub}}} \rho\, C \,\mathrm{d}V\right] = \int_{\partial\Omega_{\text{surf}}(\tilde{C})} Q_{\text{m}}\, C \,\mathrm{d}A - \tilde{C} \int_{C=\tilde{C}} \rho\,(\boldsymbol{v} - \boldsymbol{v}^{(b)}) \cdot \hat{\boldsymbol{n}} \,\mathrm{d}\mathcal{S} - \oint_{\partial\mathcal{R}_{\text{sub}}} \boldsymbol{J}^{\text{diff}} \cdot \hat{\boldsymbol{n}} \,\mathrm{d}\mathcal{S} - \int_{\text{sides}} C\,\rho\,(\boldsymbol{v} - \boldsymbol{v}^{(b)}) \cdot \hat{\boldsymbol{n}} \,\mathrm{d}\mathcal{S} - \int_{\text{sides}} [-\nabla C \wedge (\rho\, \boldsymbol{\Psi}^*)] \cdot \hat{\boldsymbol{n}} \,\mathrm{d}\mathcal{S}. \tag{47.110}$$

As for the discussion in Section 47.7.3, we introduce the internal tracer mass and make use of the fluid mass budgets (47.108a) and (47.108b) to write the advective form of the internal mass budget

$$\frac{\mathrm{d}}{\mathrm{d}t}\left[M_C - \tilde{C}\, M\right] = \int_{\partial\Omega_{\mathrm{surf}}(\tilde{C})} Q_{\mathrm{m}}\,(C-\tilde{C})\,\mathrm{d}A - \oint_{\partial\mathcal{R}} \boldsymbol{J}^{\mathrm{diff}}\cdot\hat{\boldsymbol{n}}\,\mathrm{d}\mathcal{S} - \int_{\mathrm{sides}} (C-\tilde{C})\,\rho\,(\boldsymbol{v}^{\dagger}-\boldsymbol{v}^{(b)})\cdot\hat{\boldsymbol{n}}\,\mathrm{d}\mathcal{S}, \tag{47.111}$$

and the corresponding skew flux form of the same budget

$$\frac{\mathrm{d}}{\mathrm{d}t}\left[M_C - \tilde{C}\, M\right] = \int_{\partial\Omega_{\mathrm{surf}}(\tilde{C})} Q_{\mathrm{m}}\,(C-\tilde{C})\,\mathrm{d}A - \oint_{\partial\mathcal{R}} \boldsymbol{J}^{\mathrm{diff}}\cdot\hat{\boldsymbol{n}}\,\mathrm{d}\mathcal{S} - \int_{\mathrm{sides}} (C-\tilde{C})\,\rho\,(\boldsymbol{v}-\boldsymbol{v}^{(b)})\cdot\hat{\boldsymbol{n}}\,\mathrm{d}\mathcal{S} - \int_{\mathrm{sides}} [-\nabla C\wedge(\rho\,\boldsymbol{\Psi}^{*})]\cdot\hat{\boldsymbol{n}}\,\mathrm{d}\mathcal{S}. \tag{47.112}$$

We now examine the right hand side of the budgets (47.111) and (47.112) to show they are indeed measuring the same tracer mass budget. For that purpose, consider the skew flux integral in equation (47.112) 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

$$-\int_{\mathrm{sides}} [-\nabla C\wedge(\rho\,\boldsymbol{\Psi}^{*})]\cdot\hat{\boldsymbol{n}}\,\mathrm{d}\mathcal{S} = -\oint_{\partial\mathcal{R}_{\mathrm{sub}}} [-\nabla C\wedge(\rho\,\boldsymbol{\Psi}^{*})]\cdot\hat{\boldsymbol{n}}\,\mathrm{d}\mathcal{S}, \tag{47.113}$$

where the right hand side is an integral around the full domain boundaries. Now reintroduce the eddy induced velocity and rotational flux to have

$$-\oint_{\partial\mathcal{R}_{\mathrm{sub}}} [-\nabla C\wedge(\rho\,\boldsymbol{\Psi}^{*})]\cdot\hat{\boldsymbol{n}}\,\mathrm{d}\mathcal{S} = -\oint_{\partial\mathcal{R}_{\mathrm{sub}}} [C\,\rho\,\boldsymbol{v}^{*} - \nabla\wedge(C\,\rho\,\boldsymbol{\Psi}^{*})]\cdot\hat{\boldsymbol{n}}\,\mathrm{d}\mathcal{S}. \tag{47.114}$$

The rotational flux has zero divergence, so that Gauss's divergence theorem means that the rotational flux vanishes when integrated along the domain boundaries

$$\oint_{\partial\mathcal{R}_{\mathrm{sub}}} \nabla\wedge(C\,\rho\,\boldsymbol{\Psi}^{*})\cdot\hat{\boldsymbol{n}}\,\mathrm{d}\mathcal{S} = 0. \tag{47.115}$$

The eddy advection term in equation (47.114) vanishes on the top boundary at $z=\eta$ due to the boundary condition $\boldsymbol{v}^{*}\cdot\hat{\boldsymbol{n}}=0$, thus yielding

$$\oint_{\partial\mathcal{R}_{\mathrm{sub}}} C\,\rho\,\boldsymbol{v}^{*}\cdot\hat{\boldsymbol{n}}\,\mathrm{d}\mathcal{S} = \int_{\mathrm{sides}} C\,\rho\,\boldsymbol{v}^{*}\cdot\hat{\boldsymbol{n}}\,\mathrm{d}\mathcal{S} + \int_{C=\tilde{C}} C\,\rho\,\boldsymbol{v}^{*}\cdot\hat{\boldsymbol{n}}\,\mathrm{d}\mathcal{S} \tag{47.116a}$$

$$= \int_{\mathrm{sides}} C\,\rho\,\boldsymbol{v}^{*}\cdot\hat{\boldsymbol{n}}\,\mathrm{d}\mathcal{S} + \tilde{C}\int_{C=\tilde{C}} \rho\,\boldsymbol{v}^{*}\cdot\hat{\boldsymbol{n}}\,\mathrm{d}\mathcal{S}. \tag{47.116b}$$

Again make use of the property $\nabla\cdot(\rho\,\boldsymbol{v}^{*})=0$ and $\hat{\boldsymbol{n}}\cdot\boldsymbol{v}^{*}=0$ at $z=\eta$ to write

$$0 = \oint_{\partial\mathcal{R}_{\mathrm{sub}}} \rho\,\boldsymbol{v}^{*}\cdot\hat{\boldsymbol{n}}\,\mathrm{d}\mathcal{S} = \oint_{\mathrm{sides}} \rho\,\boldsymbol{v}^{*}\cdot\hat{\boldsymbol{n}}\,\mathrm{d}\mathcal{S} + \oint_{C=\tilde{C}} \rho\,\boldsymbol{v}^{*}\cdot\hat{\boldsymbol{n}}\,\mathrm{d}\mathcal{S}, \tag{47.117}$$

which gives us

$$-\int_{\mathrm{sides}} [-\nabla C\wedge(\rho\,\boldsymbol{\Psi}^{*})]\cdot\hat{\boldsymbol{n}}\,\mathrm{d}\mathcal{S} = -\oint_{\partial\mathcal{R}_{\mathrm{sub}}} C\,\rho\,\boldsymbol{v}^{*}\cdot\hat{\boldsymbol{n}}\,\mathrm{d}\mathcal{S} = -\int_{\mathrm{sides}} (C-\tilde{C})\,\rho\,\boldsymbol{v}^{*}\cdot\hat{\boldsymbol{n}}\,\mathrm{d}\mathcal{S}. \tag{47.118}$$

Making use of this result in the skew flux formulated budget equation (47.112) brings it to the advective flux form found in equation (47.111).

We are thus led to conclude that the right hand side to equation (47.112) does indeed equal to the right hand side of equation (47.111). 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.

47.7.5 Budget for a stirred fluid in a region with interior sides

Although contained within the formalism developed in Section 47.7.4, it is revealing to specialize those budgets to the case of zero mixing, in which $\boldsymbol{J}^{\text{diff}} = 0$ and there is zero boundary mass flux, $Q_{\text{m}} = 0$. 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 skewsion.

Summary of the differential stirring formalism

As explored in this chapter, an Eulerian description of tracer stirring can arise from either advection or skewsion. 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, $\boldsymbol{v}^\dagger$

$$\rho \frac{\mathrm{D}^\dagger C}{\mathrm{D}t} = \frac{\partial(\rho\, C)}{\partial t} + \nabla \cdot (\rho\, \boldsymbol{v}^\dagger\, C) = 0. \tag{47.119}$$

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, $\boldsymbol{v}$. That is, tracer isosurfaces satisfy the residual mean impermeability condition

$$\rho\, (\boldsymbol{v}^\dagger - \boldsymbol{v}^{(b)}) \cdot \hat{\boldsymbol{n}} = 0 \qquad \text{on } C \text{ isosurfaces}, \tag{47.120}$$

with

$$\hat{\boldsymbol{n}} = \frac{\nabla C}{|\nabla C|} \qquad \text{and} \qquad \boldsymbol{v}^{(b)} \cdot \hat{\boldsymbol{n}} = -\frac{\partial_t C}{|\nabla C|}. \tag{47.121}$$

The impermeability condition (47.120) offers a geometric interpretation of the tracer equation (47.119) following from the discussion of dia-surface transport in Section 17.3. 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

$$(\boldsymbol{v}^\dagger - \boldsymbol{v}^{(b)}) \cdot \hat{\boldsymbol{n}} = 0 \Longrightarrow (\boldsymbol{v} - \boldsymbol{v}^{(b)}) \cdot \hat{\boldsymbol{n}} = -\boldsymbol{v}^* \cdot \hat{\boldsymbol{n}} \qquad \text{on } C \text{ isosurfaces}. \tag{47.122}$$

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 (47.122) 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 skewsion by the eddy induced streamfunction

$$\rho \frac{\mathrm{D}C}{\mathrm{D}t} + \nabla \cdot [-\nabla C \wedge (\rho\, \boldsymbol{\Psi}^*)] = \frac{\partial(\rho\, C)}{\partial t} + \nabla \cdot [\rho\, \boldsymbol{v}\, C - \nabla C \wedge (\rho\, \boldsymbol{\Psi}^*)] = 0. \tag{47.123}$$

In terms of the eddy streamfunction, $\rho\, \boldsymbol{\Psi}^*$, the impermeability condition (47.120) takes on the form

$$\Big[\rho\, \boldsymbol{v} + \nabla \wedge (\rho\, \boldsymbol{\Psi}^*) - \rho\, \boldsymbol{v}^{(b)}\Big] \cdot \hat{\boldsymbol{n}} = 0 \qquad \text{on } C \text{ isosurfaces}. \tag{47.124}$$

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

$$\frac{\mathrm{d}}{\mathrm{d}t}\left[\int_{\mathcal{R}_{\mathrm{sub}}} \rho \,\mathrm{d}V\right] = -\int_{C=\tilde{C}} \rho\,(\boldsymbol{v}^{\dagger} - \boldsymbol{v}^{(b)})\cdot\hat{\boldsymbol{n}}\,\mathrm{d}\mathcal{S} - \int_{\mathrm{sides}} \rho\,(\boldsymbol{v}^{\dagger} - \boldsymbol{v}^{(b)})\cdot\hat{\boldsymbol{n}}\,\mathrm{d}\mathcal{S} \tag{47.125a}$$

$$\frac{\mathrm{d}}{\mathrm{d}t}\left[\int_{\mathcal{R}_{\mathrm{sub}}} \rho\, C \,\mathrm{d}V\right] = -\tilde{C}\int_{C=\tilde{C}} \rho\,(\boldsymbol{v}^{\dagger} - \boldsymbol{v}^{(b)})\cdot\hat{\boldsymbol{n}}\,\mathrm{d}\mathcal{S} - \int_{\mathrm{sides}} C\,\rho\,(\boldsymbol{v}^{\dagger} - \boldsymbol{v}^{(b)})\cdot\hat{\boldsymbol{n}}\,\mathrm{d}\mathcal{S}. \tag{47.125b}$$

The residual mean impermeability condition (47.120) for the $C = \tilde{C}$ surface renders a simplification to the flud mass and tracer mass budgets

$$\frac{\mathrm{d}}{\mathrm{d}t}\left[\int_{\mathcal{R}_{\mathrm{sub}}} \rho \,\mathrm{d}V\right] = -\int_{\mathrm{sides}} \rho\,(\boldsymbol{v}^{\dagger} - \boldsymbol{v}^{(b)})\cdot\hat{\boldsymbol{n}}\,\mathrm{d}\mathcal{S} \tag{47.126a}$$

$$\frac{\mathrm{d}}{\mathrm{d}t}\left[\int_{\mathcal{R}_{\mathrm{sub}}} \rho\, C \,\mathrm{d}V\right] = -\int_{\mathrm{sides}} C\,\rho\,(\boldsymbol{v}^{\dagger} - \boldsymbol{v}^{(b)})\cdot\hat{\boldsymbol{n}}\,\mathrm{d}\mathcal{S}. \tag{47.126b}$$

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 skewsion

The mass budget formulated in terms of Eulerian mean advection, and the corresponding tracer mass budget using eddy skewsion, are given by

$$\frac{\mathrm{d}}{\mathrm{d}t}\left[\int_{\mathcal{R}_{\mathrm{sub}}} \rho \,\mathrm{d}V\right] = -\int_{C=\tilde{C}} \rho\,(\boldsymbol{v} - \boldsymbol{v}^{(b)})\cdot\hat{\boldsymbol{n}}\,\mathrm{d}\mathcal{S} - \int_{\mathrm{sides}} \rho\,(\boldsymbol{v} - \boldsymbol{v}^{(b)})\cdot\hat{\boldsymbol{n}}\,\mathrm{d}\mathcal{S} \tag{47.127a}$$

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}\left[\int_{\mathcal{R}_{\mathrm{sub}}} \rho\, C \,\mathrm{d}V\right] &= -\tilde{C}\int_{C=\tilde{C}} \rho\,(\boldsymbol{v} - \boldsymbol{v}^{(b)})\cdot\hat{\boldsymbol{n}}\,\mathrm{d}\mathcal{S} - \int_{\mathrm{sides}} C\,\rho\,(\boldsymbol{v} - \boldsymbol{v}^{(b)})\cdot\hat{\boldsymbol{n}}\,\mathrm{d}\mathcal{S} \\ &\quad - \int_{\mathrm{sides}} [-\nabla C \wedge (\rho\,\boldsymbol{\Psi}^{*})]\cdot\hat{\boldsymbol{n}}\,\mathrm{d}\mathcal{S}.\end{aligned} \tag{47.127b}$$

We already saw in Section 47.7.4 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 equivalance 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 (47.126b) 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, $\boldsymbol{v}^{\dagger}$, as per the residual mean impermeability conditions 47.122 and 47.124. 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

(47.122) 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 (47.128):

$$(\boldsymbol{v}^* - \boldsymbol{v}^{(b)}) \cdot \hat{\boldsymbol{n}} = 0 \qquad \text{on } C \text{ isosurfaces and with } \boldsymbol{v} = 0. \tag{47.128}$$

When formulated using the residual mean advection, the fluid mass budget (47.126a) and tracer mass budget (47.126b) reduce in this case to

$$\frac{\mathrm{d}}{\mathrm{d}t}\left[\int_{\mathcal{R}_{\text{sub}}} \rho \,\mathrm{d}V\right] = -\int_{\text{sides}} \rho\, \boldsymbol{v}^* \cdot \hat{\boldsymbol{n}} \,\mathrm{d}\mathcal{S} \tag{47.129a}$$

$$\frac{\mathrm{d}}{\mathrm{d}t}\left[\int_{\mathcal{R}_{\text{sub}}} \rho\, C \,\mathrm{d}V\right] = -\int_{\text{sides}} C\, \rho\, \boldsymbol{v}^* \cdot \hat{\boldsymbol{n}} \,\mathrm{d}\mathcal{S}, \tag{47.129b}$$

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, (47.127a), and tracer mass budget written in terms of skew fluxes, (47.127b), are given by

$$\frac{\mathrm{d}}{\mathrm{d}t}\left[\int_{\mathcal{R}_{\text{sub}}} \rho \,\mathrm{d}V\right] = +\int_{C=\tilde{C}} \rho\, \boldsymbol{v}^{(b)} \cdot \hat{\boldsymbol{n}} \,\mathrm{d}\mathcal{S} \tag{47.130a}$$

$$\frac{\mathrm{d}}{\mathrm{d}t}\left[\int_{\mathcal{R}_{\text{sub}}} \rho\, C \,\mathrm{d}V\right] = +\tilde{C}\int_{C=\tilde{C}} \rho\, \boldsymbol{v}^{(b)} \cdot \hat{\boldsymbol{n}} \,\mathrm{d}\mathcal{S} - \int_{\text{sides}} [-\nabla C \wedge (\rho\, \boldsymbol{\Psi}^*)] \cdot \hat{\boldsymbol{n}} \,\mathrm{d}\mathcal{S}. \tag{47.130b}$$

For the mass budget, (47.130a), 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 $\boldsymbol{v}^{(b)} \cdot \hat{\boldsymbol{n}} = \boldsymbol{v}^* \cdot \hat{\boldsymbol{n}}$. For the tracer mass budget, (47.130b), 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 (47.129a) and (47.130a), and tracer mass budgets (47.129b) and (47.130b), 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.

47.8 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

$$\rho\, \frac{\mathrm{D}^\dagger \theta}{\mathrm{D}t} = \rho \left[\frac{\partial}{\partial t} + \boldsymbol{v}^\dagger \cdot \nabla\right] \theta = -\nabla \cdot \boldsymbol{J}^{\text{diff}}(\theta). \tag{47.131}$$

Further nonlinearities arise when the subgrid scale diffusion tensor is itself a function of the buoyancy, as discussed at the end of Section 47.3.4, and/or when the parameterized eddy-induced velocity is a function of the buoyancy, as discussed in Section 49.2.

47.8.1 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

$$\frac{\mathrm{D}^\dagger\theta}{\mathrm{D}t} = \frac{\partial\theta}{\partial t} + \boldsymbol{v}^\dagger \cdot \nabla\theta = 0 \qquad \text{adiabatic.} \tag{47.132}$$

Furthermore, following the kinematics of Section 14.4.2, the adiabatic residual mean flow field does not penetrate surfaces of constant potential temperature (θ-isosurfaces are impermeable) since

$$\boldsymbol{v}^\dagger \cdot \nabla\theta = -\frac{\partial\theta}{\partial t}. \tag{47.133}$$

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 47.7.4.

47.8.2 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 17.3 to render the expression (17.25) for w^{dia}, which we here write as

$$w^{\mathrm{dia}} \equiv \hat{\boldsymbol{n}} \cdot (\boldsymbol{v}^\dagger - \boldsymbol{v}^{(\theta)}) = \frac{1}{|\nabla\theta|}\,\frac{\mathrm{D}^\dagger\theta}{\mathrm{D}t} \tag{47.134}$$

where

$$\hat{\boldsymbol{n}} = \frac{\nabla\theta}{|\nabla\theta|} \qquad \text{and} \qquad \boldsymbol{v}^{(\theta)} \cdot \nabla\theta = -\frac{\partial\theta}{\partial t}. \tag{47.135}$$

Rearrangement of equation (47.134) renders the kinematic identity

$$\frac{\mathrm{D}^\dagger\theta}{\mathrm{D}t} = \frac{\partial\theta}{\partial t} + \boldsymbol{v}^\dagger \cdot \nabla\theta = w^{\mathrm{dia}}\,|\nabla\theta|. \tag{47.136}$$

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 47.2)

$$\rho\,\frac{\mathrm{D}^\dagger\theta}{\mathrm{D}t} = \nabla \cdot (\kappa\,\rho\,\nabla\theta), \tag{47.137}$$

with $\kappa > 0$ the scalar kinematic diffusivity. Diffusion in turn drives a diabatic transport velocity

$$\rho\,w^{\mathrm{dia}} = \frac{\nabla \cdot (\kappa\,\rho\,\nabla\theta)}{|\nabla\theta|}. \tag{47.138}$$

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

$$\rho\, w^{\text{dia}} = \frac{\partial_z(\kappa\,\rho\,\partial_z\theta)}{|\partial_z\theta|} \neq 0. \tag{47.139}$$

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 $\boldsymbol{v}^\dagger = 0$ we have

$$\frac{\partial\theta}{\partial t} = \rho^{-1}\,\frac{\partial}{\partial z}\left[\kappa\,\rho\,\frac{\partial\theta}{\partial z}\right] = w^{\text{dia}}\left|\frac{\partial\theta}{\partial z}\right|. \tag{47.140}$$

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

$$\rho\,\boldsymbol{v}^\dagger\cdot\nabla\theta = \rho\,w^{\text{dia}}\,|\nabla\theta| = \nabla\cdot(\kappa\,\rho\,\nabla\theta). \qquad \text{steady state.} \tag{47.141}$$

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.

47.9 Exercises

EXERCISE 47.1: 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

$$\tau_{\text{diffusion}} = \frac{\Delta^2}{\kappa} \tag{47.142}$$

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 47.2: ANALYTICAL SOLUTION TO ONE-DIMENSIONAL DIFFUSION EQUATION
Consider a one-dimensional diffusion equation

$$\frac{\partial C}{\partial t} = \kappa\frac{\partial^2 C}{\partial z^2}, \tag{47.143}$$

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

$$\frac{\partial C}{\partial t} = \kappa \frac{\partial^2 C}{\partial z^2}, \tag{47.144a}$$

$$\text{no-flux boundary condition from part (b)} \tag{47.144b}$$

$$C(z, t = 0) = C_0 \cos(Kz), \tag{47.144c}$$

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 47.3: 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

$$\frac{\partial C}{\partial t} = \frac{\partial}{\partial z}\left[\kappa \frac{\partial C}{\partial z}\right] \tag{47.145a}$$

$$= \frac{\partial \kappa}{\partial z}\frac{\partial C}{\partial z} + \kappa \frac{\partial^2 C}{\partial z^2}, \tag{47.145b}$$

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

$$\frac{\partial C^*}{\partial t} = \kappa \frac{\partial^2 C^*}{\partial z^2}. \tag{47.146}$$

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

$$\frac{\partial C}{\partial t} = \frac{\partial}{\partial x^m}\left[\mathbb{K}^{mn} \frac{\partial C}{\partial x^n}\right]. \tag{47.147}$$

Now consider an extrema in the tracer field, which is defined by

$$\frac{\partial C^*}{\partial x^n} = 0 \qquad \forall\, n = 1, 2, 3. \tag{47.148}$$

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 47.4: ONE-DIMENSIONAL ADVECTION

Consider the advection equation in one space dimension without boundaries

$$\frac{\partial C}{\partial t} + u\,\frac{\partial C}{\partial x} = 0 \tag{47.149a}$$

$$C(x, z, t = 0) = C_0\,\cos(k\,x) \tag{47.149b}$$

$$u(z, t) = \alpha\,z\,\cos(\omega\,t). \tag{47.149c}$$

The specified zonal velocity is non-divergent, oscillatory in time, and vertically sheared

$$\frac{\partial u}{\partial z} = \alpha\,\cos(\omega\,t), \tag{47.150}$$

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 (47.44).

EXERCISE 47.5: SKEW FLUX FOR OCEAN MESOSCALE EDDIES

Consider a middle-latitude mesoscale ocean eddy respecting geostrophic balance (see Section 26.3) on an f-plane (constant Coriolis parameter) and incompressibility. In this case, the horizontal eddy-induced velocity at the ocean surface is non-divergent

$$\boldsymbol{u}^* = \nabla \wedge \hat{\boldsymbol{z}}\,\psi. \tag{47.151}$$

In this equation, the geostrophic streamfunction is given by

$$\psi = -\hat{\boldsymbol{z}}\,\frac{g\,\eta}{f}, \tag{47.152}$$

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 (47.66).

(b) Determine the skew tracer flux (47.79a).

EXERCISE 47.6: 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 47.5: *Illustrating the area contained between two closed tracer contours, $C_0 \leq C(x, y, t) \leq C_1$. Exercise 47.6 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 47.5. Derive the following expression for the area enclosed by the two contours

$$\mathcal{A} = \int_{C_0}^{C_1} \mathrm{d}C \oint \frac{\mathrm{d}l}{|\nabla C|}. \tag{47.153}$$

In this expression, $\mathrm{d}l$ 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

$$\mathcal{A}(C) = \int_{C_0}^{C} \mathrm{d}C' \oint \frac{\mathrm{d}l}{|\nabla C'|} \tag{47.154}$$

we have the identity

$$\frac{\partial \mathcal{A}(C)}{\partial C} = \oint \frac{\mathrm{d}l}{|\nabla C|}. \tag{47.155}$$

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

$$\frac{\partial}{\partial C}\left[\int \Phi(\boldsymbol{x})\,\mathrm{d}\mathcal{A}\right] = \frac{\partial}{\partial C}\left[\int_{C_0}^{C} \mathrm{d}C' \oint \frac{\Phi\,\mathrm{d}l}{|\nabla C'|}\right] \tag{47.156a}$$

$$= \oint \frac{\Phi\,\mathrm{d}l}{|\nabla C|}, \tag{47.156b}$$

with Φ an arbitrary function. This is a remarkable identity with many useful applications such as those in *Marshall et al.* (2006).

EXERCISE 47.7: 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

$$\langle \boldsymbol{x} \rangle^C = \frac{\int \boldsymbol{x}\, C \,\mathrm{d}V}{\int C \,\mathrm{d}V}, \tag{47.157}$$

with C the tracer concentration, $\boldsymbol{x}$ the coordinate of a point in the fluid, and integration 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

$$\mathcal{V} = \int \mathrm{d}V \qquad \text{and} \qquad \mathcal{C} = \int C \,\mathrm{d}V. \tag{47.158}$$

Furthermore, when computing the time derivative acting on the integral, make use of the kinematic results from Section 15.3.3, in which for any integrand φ

$$\frac{\mathrm{d}}{\mathrm{d}t} \int \varphi \,\mathrm{d}V = \int \frac{\partial \varphi}{\partial t} \,\mathrm{d}V, \tag{47.159}$$

which follows since the region boundaries are assumed to be static.

(a) Consider a tracer concentration whose tendency at a point in space is affected only by advection

$$\frac{\partial C}{\partial t} = -\nabla \cdot (\boldsymbol{v}\, C), \tag{47.160}$$

with $\boldsymbol{v}$ a non-divergent velocity, $\nabla \cdot \boldsymbol{v} = 0$. Show that the tracer center of mass position evolves according to the tracer center of mass velocity

$$\frac{\mathrm{d}\langle \boldsymbol{x} \rangle^C}{\mathrm{d}t} = \langle \boldsymbol{v} \rangle^C, \tag{47.161}$$

where the tracer center of mass velocity is given by

$$\langle \boldsymbol{v} \rangle^C = \frac{\int \boldsymbol{v}\, C \,\mathrm{d}V}{\int C \,\mathrm{d}V} = \frac{1}{\mathcal{C}} \int \boldsymbol{v}\, C \,\mathrm{d}V. \tag{47.162}$$

(b) Consider a tracer concentration whose tendency at a point in space affected only by diffusion

$$\frac{\partial C}{\partial t} = \nabla \cdot (K \nabla C), \tag{47.163}$$

where $K = K(\boldsymbol{x}) > 0$ is a diffusivity (dimensions of squared length per time) that is assumed to vanish at the domain boundaries. Show that the tracer center of mass drifts up the diffusivity gradient

$$\frac{\mathrm{d}\langle \boldsymbol{x} \rangle^C}{\mathrm{d}t} = \langle \nabla K \rangle^C. \tag{47.164}$$

(c) Consider an initial tracer concentration that is a function only of latitude,

$$C(x, y, z, t = 0) = C_0(y), \tag{47.165}$$

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. Introduce an eddy stirring that breaks the zonal symmetry. Qualitatively discuss the process whereby this turbulent diffusion causes the tracer center of mass to drift towards the turbulent region.

EXERCISE 47.8: 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, $\mathcal{R}(\boldsymbol{v})$. The fluid is assumed to have a tracer whose concentration is affected by an irreversible process so that

$$\frac{\mathrm{D}C}{\mathrm{D}t} = \dot{C} \neq 0. \tag{47.166}$$

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

$$\langle \boldsymbol{x} \rangle^C = \frac{\int_{\mathcal{R}(\boldsymbol{v})} \boldsymbol{x}\, C\, \rho\, \mathrm{d}V}{\int_{\mathcal{R}(\boldsymbol{v})} C\, \rho\, \mathrm{d}V}. \tag{47.167}$$

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 15.3.5 since the region has a constant mass. Consequently, we can set

$$\frac{\mathrm{d}}{\mathrm{d}t} \int_{\mathcal{R}(\boldsymbol{v})} \psi\, \rho\, \mathrm{d}V = \int_{\mathcal{R}(\boldsymbol{v})} \frac{\mathrm{D}\psi}{\mathrm{D}t}\, \rho\, \mathrm{d}V. \tag{47.168}$$

Figure 47.6: *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 47.9: DIFFUSIVE HOMOGENIZATION OF SCALARS INSIDE CLOSED CONTOURS
The advection-diffusion equation for a tracer concentration is given by

$$\frac{\partial(\rho\, C)}{\partial t} + \nabla \cdot (\rho\, \boldsymbol{v}\, C) = -\nabla \cdot \boldsymbol{J} \tag{47.169}$$

where

$$\boldsymbol{J} = -\rho\, \mathbb{K} \cdot \nabla C \tag{47.170}$$

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

$$\nabla \cdot (\rho\, \boldsymbol{v}\, C) = -\nabla \cdot \boldsymbol{J}. \tag{47.171}$$

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 47.6. 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 51.5.

48

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 49).

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*, 2014). 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 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 15.1 and the maths and physics of the advection-diffusion equation explored in Chapter 47. We focus most discussion on incompressible flows discussed in Chapter 16 and applicable to the Boussinesq fluid commonly assumed for the ocean (Chapter 25). 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 48.4 and 48.6) are posed using the isopycnal vertical coordinates detailed in Chapter 8 and further pursued in Chapter 38.

48.1 Reynold's decomposition **848**
48.2 Basic kinematics of the GLM **849**
48.2.1 Motivation 850
48.2.2 Length scales and the small parameter 851
48.2.3 Decomposing the particle trajectory 851
48.2.4 GLM and the Stokes mean 852
48.2.5 An example linear wave 853
48.2.6 GLM with a materially constant scalar 855
48.2.7 Further study 856
48.3 Kinematics of eddy tracer fluxes **856**
48.3.1 Particle displacements and eddy tracer fluxes 856
48.3.2 Symmetric and skew-symmetric tracer fluxes 857
48.3.3 Massaging the mean field tracer equation 860
48.3.4 Connection to Stokes drift 860
48.3.5 A linear rotating periodic wave example 861
48.3.6 Further study 862
48.4 Kinematics of volume transport in isopycnal layers **862**
48.4.1 Isopycnal mean 863
48.4.2 Modified mean 863
48.4.3 Transformed residual mean (TRM) 865
48.4.4 Volume conservation and the thickness equation 866
48.4.5 Ensemble mean kinematics in isopycnal coordinates 868
48.4.6 Ensemble mean kinematics in geopotential coordinates 869
48.4.7 Approximate ensemble mean kinematics in geopotential coordinates 869
48.4.8 Further study 871
48.5 Thickness transport for a linear longitudinal wave **871**
48.5.1 An undulating fluid layer 871
48.5.2 Stokes drift 872
48.5.3 Linearized thickness perturbations 873
48.5.4 Correlation between thickness and velocity 873
48.5.5 Comments 874
48.6 Mean tracer equation **874**
48.6.1 Thickness weighted means 874
48.6.2 Isopycnal mean thickness weighted tracer equation 874
48.6.3 Subgrid scale tracer transport tensor 875
48.6.4 Mean tracer transport beneath a density surface 875
48.6.5 Summary of the tracer parameterization problem 876
48.6.6 Comments 876

48.1 Reynold's decomposition

At any point in space and time, we can decompose a field into a mean, $\overline{\Phi}(\boldsymbol{x}, t)$, and a departure from the mean, $\Phi'(\boldsymbol{x}, t)$

$$\Phi(\boldsymbol{x}, t) = \overline{\Phi}(\boldsymbol{x}, t) + \Phi'(\boldsymbol{x}, t). \tag{48.1}$$

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: $\overline{\Phi}(\boldsymbol{x}, t) = \overline{\Phi}(\boldsymbol{x})$. 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 "Reynold's decomposition"

$$\overline{\Phi'} = 0 \tag{48.2a}$$

$$\overline{\overline{\Phi}} = \overline{\Phi} \tag{48.2b}$$

$$\overline{\gamma\,\overline{\Phi}} = \gamma\,\overline{\Phi} \qquad \text{for } \gamma \text{ a constant.} \tag{48.2c}$$

Equation (48.2a) says that the mean of an eddy fluctuation vanishes. The equality (48.2b) says that the mean of a mean field returns the mean field. The final equality, (48.2c), 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.

48.2 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*

$$\text{Lagrangian mean} = \text{Eulerian mean} + \text{Stokes mean}. \tag{48.3}$$

This name is motivated from the *Stokes drift* introduced in Section 45.3, which we again encounter in Section 48.2.4. 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.

48.2.1 Motivation

Consider a materially constant scalar field

$$\frac{\mathrm{D}\Phi}{\mathrm{D}t} = \frac{\partial \Phi}{\partial t} + \boldsymbol{v} \cdot \nabla \Phi = 0. \tag{48.4}$$

The scalar Φ is constant following fluid particles whose trajectories are integral curves of the fluid velocity $\boldsymbol{v}$. 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 $\mathrm{D}\Phi/\mathrm{D}t = 0$. This aspiration is not trivial.

Eulerian mean

An Eulerian mean operator considered in Section 48.1 leads to the mean field equation

$$\frac{\partial \overline{\Phi}}{\partial t} + \overline{\boldsymbol{v}} \cdot \nabla \overline{\Phi} = -\overline{\boldsymbol{v}' \cdot \nabla \Phi'}. \tag{48.5}$$

Whereas Φ is materially constant when following the instantaneous flow field $\boldsymbol{v}$, the Eulerian mean $\overline{\Phi}$ is not materially constant when following $\overline{\boldsymbol{v}}$ due to the source term $-\overline{\boldsymbol{v}' \cdot \nabla \Phi'}$ 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

$$\frac{\partial \Phi(\boldsymbol{a}, t)}{\partial t} = 0. \tag{48.6}$$

Consider a mean operator computed as an average over a region of material space coordinate $\boldsymbol{a}$. For example, if $\boldsymbol{a}$ is the initial fluid particle position, then an average coordinate, $\overline{\boldsymbol{a}}$, and corresponding averaged field, $\overline{\Phi}$, render a coarse-graining over the initial positions. Since each member of the Lagrangian average satisfies the linear equation (48.6), so too does the Lagrangian mean

$$\frac{\partial \overline{\Phi}(\overline{\boldsymbol{a}}, t)}{\partial t} = 0. \tag{48.7}$$

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

$$\frac{\partial \overline{\Phi}^{(L)}}{\partial t} + \overline{\boldsymbol{v}}^{(L)} \cdot \nabla \overline{\Phi}^{(L)} = 0. \tag{48.8}$$

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 (48.8), we will motivate the GLM average from the analysis of small amplitude eddying motions.

48.2.2 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, $|\boldsymbol{\xi}|$. In the following, we assume the particle displacements are small relative to λ

$$|\boldsymbol{\xi}| \ll \lambda \qquad \text{small amplitude waves.} \tag{48.9}$$

We thus introduce the small non-dimensional ratio of length scales for the following analysis

$$\alpha = \frac{|\boldsymbol{\xi}|}{\lambda} \ll 1. \tag{48.10}$$

48.2.3 Decomposing the particle trajectory

Recall the discussion of fluid particle trajectories given in Chapter 12. In this description, the trajectory of a particle is determined by integrating the relation between the particle trajectory and the particle velocity

$$\left[\frac{\partial \boldsymbol{X}(\boldsymbol{a},t)}{\partial t}\right]_{\boldsymbol{a}} = \boldsymbol{v}[\boldsymbol{X}(\boldsymbol{a},t)] \Longrightarrow \boldsymbol{X}(\boldsymbol{a},t) = \boldsymbol{X}(\boldsymbol{a},0) + \int_0^t \boldsymbol{v}[\boldsymbol{X}(\boldsymbol{a},t')]\,\mathrm{d}t', \tag{48.11}$$

so that the trajectory measures the position of a particle relative to a chosen origin. The material coordinate, $\boldsymbol{a}$, 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, $\boldsymbol{x}$, to be the mean position of a unique fluid particle. In turn, we introduce an Eulerian field, $\boldsymbol{\xi}(\boldsymbol{x},t)$, that measures the position of a fluid particle relative to its mean position. Correspondingly, the Eulerian mean of the disturbance field vanishes

$$\overline{\boldsymbol{\xi}(\boldsymbol{x},t)} = 0. \tag{48.12}$$

Note that the Eulerian mean operator can be any of the operators (or others) satisfying the Reynold's decomposition property discussed in Section 48.1

Specification of $\boldsymbol{\xi}(\boldsymbol{x},t)$ 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 48.1, for which the particle displacement amplitude is much smaller than the wavelength of the disturbance

$$\alpha = \frac{|\boldsymbol{\xi}|}{\lambda} \ll 1. \tag{48.13}$$

In this case the disturbance field is constructed by time integration of the eddy velocity field

$$\left[\frac{\partial \boldsymbol{\xi}(\boldsymbol{x},t)}{\partial t}\right]_{\boldsymbol{x}} = \boldsymbol{v}'(\boldsymbol{x},t) \Longrightarrow \boldsymbol{\xi}(\boldsymbol{x},t) = \int^t \boldsymbol{v}'(\boldsymbol{x},t')\,\mathrm{d}t'. \tag{48.14}$$

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

$$\nabla \cdot \boldsymbol{v}' = 0 \Longrightarrow \nabla \cdot \boldsymbol{\xi} = 0. \tag{48.15}$$

The definition (48.14) for the disturbance field, $\boldsymbol{\xi}(\boldsymbol{x}, t)$, is directly analogous to the particle trajectory position, $\boldsymbol{X}(\boldsymbol{a}, t)$, given by equation (48.11). However, there are important distinctions. Namely, the disturbance, $\boldsymbol{\xi}(\boldsymbol{x}, t)$, is an Eulerian field that measures the position of a fluid particle relative to its mean position, with each Eulerian position $\boldsymbol{x}$ corresponding to the mean position for a distinct fluid particle. In contrast, the particle position, $\boldsymbol{X}(\boldsymbol{a}, t)$, is a Lagrangian field that is attached to each fluid particle and measures the position of that particle relative to a unique origin.

Figure 48.1: *Illustrating the displacement of fluid particles at two selective vertical positions due to a linear transverse wave disturbance. The particle positions, $\boldsymbol{x}^{(\xi)} = \boldsymbol{x} + \boldsymbol{\xi}(\boldsymbol{x}, t)$, have a disturbance field of the form $\boldsymbol{\xi}(\boldsymbol{x}, t) = \hat{\boldsymbol{z}}\, \xi_0(x, z) \sin(k\, x - \omega\, t)$, with $\xi_0(x, z)$ a spatially dependent wave amplitude, $\lambda = 2\pi/k$ the wavelength, $\boldsymbol{k} = \hat{\boldsymbol{x}}\, k$ the wavevector, $\omega = c\, k$ 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: $\boldsymbol{\xi} \cdot \boldsymbol{k} = 0$ (see Section 48.2.5), whereas Stokes drift requires particle motion to have a nonzero component in the wave direction (see Figures 45.4 and 45.5). Even so, it does generally produced a Stokes mean for an arbitrary field Φ (Section 48.2.4).*

48.2.4 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

$$\underbrace{\overline{\Phi(\boldsymbol{x} + \boldsymbol{\xi}(\boldsymbol{x}, t))}}_{\text{GLM}} \neq \underbrace{\overline{\Phi(\boldsymbol{x}, t)}}_{\text{Eulerian}}, \tag{48.16}$$

where it is common to make use of the shorthand[1]

$$\boldsymbol{x}^{(\xi)}(\boldsymbol{x}, t) \equiv \boldsymbol{x} + \boldsymbol{\xi}(\boldsymbol{x}, t) \tag{48.17}$$

for the instantaneous position of the fluid particle. The average

$$\overline{\Phi}^{(L)}(\boldsymbol{x}, t) \equiv \overline{\Phi(\boldsymbol{x} + \boldsymbol{\xi}(\boldsymbol{x}, t), t)} = \overline{\Phi(\boldsymbol{x}^{(\xi)})} \tag{48.18}$$

[1] We place superscripts ξ, S, L, E inside parentheses to distinguish from tensor labels.

defines the generalized Lagrangian mean. As defined, the GLM is computed by evaluating the property Φ at the position of a fluid particle, $\boldsymbol{x}^{(\xi)}(\boldsymbol{x},t) = \boldsymbol{x} + \boldsymbol{\xi}(\boldsymbol{x},t)$, where $\boldsymbol{x}$ is both an arbitrary Eulerian field point and the mean position of a fluid particle, $\overline{\boldsymbol{x}^{(\xi)}} = \boldsymbol{x}$. In contrast, the Eulerian mean is determined by evaluating Φ at the fixed Eulerian point in space

$$\overline{\Phi}^{(E)}(\boldsymbol{x},t) \equiv \overline{\Phi(\boldsymbol{x},t)}. \tag{48.19}$$

Following our discussion at the start of Section 48.2, we define the difference between the GLM and Euleriam mean as the Stokes mean

$$\overline{\Phi}^{(S)}(\boldsymbol{x},t) = \overline{\Phi}^{(L)}(\boldsymbol{x},t) - \overline{\Phi}^{(E)}(\boldsymbol{x},t). \tag{48.20}$$

The Stokes mean arises from inhomogenities in Φ, which in turn lead to differences in its mean depending on how that field is sampled, whether sampled on a fluid particle, $\boldsymbol{x}^{(\xi)}(\boldsymbol{x},t)$, or sampled at the mean position of the fluid particle, $\boldsymbol{x}$.

We mathematically expose the origin of the Stokes mean by performing a Taylor series expansion around the mean particle position

$$\Phi(\boldsymbol{x}+\boldsymbol{\xi},t) = \Phi(\boldsymbol{x},t) + \boldsymbol{\xi}\cdot\nabla\Phi(\boldsymbol{x},t) + \frac{1}{2}\,\xi_m\,\xi_n\partial_m\partial_n\Phi(\boldsymbol{x},t) + \mathcal{O}(\alpha^3). \tag{48.21}$$

The non-dimensional ratio $\alpha = |\boldsymbol{\xi}|/\lambda \ll 1$ was introduced in equation (48.13), 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 (48.21) then leads to an expression for the Stokes mean

$$\overline{\Phi}^{(S)}(\boldsymbol{x},t) = \overline{\Phi}^{(L)}(\boldsymbol{x},t) - \overline{\Phi}^{(E)}(\boldsymbol{x},t) \tag{48.22a}$$

$$= \overline{\boldsymbol{\xi}\cdot\nabla\Phi} + \frac{1}{2}\,\overline{\xi_m\,\xi_n\partial_m\partial_n\Phi} + \mathcal{O}(\alpha^3). \tag{48.22b}$$

$$= \overline{\boldsymbol{\xi}\cdot\nabla\Phi'} + \frac{1}{2}\,\overline{\xi_m\,\xi_n}\,\partial_m\partial_n\overline{\Phi}^{(E)} + \mathcal{O}(\alpha^3), \tag{48.22c}$$

where we introduced the Eulerian fluctuation

$$\Phi'(\boldsymbol{x},t) = \Phi(\boldsymbol{x},t) - \overline{\Phi}^{(E)}(\boldsymbol{x},t) \tag{48.23}$$

and all terms on the right hand side of equation (48.22c) are evaluated at $(\boldsymbol{x},t)$. The Stokes drift (Section 45.3) associated with the GLM arises from setting Φ equal to one of the velocity components

$$\overline{v}_p^{(S)} = \overline{\boldsymbol{\xi}\cdot\nabla v_p'} + \frac{1}{2}\,\overline{\xi_m\,\xi_n}\,\partial_m\partial_n\overline{v}_p^{(E)} + \mathcal{O}(\alpha^3). \tag{48.24}$$

48.2.5 An example linear wave

We exemplify the previous discussion by considering the small amplitude linear wave

$$\boldsymbol{\xi} = -\frac{\boldsymbol{U}(\boldsymbol{x})}{\omega}\sin(\boldsymbol{k}\cdot\boldsymbol{x}-\omega\,t) \tag{48.25a}$$

$$\boldsymbol{v}' = \partial_t\boldsymbol{\xi} = \boldsymbol{U}(\boldsymbol{x})\cos(\boldsymbol{k}\cdot\boldsymbol{x}-\omega\,t) \tag{48.25b}$$

$$\nabla v_p' = \nabla U_p\cos(\boldsymbol{k}\cdot\boldsymbol{x}-\omega\,t) - \boldsymbol{k}\,U_p\sin(\boldsymbol{k}\cdot\boldsymbol{x}-\omega\,t) \tag{48.25c}$$

$$\nabla\cdot\boldsymbol{v}' = (\nabla\cdot\boldsymbol{U})\cos(\boldsymbol{k}\cdot\boldsymbol{x}-\omega\,t) - \boldsymbol{k}\cdot\boldsymbol{U}\sin(\boldsymbol{k}\cdot\boldsymbol{x}-\omega\,t) \tag{48.25d}$$

where $\boldsymbol{U}$ is the velocity amplitude that is generally a function of space, $\boldsymbol{k}$ is the wavevector, and

$$T = 2\,\pi/\omega \tag{48.26}$$

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 $\boldsymbol{x}$. The disturbance field and velocity field both have a zero mean when time integrated over a wave period

$$\overline{\boldsymbol{\xi}}(\boldsymbol{x}) = \frac{1}{T}\int_0^T \boldsymbol{\xi}(\boldsymbol{x}, t')\,\mathrm{d}t' = 0 \tag{48.27a}$$

$$\overline{\boldsymbol{v}'}(\boldsymbol{x}) = \frac{1}{T}\int_0^T \boldsymbol{v}'(\boldsymbol{x}, t')\,\mathrm{d}t' = 0. \tag{48.27b}$$

To maintain a non-divergent eddy velocity at arbitrary times requires

$$\nabla\cdot\boldsymbol{v}' = 0 \Longrightarrow \nabla\cdot\boldsymbol{U} = \boldsymbol{U}\cdot\boldsymbol{k} = 0. \tag{48.28}$$

The second condition means that the wave is transverse, so that particle displacements are orthogonal to the wavevector (e.g., Figure 48.1)

Stokes drift

Specializing to the velocity field (48.25b), substituting into the Stokes drift expression (48.24), and making use of an average over a wave period yields

$$\overline{\boldsymbol{\xi}\cdot\nabla v_p'} = \frac{U_p\,\boldsymbol{U}\cdot\boldsymbol{k}}{2\,\omega} \tag{48.29a}$$

$$\overline{v}_p^{(E)} = 0. \tag{48.29b}$$

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

$$\overline{\boldsymbol{v}}^{(S)} = \frac{\boldsymbol{U}\,(\boldsymbol{U}\cdot\boldsymbol{k})}{2\,\omega} + \mathcal{O}(\alpha^2). \tag{48.30}$$

The Stokes drift vanishes at this order of accuracy for transverse waves in which $\boldsymbol{U}\cdot\boldsymbol{k} = 0$.

As a check on our formalism we consider a one-dimensional longitudinal wave, in which the Stokes drift is given by

$$\overline{v}^{(S)} = \frac{U^2}{2\,c} + \mathcal{O}(\alpha^2). \tag{48.31}$$

This result agrees with that derived using Lagrangian trajectories in Section 45.3 (see Exercise 45.1). 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

$$\overline{\Phi}^{(S)}(\boldsymbol{x}, t) = -\omega^{-1}\,\boldsymbol{U}\cdot\overline{\nabla\Phi'\sin(\boldsymbol{k}\cdot\boldsymbol{x} - \omega t)} + \mathcal{O}(\alpha^2) \tag{48.32a}$$

$$= -\omega^{-1}\,\overline{\nabla\cdot(\boldsymbol{U}\,\Phi')\sin(\boldsymbol{k}\cdot\boldsymbol{x} - \omega t)} + \mathcal{O}(\alpha^2), \tag{48.32b}$$

where the second equality made use of the non-divergent nature of the wave field (48.28). 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, $\boldsymbol{U}$. For example, consider a transverse wave such as that shown in Figure 48.1. Even though the Stokes drift vanishes to order $\mathcal{O}(\alpha^2)$, the Stokes mean, $\overline{\Phi}^{(S)}(\boldsymbol{x},t)$, can be nonzero so long as there is a nonzero vertical gradient in the Eulerian fluctuation.

48.2.6 GLM with a materially constant scalar

Consider a materially constant scalar field, such as a tracer concentration in the absence of mixing and sources

$$\frac{\mathrm{D}C}{\mathrm{D}t} = 0. \tag{48.33}$$

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

$$\overline{C}^{(L)}(\boldsymbol{x},t) = C(\boldsymbol{x}+\boldsymbol{\xi},t). \tag{48.34}$$

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, $\boldsymbol{x}$, the GLM field $\overline{C}^{(L)}(\boldsymbol{x},t)$ equals to the concentration, C, evaluated on a fluid particle at $\boldsymbol{x}+\boldsymbol{\xi}$. 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 (48.34) when considering isopycnal kinematics in Sections 48.4 and 48.6.

Relating the particle disturbance field to Eulerian properties of C

There is a frequently used consequence of the identity (48.34) involving the disturbance field, the Eulerian fluctuation

$$C'(\boldsymbol{x},t) = C(\boldsymbol{x},t) - \overline{C}(\boldsymbol{x},t) \tag{48.35}$$

and the Eulerian mean

$$C^{(E)}(\boldsymbol{x},t) = \overline{C}(\boldsymbol{x},t). \tag{48.36}$$

To derive it, recall the Taylor series expansion (48.21) truncated here to first order accuracy

$$C(\boldsymbol{x}+\boldsymbol{\xi},t) = C(\boldsymbol{x},t) + \boldsymbol{\xi}\cdot\nabla\overline{C}(\boldsymbol{x},t) + \mathcal{O}(\alpha^2). \tag{48.37}$$

Taking the Eulerian mean of both sides renders

$$\overline{C(\boldsymbol{x}+\boldsymbol{\xi},t)} = \overline{C}(\boldsymbol{x},t) + \mathcal{O}(\alpha^2). \tag{48.38}$$

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 (48.22c). From equation (48.34) we know that $C(\boldsymbol{x}+\boldsymbol{\xi},t) = \overline{C(\boldsymbol{x}+\boldsymbol{\xi},t)}$, so that we can subtract equations (48.37) and (48.38) to find

$$C'(\boldsymbol{x},t) = -\boldsymbol{\xi}\cdot\nabla\overline{C}(\boldsymbol{x},t) + \mathcal{O}(\alpha^2). \tag{48.39}$$

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 48.3. Furthermore, for the isopycnal kinematics in Sections 48.4 and 48.6, we focus on vertical particle displacements, $\boldsymbol{\xi} = \xi\,\hat{\boldsymbol{z}}$, in which case the Eulerian fluctuation is given by

$$C'(z,t) = -\xi\,\partial_z \overline{C}(z,t) + \mathcal{O}(\alpha^2). \tag{48.40}$$

48.2.7 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* (2014).

48.3 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

$$\frac{\partial C}{\partial t} + \nabla\cdot(\boldsymbol{v}\,C) = 0, \tag{48.41}$$

and its Eulerian mean is

$$\frac{\partial \overline{C}}{\partial t} + \nabla\cdot(\overline{\boldsymbol{v}}\,\overline{C}) = -\nabla\cdot\overline{\boldsymbol{v}'\,C'}. \tag{48.42}$$

The eddy advective flux, $\boldsymbol{v}'\,C'$, is the product of the eddy velocity and eddy tracer concentration. Its mean provides the correlation or mean eddy flux, $\overline{\boldsymbol{v}'\,C'}$. 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 48.2 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.

48.3.1 Particle displacements and eddy tracer fluxes

Following Section 48.2, we here introduce a particle disturbance vector corresponding to small amplitude eddy fluctuations

$$\partial_t \boldsymbol{\xi}(\boldsymbol{x},t) = \boldsymbol{v}'(\boldsymbol{x},t) + \mathcal{O}(\alpha^2) \tag{48.43a}$$

$$\overline{\boldsymbol{\xi}} = 0. \tag{48.43b}$$

Correspondingly, each spatial point, $\boldsymbol{x}$, is the mean position of a fluid particle whose instantaneous position is $\boldsymbol{x} + \boldsymbol{\xi}(\boldsymbol{x},t)$. Following the results from Section 48.2.6, to leading order we can write the Eulerian fluctuation in terms of the particle displacement (equation (48.39))

$$C'(\boldsymbol{x},t) = -\boldsymbol{\xi}\cdot\nabla\overline{C}(\boldsymbol{x},t) + \mathcal{O}(\alpha^2) \tag{48.44}$$

Notice that if the particle displacement is oriented along a mean tracer iso-surface, then $\boldsymbol{\xi}\cdot\nabla\overline{C}(\boldsymbol{x},t) = 0$ 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

$$\boldsymbol{v}'\,C' = -\partial_t\boldsymbol{\xi}\,(\boldsymbol{\xi}\cdot\nabla)\overline{C} + \mathcal{O}(\alpha^2). \tag{48.45}$$

We unpack this expression for the purpose of characterizing kinematic properties of the eddy tracer flux

48.3.2 Symmetric and skew-symmetric tracer fluxes

From equation (48.45), the m'th component of the eddy tracer flux is given by

$$v'_m\, C' = -\left[(\partial_t\, \xi_m)\, \xi_n\right] \partial_n \overline{C}. \tag{48.46}$$

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 components[2]

$$2\,(\partial_t\, \xi_m)\, \xi_n = \left[(\partial_t\, \xi_m)\, \xi_n + (\partial_t\, \xi_n)\, \xi_m\right] + \left[(\partial_t\, \xi_m)\, \xi_n - (\partial_t\, \xi_n)\, \xi_m\right] \tag{48.47a}$$

$$= \partial_t(\xi_m\, \xi_n) + \left[(\partial_t\, \xi_m)\, \xi_n - (\partial_t\, \xi_n)\, \xi_m\right]. \tag{48.47b}$$

Introducing the symmetric and anti-symmetric correlation tensors

$$2\,\mathbb{K}_{mn} \equiv \overline{\partial_t\,(\xi_m\, \xi_n)} \tag{48.48a}$$

$$2\,\mathbb{A}_{mn} \equiv \overline{(\partial_t\, \xi_m)\, \xi_n} - \overline{(\partial_t\, \xi_n)\, \xi_m} \tag{48.48b}$$

allows us to write the mean eddy tracer flux

$$\overline{v'_m\, C'} = -(\mathbb{K}_{mn} + \mathbb{A}_{mn})\, \partial_n \overline{C} \tag{48.49}$$

and the mean field tracer equation (48.42)

$$\frac{\partial \overline{C}}{\partial t} + \nabla \cdot (\overline{\boldsymbol{v}}\, \overline{C}) = \nabla \cdot \left[(\mathbb{K} + \mathbb{A}) \cdot \nabla \overline{C}\right]. \tag{48.50}$$

The right hand side of this equation equals to the convergence of the symmetric and skew-symmetric tracer fluxes[3]

$$\nabla \cdot \left[(\mathbb{K} + \mathbb{A}) \cdot \nabla \overline{C}\right] = -\nabla \cdot (\boldsymbol{F}^{(\mathrm{sym})} + \boldsymbol{F}^{(\mathrm{skew})}), \tag{48.51}$$

where

$$\boldsymbol{F}^{(\mathrm{sym})} = -\mathbb{K} \cdot \nabla \overline{C} \tag{48.52a}$$

$$\boldsymbol{F}^{(\mathrm{skew})} = -\mathbb{A} \cdot \nabla \overline{C} \tag{48.52b}$$

$$\overline{\boldsymbol{v}'\, C'} = \boldsymbol{F}^{(\mathrm{sym})} + \boldsymbol{F}^{(\mathrm{skew})}. \tag{48.52c}$$

[2]See Section 13.2.4 for a similar decomposition of the velocity gradient tensor.

[3]We place parentheses around "skew" and "sym" to distinguish the name for these vectors from what may otherwise appear to be tensor labels.

The symmetric flux

In terms of particle displacements, the symmetric flux (48.52a) is given by

$$F_m^{(\mathrm{sym})} = -\mathbb{K}_{mn}\,\partial_n \overline{C} = -\frac{1}{2}\,\overline{\partial_t(\xi_m\,\xi_n)}\,\partial_n \overline{C}. \tag{48.53}$$

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 48.3.5). 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 49.3.

The skew, advective, and rotational fluxes

Following our discussion in Section 47.6, we write the skew flux as

$$F_m^{(\mathrm{skew})} = -\mathbb{A}_{mn}\,\partial_n \overline{C} = -\epsilon_{mnp}\,\Psi_p\,\partial_n \overline{C} = -\left(\nabla \overline{C} \wedge \boldsymbol{\Psi}\right)_m, \tag{48.54}$$

where we introduced the vector streamfunction (dimensions squared length per time)[4]

$$\boldsymbol{\Psi} = \frac{1}{2}\,\overline{\partial_t \boldsymbol{\xi} \wedge \boldsymbol{\xi}} = \frac{1}{2}\,\overline{\boldsymbol{v}' \wedge \boldsymbol{\xi}}. \tag{48.55}$$

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

$$\begin{aligned}
\boldsymbol{F}^{(\mathrm{skew})} &= -\nabla \overline{C} \wedge \boldsymbol{\Psi} && (48.56\mathrm{a})\\
&= (\nabla \wedge \boldsymbol{\Psi})\,\overline{C} - \nabla \wedge (\overline{C}\,\boldsymbol{\Psi}) && (48.56\mathrm{b})\\
&= \mathcal{U}^{(\Psi)}\,\overline{C} - \nabla \wedge (\overline{C}\,\boldsymbol{\Psi}) && (48.56\mathrm{c})\\
&= \boldsymbol{F}^{(\mathrm{adv})} - \boldsymbol{F}^{(\mathrm{rot})}, && (48.56\mathrm{d})
\end{aligned}$$

so that the skew flux equals to an advective flux minus a rotational flux. We here introduced the non-divergent velocity

$$\mathcal{U}^{(\Psi)} = \nabla \wedge \boldsymbol{\Psi} \tag{48.57}$$

and the non-divergent rotational flux

$$\boldsymbol{F}^{(\mathrm{rot})} = \nabla \wedge (\overline{C}\,\boldsymbol{\Psi}). \tag{48.58}$$

The divergence of the skew flux equals to the divergence of the advective flux

$$\nabla \cdot \boldsymbol{F}^{(\mathrm{skew})} = \nabla \cdot (\boldsymbol{F}^{(\mathrm{adv})} - \boldsymbol{F}^{(\mathrm{rot})}) = \nabla \cdot \boldsymbol{F}^{(\mathrm{adv})}, \tag{48.59}$$

so that the rotational flux has no impact on evolution of the mean tracer field.

[4] *Middleton and Loder* (1989) and *Garrett* (2006) introduce a skew-diffusivity, $\boldsymbol{D}$, which is opposite in sign to the vector streamfunction: $\boldsymbol{\Psi} = -\boldsymbol{D}$.

What does a point measurement estimate?

From equation (48.52c), we see that a point measurement of the correlation $\overline{\boldsymbol{v}' C'}$ provides an estimate of the diffusive and skew diffusive tracer fluxes

$$\overline{\boldsymbol{v}' C'} = \boldsymbol{F}^{(\mathrm{sym})} + \boldsymbol{F}^{(\mathrm{skew})} = -(\mathbb{K} + \mathbb{A}) \cdot \nabla \overline{C}. \tag{48.60}$$

Furthermore, for a periodic wave field, where the symmetric tensor vanishes, the correlation, $\overline{\boldsymbol{v}' C'}$, provides a direct estimate of the skew flux, $-\nabla \overline{C} \wedge \boldsymbol{\Psi}$. One might instead presume that the point measurement offers a direct estimate of the advective flux, $\overline{C}\, \mathcal{U}^{(\Psi)}$, 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

$$\overline{\boldsymbol{v}' C'} = -\mathbb{K} \cdot \nabla \overline{C} - \nabla \overline{C} \wedge \boldsymbol{\Psi} \tag{48.61a}$$

$$= -\mathbb{K} \cdot \nabla \overline{C} - \nabla \wedge (\overline{C}\, \boldsymbol{\Psi}) + \overline{C}\, \nabla \wedge \boldsymbol{\Psi} \tag{48.61b}$$

$$= -\mathbb{K} \cdot \nabla \overline{C} - \nabla \wedge (\overline{C}\, \boldsymbol{\Psi}) + \overline{C}\, \mathcal{U}^{(\Psi)}. \tag{48.61c}$$

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 $\mathcal{S}$

$$\mathcal{T} = \overline{\int_{\mathcal{S}} \boldsymbol{v}\, C \cdot \hat{\boldsymbol{n}}\, \mathrm{d}S} = \int_{\mathcal{S}} \overline{\boldsymbol{v}\, C} \cdot \hat{\boldsymbol{n}}\, \mathrm{d}S = \int_{\mathcal{S}} [\overline{\boldsymbol{v}}\, \overline{C} + \overline{\boldsymbol{v}' C'}] \cdot \hat{\boldsymbol{n}}\, \mathrm{d}S. \tag{48.62}$$

In terms of particle displacements, the eddy correlation, $\overline{\boldsymbol{v}' C'}$, equals to the sum of the symmetric flux and the skew flux as in equation (48.60). Introducing the diffusive, advective, and rotational flux then renders

$$\mathcal{T} = \int_{\mathcal{S}} [\overline{\boldsymbol{v}}\, \overline{C} + \mathcal{U}^{(\Psi)}\, \overline{C} - \nabla \wedge (\overline{C}\, \boldsymbol{\Psi}) - \mathbb{K} \cdot \nabla \overline{C}] \cdot \hat{\boldsymbol{n}}\, \mathrm{d}S. \tag{48.63}$$

Use of Stokes' Theorem transforms the rotational term to a line integral around the boundary of the area

$$\mathcal{T} = \int_{\mathcal{S}} [\overline{\boldsymbol{v}}\, \overline{C} + \mathcal{U}^{(\Psi)}\, \overline{C} - \mathbb{K} \cdot \nabla \overline{C}] \cdot \hat{\boldsymbol{n}}\, \mathrm{d}S - \oint_{\partial \mathcal{S}} \overline{C}\, \boldsymbol{\Psi} \cdot \mathrm{d}\boldsymbol{l}. \tag{48.64}$$

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

$$\oint_{\partial \mathcal{S}} \overline{C}\, \boldsymbol{\Psi} \cdot \mathrm{d}\boldsymbol{l} = (1/2) \oint_{\partial \mathcal{S}} \overline{C}\, (\overline{\boldsymbol{v}' \wedge \boldsymbol{\xi}}) \cdot \mathrm{d}\boldsymbol{l}. \tag{48.65}$$

We further this interpretation when considering the transport beneath a fluctuating isopycnal surface in Section 48.6.4.

48.3.3 Massaging the mean field tracer equation

We here write the mean tracer equation (48.50) in various forms that can be found throughout the literature. For this purpose, write the right hand side of equation (48.50) in the form

$$\nabla \cdot \left[(\mathbb{K} + \mathbb{A}) \cdot \nabla \overline{C}\right] = \partial_m \left[(\mathbb{K}_{mn} + \mathbb{A}_{mn})\, \partial_n \overline{C}\right] \tag{48.66a}$$

$$= \partial_m (\mathbb{K}_{mn} + \mathbb{A}_{mn})\, \partial_n \overline{C} + (\mathbb{K}_{mn} + \mathbb{A}_{mn})\, \partial_m \partial_n \overline{C} \tag{48.66b}$$

$$= \partial_m (\mathbb{K}_{mn} + \mathbb{A}_{mn})\, \partial_n \overline{C} + \mathbb{K}_{mn}\, \partial_m \partial_n \overline{C}. \tag{48.66c}$$

The final equality follows from the identity

$$\mathbb{A}_{mn}\, \partial_m \partial_n \overline{C} = 0, \tag{48.67}$$

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 (48.66c) can be interpreted as an advection operator through the action of a non-divergent plus a divergent advection velocity

$$\partial_m (\mathbb{K}_{mn} + \mathbb{A}_{mn})\, \partial_n \overline{C} = [\mathcal{U}^{(K)} + \mathcal{U}^{(\Psi)}] \cdot \nabla \overline{C}, \tag{48.68}$$

where[5]

$$\mathcal{U}_n^{(K)} \equiv -\partial_m \mathbb{K}_{mn} \Longrightarrow \nabla \cdot \mathcal{U}^{(K)} = -\partial_n \partial_m \mathbb{K}_{mn} \neq 0 \tag{48.69a}$$

$$\mathcal{U}_n^{(\Psi)} \equiv -\partial_m \mathbb{A}_{mn} \Longrightarrow \nabla \cdot \mathcal{U}^{(\Psi)} = \partial_n \partial_m \mathbb{A}_{mn} = 0. \tag{48.69b}$$

Bringing the above results together allows us to write the mean field tracer equation (48.50) in the following equivalent forms

$$\frac{\partial \overline{C}}{\partial t} + [\overline{\boldsymbol{v}} + \mathcal{U}^{(\Psi)} + \mathcal{U}^{(K)}] \cdot \nabla \overline{C} = \mathbb{K}_{mn}\, \partial_m \partial_n \overline{C} \qquad \text{advective form} \tag{48.70a}$$

$$\frac{\partial \overline{C}}{\partial t} + \nabla \cdot \left([\overline{\boldsymbol{v}} + \mathcal{U}^{(\Psi)}]\, \overline{C}\right) = \nabla \cdot (\mathbb{K} \cdot \nabla \overline{C}) \qquad \text{flux form,} \tag{48.70b}$$

where we made use of the identities

$$\nabla \cdot \overline{\boldsymbol{v}} = 0 \qquad \nabla \cdot \mathcal{U}^{(\Psi)} = 0 \qquad \nabla \cdot \mathcal{U}^{(K)} \neq 0. \tag{48.71}$$

48.3.4 Connection to Stokes drift

From equation (48.24) we have the leading order expression for the Stokes drift

$$\overline{v}_p^{(S)} = \overline{\xi_n\, \partial_n\, \partial_t\, \xi_p} + \mathcal{O}(\alpha^2). \tag{48.72}$$

As noted in equation (48.15), with $\partial_t \boldsymbol{\xi} = \boldsymbol{v}'$ and with $\nabla \cdot \boldsymbol{v}' = 0$, the corresponding particle displacements are non-divergent, $\nabla \cdot \boldsymbol{v}' = 0 \Rightarrow \nabla \cdot \boldsymbol{\xi} = 0$. Consequently, to second order accuracy, the Stokes drift velocity can be written

$$\overline{v}_p^{(S)} = \overline{\xi_n\, \partial_n\, \partial_t\, \xi_p} \tag{48.73a}$$

$$= \partial_n \left[\overline{(\partial_t\, \xi_p)\, \xi_n}\right] \tag{48.73b}$$

$$= \partial_n (\mathbb{K}_{pn} + \mathbb{A}_{pn}) \tag{48.73c}$$

$$= \partial_n (\mathbb{K}_{np} - \mathbb{A}_{np}) \tag{48.73d}$$

$$= -\mathcal{U}_p^{(K)} + \mathcal{U}_p^{(\Psi)}. \tag{48.73e}$$

[5]Note 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 (48.69a), whereas they define $\mathcal{U}_n^{(\Psi)} = -\partial_m \mathbb{A}_{mn}$ as in equation (48.69b).

For the case of periodic waves, the Stokes drift velocity equals to the non-divergent skew velocity

$$\overline{\boldsymbol{v}}^{(S)} = \mathcal{U}^{(\Psi)} \Rightarrow \nabla \cdot \overline{\boldsymbol{v}}^{(S)} = 0 \qquad \text{periodic waves.} \tag{48.74}$$

More generally, for non-periodic waves the divergent velocity is non-zero so that the Stokes velocity is also divergent

$$\overline{\boldsymbol{v}}^{(S)} = \mathcal{U}^{(\Psi)} - \mathcal{U}^{(K)} \Rightarrow \nabla \cdot \overline{\boldsymbol{v}}^{(S)} = -\nabla \cdot \mathcal{U}^{(K)} \neq 0 \qquad \text{non-periodic waves.} \tag{48.75}$$

48.3.5 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

$$\boldsymbol{\xi}(\boldsymbol{x}, t) = \Gamma\,[\hat{\boldsymbol{x}}\,\cos(k\,x - \omega\,t) + \hat{\boldsymbol{y}}\,\sin(k\,x - \omega\,t)] \tag{48.76a}$$

$$\partial_t \boldsymbol{\xi}(\boldsymbol{x}, t) = \omega\,\Gamma\,[\hat{\boldsymbol{x}}\,\sin(k\,x - \omega\,t) - \hat{\boldsymbol{y}}\,\cos(k\,x - \omega\,t)], \tag{48.76b}$$

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

$$\boldsymbol{\xi} \cdot \boldsymbol{\xi} = \Gamma^2. \tag{48.77}$$

We are motivated to let the mean operator be a phase average

$$\overline{\phi} = \frac{1}{T} \int_0^T \phi(t)\,\mathrm{d}t, \tag{48.78}$$

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, $\overline{C}$, remains unchanged by these waves. The absence of a rectified change to $\overline{C}$ reflects the linear periodic nature of the wave field.

Symmetric mixing tensor

The symmetric mixing tensor

$$\mathbb{K}_{mn} = \frac{\Gamma^2}{2\,T} \int_0^T \mathrm{d}t\, \frac{\partial}{\partial t} \begin{bmatrix} \cos^2(k\,x - \omega\,t) & \cos(k\,x - \omega\,t)\,\sin(k\,x - \omega\,t) \\ \cos(k\,x - \omega\,t)\,\sin(k\,x - \omega\,t) & \sin^2(k\,x - \omega\,t) \end{bmatrix} \tag{48.79}$$

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

$$\mathbb{A}_{12} = -\mathbb{A}_{21} = \frac{\Gamma^2\,\omega}{2}\,[\sin^2(k\,x - \omega\,t) + \cos^2(k\,x - \omega\,t)] = \frac{\Gamma^2\,\omega}{2}, \tag{48.80}$$

which reflects the counter-clockwise polarization. The corresponding vector streamfunction is vertical

$$\boldsymbol{\Psi} = \frac{\Gamma^2\,\omega}{2}\,\hat{\boldsymbol{z}}, \tag{48.81}$$

Figure 48.2: *Sketch of the various tracer fluxes associated with the polarized displacement vector (48.76a). The particles are moving on the horizontal plane in a circle with time-independent radius Γ. The vector streamfunction (48.81) points in the negative $\hat{\boldsymbol{z}}$ direction. The mean concentration gradient, $\nabla\overline{C}$, 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.*

and the skew flux is horizontal

$$\boldsymbol{F}^{(\mathrm{skew})} = \frac{\Gamma^2\,\omega}{2}\,(\hat{\boldsymbol{z}} \wedge \nabla\overline{C}). \tag{48.82}$$

Finally, the advective velocity is given by

$$\nabla \wedge \boldsymbol{\Psi} = \omega\,\Gamma\,\nabla\Gamma \wedge \hat{\boldsymbol{z}}, \tag{48.83}$$

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 $\overline{C}$. Figure 48.2 offers a schematic of the skew, advective, and rotational fluxes induced by a linear rotating particle wave in the horizontal plane.

48.3.6 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.

48.4 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 50.2.4 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 48.6, at which point we also consider eddy correlations between velocity and tracer as in Section 48.3.

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 48.2. 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.

48.4.1 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 48.4.3, we relate this isopycnal approach to the vertical/isopycnal GLM.

Defining the isopycnal ensemble

An overbar with a potential density label, $\overline{(\ )}^{(\sigma)}$, 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 8 and 38), with the mean field denoted by

$$\overline{\Phi}^{(\sigma)}(x, y, \sigma, t) \equiv \text{ensemble mean using isopycnal vertical coordinates.} \tag{48.84}$$

This average is straightforward to compute when using isopycnal coordinates, thus producing an isopycnal mean that is a function of the potential density, σ.

48.4.2 Modified mean

As a complement to the isopycnal approach in Section 48.4.1, we here introduce the vertical/isopycnal GLM, also known as the modified mean.

Figure 48.3: *Schematic of the ensemble mean depth $\overline{z}^{\rho}(x, y, \rho, t)$ 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, $2\,\overline{z}^{\rho}(\rho) = z(1, \rho) + z(2, \rho)$, where the depth $z(1, \rho)$ is generally different from $z(2, \rho)$.*

Vertical/isopycnal GLM

The discussion in Section 48.2 considered a three-vector particle displacement vector $\boldsymbol{\xi}(\boldsymbol{x}, t)$. In contrast, we are here interested just in the vertical displacement of an isopycnal layer interface

$$\boldsymbol{\xi}(x, y, \sigma, t) = \hat{\boldsymbol{z}}\, \xi(x, y, \sigma, t). \tag{48.85}$$

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)

$$z(\sigma) = \overline{z}^{(\sigma)} + \xi(\sigma), \tag{48.86}$$

where

$$\overline{z}^{(\sigma)} = \overline{z(\sigma)}^{(\sigma)} \tag{48.87}$$

is the isopycnal ensemble mean depth, and the displacement field has a zero ensemble mean

$$\overline{\xi(\sigma)}^{(\sigma)} = 0. \tag{48.88}$$

Given the above definitions for the vertical position, we define the *vertical/isopycnal GLM* for an arbitrary function

$$\widetilde{\Phi}(x, y, \overline{z}^{(\sigma)}, t) \equiv \overline{\Phi(x, y, \overline{z}^{(\sigma)} + \xi(x, y, \sigma, t), t)}^{(\sigma)}. \tag{48.89}$$

As defined, the vertical/isopycnal GLM, $\widetilde{\Phi}$, is a function of the ensemble mean vertical position, $\overline{z}^{(\sigma)}$ (left hand side), and is determined by an ensemble mean of Φ sampled at the depth of each ensemble member, $\overline{z}^{(\sigma)} + \xi(\sigma)$. *McDougall and McIntosh* (2001) refer to the vertical/isopycnal GLM (48.89) as the *modified mean*.

Relating the modified mean to the isopycnal mean

Following the general result (48.34), we know that the modified mean potential density, $\widetilde{\sigma}(x, y, \overline{z}^{(\sigma)}, t)$, equals to the potential density of each ensemble member so that

$$\widetilde{\sigma}(x, y, \overline{z}^{(\sigma)}, t) = \sigma(x, y, \overline{z}^{(\sigma)} + \xi(x, y, \sigma, t), t). \tag{48.90}$$

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 (48.84)), when evaluated at the modified mean potential density, $\widetilde{\sigma}$, equals to the modified mean $\widetilde{\Phi}$ when evaluated at the vertical position of the mean density

$$\overline{\Phi}^{(\sigma)}(x, y, \widetilde{\sigma}, t) = \widetilde{\Phi}(x, y, \overline{z}^{(\sigma)}, t). \tag{48.91}$$

This is a very important identity that will be used in the following.

48.4.3 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

$$\widehat{\Phi}(\sigma) = \frac{\overline{h\,\Phi}^{(\sigma)}}{\overline{h}^{(\sigma)}}. \tag{48.92}$$

The identity (48.90) then renders

$$\overline{\Phi}^{\#}(x, y, \overline{z}^{(\sigma)}, t) \equiv \widehat{\Phi}(x, y, \widetilde{\sigma}, t), \tag{48.93}$$

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

$$\widehat{\boldsymbol{u}}(x, y, \widetilde{\sigma}, t) = \overline{\boldsymbol{u}}^{\#}(x, y, \overline{z}^{(\sigma)}, t). \tag{48.94}$$

Following the discussion of the vertical gauge in Section 47.6.1 (see in particular equation (47.77)), we are led to define the depth integrated TRM transport

$$\overline{\boldsymbol{U}}^{\#}(\overline{z}^{(\sigma)}) = \int_{-H}^{\overline{z}^{(\sigma)}} \overline{\boldsymbol{u}}^{\#}(z)\,\mathrm{d}z = \int_{\sigma(-H)}^{\widetilde{\sigma}(\overline{z}^{(\sigma)})} \widehat{\boldsymbol{u}}(\gamma)\,\overline{h}^{(\gamma)}\,\mathrm{d}\gamma, \tag{48.95}$$

with the second equality following from a change of coordinates from geopotential to isopycnal. We can go further with this expression by writing

$$\overline{\boldsymbol{U}}^{\#}(\overline{z}^{(\sigma)}) = \int_{\sigma(-H)}^{\widetilde{\sigma}(\overline{z}^{(\sigma)})} \widehat{\boldsymbol{u}}(\gamma)\,\overline{h}^{(\gamma)}\,\mathrm{d}\gamma \qquad \text{from equation (48.95)} \tag{48.96a}$$

$$= \int_{\sigma(-H)}^{\widetilde{\sigma}(\overline{z}^{(\sigma)})} \overline{\boldsymbol{u}\,h}^{(\gamma)}\,\mathrm{d}\gamma \qquad \text{from equation (48.92)} \tag{48.96b}$$

$$= \int_{\sigma(-H)}^{\sigma(\overline{z}^{(\sigma)}+\xi)} \overline{\boldsymbol{u}\,h}^{(\gamma)}\,\mathrm{d}\gamma \qquad \text{from equation (48.90).} \tag{48.96c}$$

The final equality makes it clear that the TRM transport, $\overline{\boldsymbol{U}}^{\#}(\overline{z}^{(\sigma)})$, is the ensemble mean volume transport for fluid denser than $\sigma(\overline{z}^{(\sigma)} + \xi) = \widetilde{\sigma}(\overline{z}^{(\sigma)})$. This transport can also be written using geopotential coordinates

$$\overline{\boldsymbol{U}}^{\#}(\overline{z}^{(\sigma)}) = \overline{\int_{-H}^{\overline{z}^{(\sigma)}+\xi} \boldsymbol{u}\,\mathrm{d}z}. \tag{48.97}$$

The transport from each ensemble member is determined by integrating from the bottom to the depth, $\overline{z}^{(\sigma)} + \xi$, and then the TRM transport is determined by computing the ensemble mean for this transport.

Quasi-Stokes transport

The TRM transport (48.97) can be decomposed into an Eulerian mean plus the correlation of a fluctuation

$$\overline{\boldsymbol{U}}^{\#}(\overline{z}^{(\sigma)}) \equiv \overline{\boldsymbol{U}}(\overline{z}^{(\sigma)}) + \overline{\boldsymbol{U}}^{\mathrm{qs}}(\overline{z}^{(\sigma)}). \tag{48.98}$$

The first term,

$$\overline{\boldsymbol{U}}(\overline{z}^{(\sigma)}) = \int_{-H}^{\overline{z}^{(\sigma)}} \boldsymbol{u}\,\mathrm{d}z \tag{48.99}$$

is the ensemble mean transport between the bottom and the ensemble mean depth, $\overline{z}^{(\sigma)}$. We interpret this transport as an Eulerian mean since the depth ranges are fixed. In contrast, the *quasi-Stokes* transport

$$\overline{\boldsymbol{U}}^{\mathrm{qs}}(\overline{z}^{(\sigma)}) \equiv \overline{\int_{\overline{z}^{(\sigma)}}^{\overline{z}^{(\sigma)}+\xi} \boldsymbol{u}\,\mathrm{d}z} \tag{48.100}$$

measures the ensemble mean transport between the mean vertical position of an isopycnal, $\overline{z}^{(\sigma)}$, and that of each ensemble member, $\overline{z}^{(\sigma)} + \xi(\sigma)$. 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 48.2)

$$\overline{\boldsymbol{U}}^{\mathrm{qs}} = \overline{\boldsymbol{U}}^{\#} - \overline{\boldsymbol{U}}. \tag{48.101}$$

As for the traditional Stokes drift discussed in Sections 45.3, and 48.2.5, 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 (47.76), we introduce the TRM vector streamfunction

$$\overline{\boldsymbol{\Psi}}^{\#} = \overline{\boldsymbol{U}}^{\#} \wedge \hat{\boldsymbol{z}}, \tag{48.102}$$

and the corresponding three-dimensional non-divergent TRM velocity

$$\overline{\boldsymbol{v}}^{\#} = \nabla \wedge \overline{\boldsymbol{\Psi}}^{\#}. \tag{48.103}$$

The vertical component,

$$\overline{w}^{\#} = \hat{\boldsymbol{z}} \cdot (\nabla \wedge \overline{\boldsymbol{\Psi}}^{\#}), \tag{48.104}$$

has no corresponding component in an isopycnal description, which only requires the horizontal thickness weighted transport, $\widehat{\boldsymbol{u}}$. However, the TRM vector streamfunction only requires the horizontal TRM transport, $\overline{\boldsymbol{U}}^{\#}$, so the two descriptions in effect make use of the same number of degrees of freedom.

48.4.4 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

$$\delta V = \delta x \, \delta y \, \delta z = \delta x \, \delta y \, \delta \sigma \, h, \tag{48.105}$$

where we introduced the specific thickness

$$h = \frac{\partial z}{\partial \sigma}. \tag{48.106}$$

As discussed in Section 8.8.1, 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 21.4.5)

$$N^2 = -\frac{g}{\rho_0} \frac{\partial \sigma}{\partial z} = -\frac{g}{\rho_0 \, h} \tag{48.107}$$

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)

$$\frac{\partial h}{\partial t} + \nabla_\sigma \cdot (h \, \boldsymbol{u}) = 0, \tag{48.108}$$

with $\boldsymbol{u}$ the horizontal velocity field, the time derivative is computed with σ held fixed, and

$$\nabla_\sigma = \nabla_z + \boldsymbol{S} \frac{\partial}{\partial z} \tag{48.109}$$

is the horizontal derivative operator with σ held fixed and

$$\boldsymbol{S} = \nabla_\sigma z \tag{48.110}$$

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, $\nabla \cdot \boldsymbol{v} = 0$, and material conservation of potential density, $\mathrm{D}\sigma/\mathrm{D}t = 0$. When written as skewsion rather than advection, the natural gauge is the vertical gauge introduced in Section 47.6.1 (equation (47.74)), since this gauge only requires the same horizontal velocity field $\boldsymbol{u}$ used with the isopycnal coordinate description. This gauge has an associated skew flux of potential density $\boldsymbol{F}^{(\mathrm{skew})} = -\nabla \sigma \wedge \boldsymbol{\Psi}$, which leads to the evolution

$$\frac{\partial \sigma}{\partial t} = \nabla \cdot (\nabla \sigma \wedge \boldsymbol{\Psi}), \tag{48.111}$$

where all derivatives are here taken with fixed Eulerian (geopotential) coordinates, (x, y, z), and the divergence operator is three-dimensional.

Figure 48.4: *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, $2\,\overline{\rho}^z(z) = \rho(1, z) + \rho(2, z)$, where $\rho(1, z)$ is generally different from $\rho(2, z)$.*

48.4.5 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 (48.108) is satisfied by each ensemble member

$$\frac{\partial h}{\partial t} + \nabla_\sigma \cdot (h\, \boldsymbol{u}) = 0. \tag{48.112}$$

The ensemble mean computed over these fluid parcels with potential density σ satisfies

$$\partial_t\, \overline{h}^{(\sigma)} + \nabla_\sigma \cdot \Big(\overline{h}^{(\sigma)}\, \overline{\boldsymbol{u}}^{(\sigma)} + \overline{h'\, \boldsymbol{u}'}^{(\sigma)} \Big) = 0, \tag{48.113}$$

where primed variables represent deviations from the isopycnal mean. The mean specific thickness $\overline{h}^{(\sigma)}$ of parcels with potential density σ therefore satisfies the conservation equation

$$\partial_t\, \overline{h}^{(\sigma)} + \nabla_\sigma \cdot (\overline{h}^{(\sigma)}\, \widehat{\boldsymbol{u}}) = 0. \tag{48.114}$$

In this equation we introduced the thickness weighted isopycnal ensemble mean horizontal velocity

$$\widehat{\boldsymbol{u}} = \frac{\overline{h\, \boldsymbol{u}}^{(\sigma)}}{\overline{h}^{(\sigma)}} = \overline{\boldsymbol{u}}^{(\sigma)} + \frac{\overline{h'\, \boldsymbol{u}'}^{(\sigma)}}{\overline{h}^{(\sigma)}} \equiv \overline{\boldsymbol{u}}^{(\sigma)} + \boldsymbol{u}^{\text{bolus}}, \tag{48.115}$$

along with the isopycnal ensemble mean horizontal velocity, $\overline{\boldsymbol{u}}^{(\sigma)}$, and the horizontal *bolus velocity*, $\boldsymbol{u}^{\text{bolus}}$ originally introduced by *Rhines* (1982). The bolus velocity for an isopycnal layer corresponds to the transport

$$\overline{h}^{(\sigma)}\, \boldsymbol{u}^{\text{bolus}} = \overline{h}^{(\sigma)}\, (\widehat{\boldsymbol{u}} - \overline{\boldsymbol{u}}^{(\sigma)}) = \overline{h'\, \boldsymbol{u}'}^{(\sigma)} \tag{48.116}$$

arises from the along-isopycnal correlations between specific thickness and horizontal velocity.

Quite conveniently, the mean conservation equation (48.114) takes the *same* mathematical form as the conservation equation (48.112) satisfied by each ensemble member. The key difference is that the isopycnal ensemble mean thickness $\overline{h}^{(\sigma)}$ is stirred by the thickness weighted isopycnal ensemble mean horizontal velocity $\widehat{\boldsymbol{u}}$, whereas the thickness of each ensemble member is stirred by a randomly different realization of the horizontal velocity $\boldsymbol{u}$. The simplicity of the mean field description (48.114) is afforded by use of the Lagrangian vertical coordinate σ.

48.4.6 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, $\overline{z}^{(\sigma)}$. Consequently, mean fields defined at the fixed vertical position correspond to either modified mean fields when not thickness weighted (equation (48.89)), or TRM fields when thickness weighted (equation (48.93)).

Evolution of modified mean density

Following the skewsion formulation from Section 47.6, at the ensemble mean depth $z = \overline{z}^{\rho}$, the streamfunction $\overline{\boldsymbol{\Psi}}^{\#}$ defines an effective skew flux of the modified mean potential density given by

$$\overline{\boldsymbol{F}}^{\#} = -\nabla\widetilde{\sigma} \wedge \overline{\boldsymbol{\Psi}}^{\#}. \tag{48.117}$$

Using the identity $\overline{\boldsymbol{\Psi}}^{\#} = \overline{\boldsymbol{U}}^{\#} \wedge \hat{\boldsymbol{z}}$, we can write this expression in one of the forms

$$\overline{\boldsymbol{F}}^{\#} = -\overline{\boldsymbol{U}}^{\#}\, \partial_z\widetilde{\sigma} + \hat{\boldsymbol{z}}\, \overline{\boldsymbol{U}}^{\#} \cdot \nabla_z\widetilde{\sigma} \tag{48.118a}$$

$$= -(\overline{\boldsymbol{U}}^{\#} + \hat{\boldsymbol{z}}\, \boldsymbol{S} \cdot \overline{\boldsymbol{U}}^{\#})\, \partial_z\widetilde{\sigma}, \tag{48.118b}$$

where

$$\boldsymbol{S} = -\frac{\nabla_z\widetilde{\sigma}}{\partial_z\widetilde{\sigma}} \tag{48.119}$$

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 $z = \overline{z}^{(\sigma)}$. The convergence of the effective skew flux leads to a stirring of the modified mean density $\widetilde{\sigma}$ at the mean depth $z = \overline{z}^{(\sigma)}$,

$$\frac{\partial\widetilde{\sigma}}{\partial t} = \nabla \cdot (\nabla\widetilde{\sigma} \wedge \overline{\boldsymbol{\Psi}}^{\#}). \tag{48.120}$$

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 (48.111) satisfied at depth z by a single member of the ensemble.

48.4.7 Approximate ensemble mean kinematics in geopotential coordinates

Equation (48.120) 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 $\overline{\boldsymbol{U}}^{\mathrm{qs}}$ defined by equation (48.100). We addressed a similar estimation in Section 48.2.4 when discussing the Stokes mean. Here, we expand the TRM transport in a Taylor series about the geopotential $z = \overline{z}^{(\sigma)}$

$$\overline{\boldsymbol{U}}^{\#}(z) = \overline{\int_{-H}^{z+\xi} \boldsymbol{u}(s)\, \mathrm{d}s} \tag{48.121a}$$

$$= \overline{\boldsymbol{U}}(z) + \overline{\boldsymbol{u}\,\xi}^{(z)} + \frac{1}{2}\, \overline{\partial_z \boldsymbol{u}\, \xi\, \xi}^{(z)} + \mathcal{O}(\alpha^3), \tag{48.121b}$$

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 $z = \overline{z}^{(\sigma)}$.

The ensemble means in equation (48.121b) are interpreted as follows. The first term is the Eulerian mean horizontal transport passing beneath the ensemble mean depth, $z = \overline{z}^{(\sigma)}$. The second term, $\overline{\boldsymbol{u}\,\xi}$ 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 $\boldsymbol{u}$ into its Eulerian mean $\overline{\boldsymbol{u}}^{(z)}$ and deviation $\boldsymbol{u}'$ leads to the correlation

$$\overline{\boldsymbol{u}\,\xi}^{(z)} = \overline{\boldsymbol{u}'\,\xi}^{(z)}. \tag{48.122}$$

For the second order term, similar considerations lead to

$$\overline{\partial_z \boldsymbol{u}\,\xi\,\xi}^{(z)} \approx \partial_z \overline{\boldsymbol{u}}^z\,\overline{\xi\,\xi}^{(z)}, \tag{48.123}$$

where neglected terms are third order and higher. Combining these relations leads to the second order accurate expression

$$\overline{\boldsymbol{U}}^{\#} \approx \overline{\boldsymbol{U}} + \overline{\boldsymbol{u}'\,\xi}^{(z)} + \frac{1}{2}\,\overline{\xi\,\xi}^{z}\,\partial_z \overline{\boldsymbol{u}}^{(z)}. \tag{48.124}$$

The disturbance field

Following the discussion in Section 48.2.6, we here determine the disturbance field, ξ, in terms of fields at constant depth. For this purpose, use the identity (48.90) to give

$$\widetilde{\sigma}(z) = \sigma(z+\xi) \tag{48.125a}$$

$$= \sigma(z) + \partial_z \sigma(z)\,\xi + \frac{1}{2}\,\partial_{zz}\sigma(z)\,\xi^2 + \mathcal{O}(\alpha^3). \tag{48.125b}$$

Subtracting the Eulerian mean of equation (48.125b) from the unaveraged equation (48.125b), and noting that $\widetilde{\sigma}$ is already a mean field, leads to the second order accurate expression for the deviation

$$\xi = -\sigma'(z)/\partial_z \overline{\sigma}^{(z)} + \mathcal{O}(\alpha^2), \tag{48.126}$$

where

$$\sigma(z) = \overline{\sigma}^{(z)} + \sigma'(z). \tag{48.127}$$

To within the same order, the deviation can be written

$$\xi = -\sigma'(z)/\partial_z \widetilde{\sigma}(z) + \mathcal{O}(\alpha^2). \tag{48.128}$$

Approximate quasi-Stokes transport

Substituting the deviation (48.128) into the approximate expression (48.121b) for the TRM transport yields an approximate expression for the Stokes transport

$$\boldsymbol{U}^{\mathrm{qs}} = -\frac{\overline{\boldsymbol{u}'\,\sigma'}^{(z)}}{\partial_z \widetilde{\sigma}} + \frac{\overline{\phi}^{(z)}\,\partial_z \overline{\boldsymbol{u}}^{(z)}}{(\partial_z \widetilde{\sigma})^2} + \mathcal{O}(\alpha^3), \tag{48.129}$$

where

$$\overline{\phi}^{(z)} = \frac{1}{2}\,\overline{\sigma'\,\sigma'}^{(z)} \tag{48.130}$$

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 (48.129). We have more to say regarding this parameterization in Section 49.1.

Substituting the deviation (48.128) into the approximate expression (48.125b) yields, to within terms of third order, the relation

$$\widetilde{\sigma} = \overline{\sigma}^{(z)} - \partial_z \left[\frac{\overline{\phi}^{(z)}}{\partial_z \overline{\sigma}^{(z)}} \right] + \mathcal{O}(\alpha^3). \tag{48.131}$$

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.

48.4.8 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).

48.5 Thickness transport for a linear longitudinal wave

We here study the Stokes drift within a layer of constant density fluid, showing that this drift leads to a net transport of volume per horizontal area; i.e., the layer thickness. To simplify the maths, we focus on transport from linear longitudinal waves. This discussion provides a specific example of the general considerations of thickness transport by eddies presented in Section 48.4.5. Indeed, much of our intuition for the general case is based on this relatively simple worked example.[6]

48.5.1 An undulating fluid layer

Figure 48.5 shows a layer of constant density fluid within a stably stratified fluid. The total volume of fluid within this layer is assumed to remain constant, which means the layer does not mix with surrounding fluid layers; i.e., it is an immiscible fluid layer. 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

$$h(y,t) = h_o + h'(y,t). \tag{48.132}$$

The layer thickness changes in time according to the convergence of the advective transport of thickness

$$\frac{\partial h}{\partial t} = -\nabla \cdot (h\,\mathbf{u}), \tag{48.133}$$

where the convergence is computed within the layer. We offer a derivation of this volume conservation equation in Section 29.1.3. Even without working through that derivation, the truth of this equation follows from Figure 48.5, whereby 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

$$\frac{\partial h}{\partial t} = -\frac{\partial\,(h\,v)}{\partial y}. \tag{48.134}$$

[6]This worked example is based Section 2 of *Lee et al.* (1997).

Figure 48.5: *Shown here is a single layer of constant density fluid, with resting thickness $h = h_o$ and instantanous thickness $h = h_o + h'(y,t)$. Associated with the undulations in thickness are fluctuations in the meridional velocity $v' = v_o \sin(k\,y - \omega\,t)$, 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.*

48.5.2 Stokes drift

Consider a linear wave perturbation in the meridional velocity that propagates in the meridional direction

$$v'(y,t) = v_o \sin(k\,y - \omega\,t), \tag{48.135}$$

where k is a constant wave number and ω is a constant frequency. This longitudinal wave is depicted in Figure 48.5. We now follow the general formalism developed in Section 45.3 or equivalently in Section 48.2.4 to determine the Stokes drift associated with this wave.

We are only concerned with the meridional component of the velocity, so the fluid particle trajectory equation is given by

$$\frac{\mathrm{d}Y}{\mathrm{d}t} = v_o \sin(k\,Y - \omega\,t), \tag{48.136}$$

where $Y = Y(Y_o, t)$ is the meridional trajectory with initial postion Y_o. Following equation (45.43) 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)

$$\frac{\mathrm{d}Y}{\mathrm{d}t} - v(y,t) = v_o^2\,k\,\cos(ky - \omega t)\int_0^t \sin(ky - \omega t')\,\mathrm{d}t' \tag{48.137a}$$

$$= \frac{v_o^2\,k}{\omega}\left(\cos^2(ky - \omega t) - \cos(ky - \omega t)\;\cos(ky)\right). \tag{48.137b}$$

Time averaging over a single wave period,

$$T = 2\pi/\omega \tag{48.138}$$

leads to the Stokes drift as per the general expression in equation (45.45)

$$V_{\text{Stokes}} = \frac{v_o^2\,k}{2\omega}. \tag{48.139}$$

Introducing the phase speed for the wave $c = \omega/k$ allows us to write the Stokes drift as

$$V_{\text{Stokes}} = \frac{v_o^2}{2\,c}. \tag{48.140}$$

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.

48.5.3 Linearized thickness perturbations

The velocity and thickness are written in terms of their rest state plus a perturbation due to the wave

$$h = h_o + h' \tag{48.141a}$$
$$v = v', \tag{48.141b}$$

where the velocity vanishes when the wave is absent. The thickness equation (48.134) thus takes the form

$$\frac{\partial h'}{\partial t} + h_o \frac{\partial v'}{\partial y} + v' \frac{\partial h'}{\partial y} = 0. \tag{48.142}$$

Linearizing this equation, and using the wave perturbation (48.135), leads to

$$\frac{\partial h'}{\partial t} + h_o \, v_o \, k \, \cos(ky - \omega t) = 0, \tag{48.143}$$

thus yielding the thickness perturbation

$$h' = h_o \frac{v'}{c}. \tag{48.144}$$

Hence, to leading order, the thickness perturbation is directly proportional to and in phase with the velocity perturbation.

48.5.4 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

$$\overline{h' \, v'} = \frac{1}{T} \int_0^T h' \, v' \, \mathrm{d}t \tag{48.145a}$$
$$= \frac{h_o}{c \, T} \int_0^T v' \, v' \, \mathrm{d}t \tag{48.145b}$$
$$= \frac{v_o^2 \, h_o}{c \, T} \int_0^T \sin^2(ky - \omega t) \, \mathrm{d}t \tag{48.145c}$$
$$= \frac{v_o^2 \, h_o}{2 \, c} \tag{48.145d}$$
$$= h_o \, V_{\text{Stokes}}, \tag{48.145e}$$

where we introduced the Stokes drift (48.140) 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.

48.5.5 Comments

The nonzero correlation in equation (48.145e) 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, such as considered in the ensemble mean isopycnal layer transport discussed in Section 48.4.5. For the general case, a nonzero bolus velocity (48.116), 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.

48.6 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.

48.6.1 Thickness weighted means

Equation (48.115) 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

$$\Phi(\sigma) = \widehat{\Phi}(\sigma) + \Phi''(\sigma) \tag{48.146a}$$

$$= \frac{\overline{h\,\Phi}^{(\sigma)}}{\overline{h}^{(\sigma)}} + \Phi''. \tag{48.146b}$$

It follows by definition that

$$\overline{h\,\Phi''}^{(\sigma)} = 0. \tag{48.147}$$

48.6.2 Isopycnal mean thickness weighted tracer equation

When attaching a tracer to fluid parcels, each member of the ensemble satisfies the isopycnal tracer equation

$$\frac{\partial C}{\partial t} + \boldsymbol{u} \cdot \nabla_\sigma C = 0. \tag{48.148}$$

Combining the tracer and thickness equations leads to the thickness weighted tracer equation

$$\frac{\partial (h\,C)}{\partial t} + \nabla_\sigma \cdot (h\,\boldsymbol{u}\,C) = 0. \tag{48.149}$$

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, $h\,\boldsymbol{u}\,C$.

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

$$\partial_t [h\,(\widehat{C} + C'')] + \nabla_\sigma \cdot [h\,(\widehat{\boldsymbol{u}} + \boldsymbol{u}'')\,(\widehat{C} + C'')] = 0. \tag{48.150}$$

Taking an ensemble average over fluid elements with the same potential density, and using equation (48.147), yield the mean thickness weighted tracer equation

$$\partial_t(\overline{h}^{(\sigma)}\,\widehat{C}) + \nabla_\sigma \cdot (\overline{h}^{(\sigma)}\,\widehat{C}\,\widehat{\boldsymbol{u}}) = -\nabla_\sigma \cdot (\overline{h\,C''\,\boldsymbol{u}''}^{(\sigma)}). \tag{48.151}$$

Now introduce the correlation

$$\overline{h\,C''\,\boldsymbol{u}''}^{(\sigma)} = \overline{h}^{(\sigma)}\,\widehat{C''\,\boldsymbol{u}''} \tag{48.152}$$

(see equation (48.146b)), and recall that the mean thickness $\overline{h}^{(\sigma)}$ satisfies the mean thickness equation (48.114). These two points lead to the evolution equation for the mean thickness weighted tracer concentration

$$(\partial_t + \widehat{\boldsymbol{u}} \cdot \nabla_\sigma)\,\widehat{C} = -\frac{1}{\overline{h}^{(\sigma)}}\,\nabla_\sigma \cdot (\overline{h}^{(\sigma)}\,\widehat{C''\,\boldsymbol{u}''}). \tag{48.153}$$

48.6.3 Subgrid scale tracer transport tensor

The correlation between tracer and velocity found on the right-hand side of the mean thickness weighted tracer equation (48.153) is typically written in terms of a subgrid scale tracer transport tensor

$$\widehat{C''\,\boldsymbol{u}''} = -\mathbb{J} \cdot \nabla_\sigma \widehat{C}. \tag{48.154}$$

This definition leads to the evolution equation

$$(\partial_t + \widehat{\boldsymbol{u}} \cdot \nabla_\sigma)\,\widehat{C} = \frac{1}{\overline{h}^{(\sigma)}}\,\nabla_\sigma \cdot (\overline{h}^{(\sigma)}\,\mathbb{J} \cdot \nabla_\sigma \widehat{C}). \tag{48.155}$$

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 8.14. However, in addition to symmetric diffusion processes, this operator includes skewed fluxes that lead to skew diffusion as discussed in Section 48.3.2. Whereas the diffusive aspect is commonly parameterized as dia-neutral diffusion and neutral diffusion (Section 49.1), there is no parameterization for the skewed correlations for use in ocean models. We comment further on this situation in Section 49.3.8.

48.6.4 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 $\sigma = \widetilde{\sigma}$,

$$\overline{\boldsymbol{C}}^{\#}(\overline{z}^{(\sigma)}) = \overline{\int_{-H}^{\overline{z}^{\sigma}+\xi} C\,\boldsymbol{u}\,\mathrm{d}z}. \tag{48.156}$$

Setting tracer concentration to unity recovers the expression (48.97) for the TRM transport. Changing coordinates and making use of the tracer correlation tensor renders

$$\overline{\boldsymbol{C}}^{\#}(\overline{z}^{(\sigma)}) = \int_{\sigma(-H)}^{\widetilde{\sigma}(\overline{z}^{(\sigma)})} \overline{C\,\boldsymbol{u}\,h}^{(\sigma)}\,\mathrm{d}\sigma \tag{48.157a}$$

$$= \int_{\sigma(-H)}^{\widetilde{\sigma}(\overline{z}^{(\sigma)})} \overline{h}^{(\sigma)}\,\mathrm{d}\sigma\,(\widehat{C}\,\widehat{\boldsymbol{u}} + \widehat{C''\,\boldsymbol{u}''}) \tag{48.157b}$$

$$= \int_{\sigma(-H)}^{\widetilde{\sigma}(\overline{z}^{(\sigma)})} \overline{h}^{(\sigma)}\,\mathrm{d}\sigma\,(\widehat{C}\,\widehat{\boldsymbol{u}} - \mathbb{J} \cdot \nabla_\sigma \widehat{C}) \tag{48.157c}$$

$$= \int_{-H}^{\overline{z}^{(\sigma)}} \mathrm{d}z\,(\widehat{C}\,\widehat{\boldsymbol{u}} - \mathbb{J} \cdot \nabla_\sigma \widehat{C}). \tag{48.157d}$$

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.

48.6.5 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 $\boldsymbol{u}^{\text{bolus}}$, which again is related to the thickness weighted horizontal velocity via

$$\widehat{\boldsymbol{u}}((\sigma)) = \frac{\overline{h\,\boldsymbol{u}}^{(\sigma)}}{\overline{h}^{(\sigma)}} = \overline{\boldsymbol{u}}^{(\sigma)} + \frac{\overline{h'\,\boldsymbol{u}'}^{(\sigma)}}{\overline{h}^{(\sigma)}} = \overline{\boldsymbol{u}}^{(\sigma)} + \boldsymbol{u}^{\text{bolus}}. \tag{48.158}$$

In addition to the bolus velocity, it is necessary to parameterize the subgrid scale tracer transport tensor

$$\widehat{C''\,\boldsymbol{u}''} = -\mathbb{J} \cdot \nabla_\sigma \widehat{C}, \tag{48.159}$$

which generally has symmetric (diffusive) and antisymmetric (stirring) components (Section 48.3).

For a geopotential coordinate description, equation (48.93) 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

$$\widehat{C}(x, y, \widetilde{\sigma}, t) = \overline{C}^{\#}(x, y, \overline{z}^{(\sigma)}, t). \tag{48.160}$$

Equation (48.160), and the developed formalism, leads to the mean field tracer equation in geopotential coordinates

$$\partial_t \overline{C}^{\#} = \nabla \cdot (\nabla \overline{C}^{\#} \wedge \overline{\boldsymbol{\Psi}}^{\#}) + R(\overline{C}^{\#}), \tag{48.161}$$

where $R(\overline{C}^{\#})$ is the geopotential coordinate form of the mixing/stirring operator on the right-hand side of equation (48.155). Details for the transformation of the mixing/stirring opertor from isopycnal to geopotential coordinates are provided in Section 8.14.

48.6.6 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.

49

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 47 and further unpacked in the tracer kinematics Chapter 48.

- add section on anisotropic neutral diffusion
- add section on anisotropic GM

49.1 Parameterizing eddy induced tracer transport **878**
49.1.1 Framework based on tracer variance cascade 878
49.1.2 Density changes from molecular diffusion 878
49.1.3 Mixing from small (or fine) scale processes 879
49.1.4 Mesoscale eddy-induced stirring and mixing 879
49.1.5 Synthesis 882
49.1.6 Lateral versus diapycnal diffusion 882
49.2 Quasi-Stokes induced tracer stirring **883**
49.2.1 Gent and McWilliams skewsion 884
49.2.2 Local adiabatic dissipation of potential energy 884
49.2.3 Meridional overturning streamfunction 886
49.2.4 Connection to form stress 886
49.2.5 Isopycnal thickness diffusion and GM 888
49.3 Neutral diffusion **889**
49.3.1 Redi neutral diffusion 889
49.3.2 Small slope neutral diffusion 890
49.3.3 Neutral tangent plane neutral diffusion 891
49.3.4 Neutrality condition 892

49.3.5 Symmetry condition 892
49.3.6 GM skewsion plus small slope neutral diffusion 892
49.3.7 Small slope neutral diffusion in generalized vertical coordinates . . . 893
49.3.8 Comments 894
49.4 Cabbeling and thermobaricity **894**
49.4.1 Basic manipulations 895
49.4.2 A tidy form 895
49.4.3 Cabbeling 896
49.4.4 Thermobaricity 896
49.4.5 Comments 897

49.1 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 50.2 to further detail how *in situ* density evolves in the presence of eddy parameterizations.

49.1.1 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.

49.1.2 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

$$\rho \, \frac{\mathrm{D}\Theta}{\mathrm{D}t} = \nabla \cdot [\rho \, \kappa_\Theta \, \nabla \Theta] \tag{49.1a}$$

$$\rho \, \frac{\mathrm{D}S}{\mathrm{D}t} = \nabla \cdot [\rho \, \kappa_{\mathrm{S}} \, \nabla S] \, , \tag{49.1b}$$

where $\kappa_\Theta > 0$ and $\kappa_{\mathrm{S}} > 0$ are the molecular kinematic diffusivities for Θ and S, respectively. Following equation (50.30), we see that these molecular tracer fluxes lead to the material evolution of *in situ* density

$$\frac{\mathrm{D}\rho}{\mathrm{D}t} - \frac{\omega}{c_{\mathrm{s}}^2} = -\alpha \, \nabla \cdot (\rho \, \kappa_\Theta \, \nabla \Theta) + \beta \, \nabla \cdot (\rho \, \kappa_{\mathrm{S}} \, \nabla S) \tag{49.2a}$$

$$= -\nabla \cdot [\rho \, (\kappa_\Theta \, \alpha \, \nabla \Theta - \kappa_{\mathrm{S}} \, \beta \, \nabla S)] + \rho \, (\kappa_\Theta \, \nabla \Theta \cdot \nabla \alpha - \kappa_{\mathrm{S}} \, \nabla S \cdot \nabla \beta). \tag{49.2b}$$

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.

49.1.3 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 (49.1b), 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

$$\frac{D\rho}{Dt} - \frac{\omega}{c_s^2} = -\alpha\, \nabla \cdot (\rho\kappa\, \nabla\Theta) + \beta\, \nabla \cdot (\rho\kappa\, \nabla S) \tag{49.3a}$$

$$= -\nabla \cdot [\rho\, \kappa\, (\alpha\, \nabla\Theta - \beta\, \nabla S)] + \rho\, \kappa\, (\nabla\Theta \cdot \nabla\alpha - \nabla S \cdot \nabla\beta). \tag{49.3b}$$

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).

49.1.4 Mesoscale eddy-induced stirring and mixing

Stirring from turbulent scales smaller than the grid scale is commonly parameterized by an eddy-induced stirring velocity, $\boldsymbol{v}^*$. 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 49.1.3. 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

$$\rho\, \frac{DS}{Dt} = \nabla \cdot (\rho\, \mathbb{M} \cdot \nabla S) \tag{49.4a}$$

$$\rho\, \frac{D\Theta}{Dt} = \nabla \cdot (\rho\, \mathbb{M} \cdot \nabla \Theta). \tag{49.4b}$$

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 47, it is useful to decompose the second order transport tensor into the sum of its symmetric and anti-symmetric components

$$\mathbb{M} = \mathbb{K} + \mathbb{A}. \tag{49.5}$$

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

$$\nabla \cdot (\rho\, \mathbb{A} \cdot \nabla S) = \partial_m \left(\rho A^{mn}\, \partial_n S\right) \tag{49.6a}$$
$$= \partial_m \left(\rho A^{mn}\right) \partial_n S + \rho A^{mn}\, \partial_m \partial_n S \tag{49.6b}$$
$$= -\rho v^{*n}\, \partial_n S, \tag{49.6c}$$

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

$$\rho A^{mn}\, \partial_m \partial_n S = 0 \tag{49.7}$$

since A^{mn} is anti-symmetric whereas $\partial_m \partial_n S$ is symmetric. Finally, we introduced a density-weighted eddy-induced velocity

$$\rho v^{*n} = -\partial_m \left(\rho A^{mn}\right). \tag{49.8}$$

Importantly, $\rho\, \boldsymbol{v}^*$ has a zero divergence, again due to anti-symmetry of A^{mn}

$$\nabla \cdot (\rho\, \boldsymbol{v}^*) = \partial_n \left(\rho v^{*n}\right) = -\partial_n \partial_m \left(\rho A^{mn}\right) = 0. \tag{49.9}$$

A zero-divergence for $\rho\, \boldsymbol{v}^*$ 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 47.5). 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

$$\rho \frac{\mathrm{D}^\dagger S}{\mathrm{D}t} = \nabla \cdot (\rho\, \mathbb{K} \cdot \nabla S) \tag{49.10a}$$
$$\rho \frac{\mathrm{D}^\dagger \Theta}{\mathrm{D}t} = \nabla \cdot (\rho\, \mathbb{K} \cdot \nabla \Theta), \tag{49.10b}$$

where the residual mean material time derivative is given by

$$\frac{\mathrm{D}^\dagger}{\mathrm{D}t} = \frac{\partial}{\partial t} + \boldsymbol{v}^\dagger \cdot \nabla \tag{49.11}$$

and the residual mean velocity is

$$\boldsymbol{v}^\dagger = \boldsymbol{v} + \boldsymbol{v}^*. \tag{49.12}$$

Making use of the residual mean velocity then leads to the material evolution of the *in situ* density

$$\frac{\mathrm{D}\rho}{\mathrm{D}t} - \frac{1}{c_{\mathrm{s}}^2} \frac{\mathrm{D}p}{\mathrm{D}t} = -\rho \boldsymbol{v}^* \cdot (-\alpha \nabla \Theta + \beta \nabla S) - \alpha \nabla \cdot (\rho\, \mathbb{K} \cdot \nabla \Theta) + \beta \nabla \cdot (\rho\, \mathbb{K} \cdot \nabla S), \tag{49.13}$$

which can be written in terms of the residual mean material time derivative

$$\frac{\mathrm{D}^\dagger \rho}{\mathrm{D}t} - \frac{1}{c_{\mathrm{s}}^2} \frac{\mathrm{D}^\dagger p}{\mathrm{D}t} = -\alpha \nabla \cdot (\rho\, \mathbb{K} \cdot \nabla \Theta) + \beta \nabla \cdot (\rho\, \mathbb{K} \cdot \nabla S). \tag{49.14}$$

[1] For a Boussinesq fluid, the density factor is replaced by the constant reference density, ρ_0, so that $\nabla \cdot \boldsymbol{v}^* = 0$ 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 (49.14) can be written

$$-\alpha\nabla\cdot(\rho\,\mathbb{K}\cdot\nabla\Theta)+\beta\nabla\cdot(\rho\,\mathbb{K}\cdot\nabla S) = \nabla\cdot[\rho\,\mathbb{K}\cdot(-\alpha\nabla\Theta+\beta\nabla S)]+\rho\,\nabla\alpha\cdot\mathbb{K}\cdot\nabla\Theta-\rho\,\nabla\beta\cdot\mathbb{K}\cdot\nabla S, \quad (49.15)$$

so that the *in situ* density evolves according to

$$\frac{\mathrm{D}^\dagger\rho}{\mathrm{D}t} - \frac{1}{c_{\mathrm{s}}^2}\frac{\mathrm{D}^\dagger p}{\mathrm{D}t} = -\underbrace{\nabla\cdot[\rho\,\mathbb{K}\cdot(\alpha\nabla\Theta-\beta\nabla S)]}_{\text{conservative processes}} + \underbrace{\rho\,\nabla\alpha\cdot\mathbb{K}\cdot\nabla\Theta-\rho\,\nabla\beta\cdot\mathbb{K}\cdot\nabla S.}_{\text{sources from nonlinear EOS processes}} \quad (49.16)$$

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 49.4.

- CONSERVATIVE PROCESSES AND NEUTRAL DIFFUSION: A linear equation of state is independent of pressure, so that the evolution equation (49.16) takes the form

$$\frac{\mathrm{D}^\dagger\rho}{\mathrm{D}t} = -\nabla\cdot[\rho\,\mathbb{K}\cdot(\alpha\nabla\Theta-\beta\nabla S)]\,. \quad (49.17)$$

  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

$$\mathbb{K}\cdot(\alpha\,\nabla\Theta-\beta\,\nabla S) = 0. \quad (49.18)$$

  As detailed in Section 49.3, 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

$$\frac{\mathrm{D}^\dagger\rho}{\mathrm{D}t} - \frac{1}{c_{\mathrm{s}}^2}\frac{\mathrm{D}^\dagger p}{\mathrm{D}t} = \underbrace{\rho\,\nabla\alpha\cdot\mathbb{K}\cdot\nabla\Theta-\rho\,\nabla\beta\cdot\mathbb{K}\cdot\nabla S.}_{\text{sources from nonlinear EOS processes}} \quad (49.19)$$

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 (49.18). 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 (49.18). In Section 49.3 we dive into the details of neutral diffusion.

49.1.5 Synthesis

In summary, the proposed evolution equation for *in situ* density in the presence of subgrid scale processes takes the form

$$\frac{\mathrm{D}\rho}{\mathrm{D}t} - \underbrace{\frac{1}{c_{\mathrm{s}}^2}\frac{\mathrm{D}p}{\mathrm{D}t}}_{\text{compressibility}} = -\underbrace{\boldsymbol{v}^* \cdot (-\alpha\nabla\Theta + \beta\nabla S)}_{\text{eddy-induced stirring}} \underbrace{-\alpha\nabla\cdot(\rho\kappa\,\nabla\Theta) + \beta\nabla\cdot(\rho\kappa\,\nabla S)}_{\text{small scale mixing and nonlinear EOS}}$$
$$+ \underbrace{\rho\,\nabla\alpha\cdot\mathbb{K}\cdot\nabla\Theta - \rho\,\nabla\beta\cdot\mathbb{K}\cdot\nabla S.}_{\text{nonlinear EOS processes from eddy mixing}} \quad (49.20)$$

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 (49.3b) 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, $\mathrm{D}^\dagger/\mathrm{D}t$.

- 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 (49.18).

- 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 49.4). 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

$$-\alpha\nabla\cdot(\rho\kappa\,\nabla\Theta) + \beta\nabla\cdot(\rho\kappa\,\nabla S) = -\nabla\cdot[\rho\kappa\,(\alpha\nabla\Theta - \beta\nabla S)] + \rho\,\kappa\,(\nabla\alpha\cdot\nabla\Theta - \nabla\beta\cdot\nabla S). \quad (49.21)$$

49.1.6 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

$$\text{horizontal diffusion} = \kappa_h \nabla_z^2 C \qquad \text{vertical diffusion} = \kappa_v\,\partial_{zz} C. \quad (49.22)$$

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

$$\text{horizontal diffusion} \sim (\kappa_h / L^2)\, \delta C \qquad \text{vertical diffusion} \sim (\kappa_v / H^2)\, \delta C. \tag{49.23}$$

These operators have the same scale when

$$\kappa_v = (H/L)^2\, \kappa_h. \tag{49.24}$$

Choosing $L = 10^5$ m and $H = 10^1$ m leads to

$$\kappa_v = 10^{-8}\, \kappa_h. \tag{49.25}$$

Furthermore, if $\kappa_h = 10^3$ m^2 s^{-1}, then the two operators provide a similar contribution to tracer evolution if $\kappa_v = 10^{-5}$ m^2 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.

49.2 Quasi-Stokes induced tracer stirring

As mentioned in Section 48.6.3, 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 48.3, 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 48.4, 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 dynanical 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 21.3.6). 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).

[2] Molecular diffusivities are roughly $\approx 10^{-6}$ m^2 s^{-1} for temperature and $\approx 10^{-7}$ m^2 s^{-1} for salinity and other material tracers. (SMG: check these values).

49.2.1 Gent and McWilliams skewsion

Gent et al. (1995) parameterize the divergent part of the quasi-Stokes transport by setting

$$\boldsymbol{U}^{\mathrm{qs}} = -\kappa \boldsymbol{S}. \tag{49.26}$$

In this expression, $\boldsymbol{S}$ is the slope of the potential density surfaces (equation (48.119)), 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

$$\boldsymbol{v}^* = -\partial_z \left(\kappa \, \boldsymbol{S}\right) + \hat{\boldsymbol{z}} \, \nabla_z \cdot \left(\kappa \, \boldsymbol{S}\right), \tag{49.27}$$

and the antisymmetric stirring tensor is

$$\mathbb{A}_{mn} = \left[\begin{array}{ccc} 0 & 0 & -\kappa S_x \\ 0 & 0 & -\kappa S_y \\ \kappa S_x & \kappa S_y & 0 \end{array} \right]. \tag{49.28}$$

The parameterized skew flux of potential density, ρ, due to the quasi-Stokes transport is given by

$$\boldsymbol{F}^{\mathrm{skew}} = -\boldsymbol{U}^{\mathrm{qs}} \, \partial_z \rho + \hat{\boldsymbol{z}} \, (\boldsymbol{U}^{\mathrm{qs}} \cdot \nabla_z) \rho \tag{49.29a}$$

$$= -\kappa \, \nabla_z \rho + \hat{\boldsymbol{z}} \, S^2 \, \kappa \, \partial_z \rho \tag{49.29b}$$

$$= -\kappa \, \nabla_z \rho - \hat{\boldsymbol{z}} \, (\kappa \, \rho_0 / g) \, (S \, N)^2. \tag{49.29c}$$

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 (48.100).

Figure 49.1: *Oriention 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).*

49.2.2 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 isopyncal layers, (II) the subgrid operator locally dissipates potential energy through an abiabatic 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

$$\boldsymbol{F} = -\nabla\rho \wedge \boldsymbol{\Psi}, \tag{49.30}$$

where $\boldsymbol{\Psi}$ remains to be found. To see what the local dissipation of APE imposes, consider the gravitational potential energy of the adiabatic Boussinesq system

$$P = g \int \rho\, z \,\mathrm{d}V, \tag{49.31}$$

where we assume the *in situ* density equals to the potential density as per a linear equation of state (Section 21.3.6). Assuming all boundaries are material and static allows us to focus on the time tendency of potential energy associated with the unknown flux

$$\frac{\mathrm{d}P}{\mathrm{d}t} = g \int z \frac{\partial\rho}{\partial t} \,\mathrm{d}V \tag{49.32a}$$

$$= -g \int (z\, \nabla\cdot\mathbf{F})\,\mathrm{d}V \tag{49.32b}$$

$$= -g \int (z\, \partial_z F^{(z)})\,\mathrm{d}V \tag{49.32c}$$

$$= g \int F^{(z)}\,\mathrm{d}V, \tag{49.32d}$$

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

$$F^{(z)} \leq 0, \tag{49.33}$$

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

$$F^{(z)} = \kappa\, S^2 \frac{\partial\rho}{\partial z} = -(\kappa\,\rho_0/g)\,(S\,N)^2 \leq 0, \tag{49.34}$$

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

$$\mathbf{F}^{(h)} = -\kappa\,\nabla_z\rho. \tag{49.35}$$

We have thus recovered the skew flux (49.29c) 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.

Advective 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 (49.32d) and make use of the vertical component of the advective flux rather than the skew flux

$$\frac{\mathrm{d}P}{\mathrm{d}t} = g \int F^{(z)} \, \mathrm{d}V \tag{49.36a}$$

$$= g \int \rho \, w^* \, \mathrm{d}V \tag{49.36b}$$

$$= g \int \rho \, \nabla_z \cdot (\kappa \, \boldsymbol{S}) \, \mathrm{d}V \tag{49.36c}$$

$$= g \int \nabla_z \cdot (\rho \, \kappa \, \boldsymbol{S}) \, \mathrm{d}V - g \int \nabla_z \rho \cdot \kappa \, \boldsymbol{S} \, \mathrm{d}V \tag{49.36d}$$

$$= -\rho_0 \int \kappa \, (S \, N)^2 \, \mathrm{d}V, \tag{49.36e}$$

which is the same result as for the skew flux.

49.2.3 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)

$$\mathcal{T}(y, z, t) = -\int \mathrm{d}x \int_{-H}^{z} (v + v^*) \, \mathrm{d}z \tag{49.37a}$$

$$= -\int \mathrm{d}x \int_{-H}^{z} v \, \mathrm{d}z + \int \mathrm{d}x \int_{-H}^{z} \partial_z (\kappa \, S_y) \, \mathrm{d}z \tag{49.37b}$$

$$= -\int \mathrm{d}x \int_{-H}^{z} v \, \mathrm{d}z + \int \kappa \, S_y \, \mathrm{d}x \tag{49.37c}$$

$$\equiv \mathcal{T}^{\mathrm{eulerian}}(y, z, t) + \mathcal{T}^{\mathrm{qS}}(y, z, t). \tag{49.37d}$$

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,

$$\mathcal{T}^{\mathrm{qs}} \sim L \, S \, \kappa. \tag{49.38}$$

As an example, let $\kappa = 10^3 \, \mathrm{m}^2 \, \mathrm{s}^{-1}$, $S = 10^{-3}$, and $L = 10^7$ m, which yields $\mathcal{T} \approx 10 \times 10^6 \, \mathrm{m}^3 \, \mathrm{s}^{-1} = 10 \, \mathrm{Sv}$. Such transport can represent a nontrivial addition to that from the resolved scale velocity field.

49.2.4 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 26.5 and 30.2. 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 26.5 and 30.2.

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 26.4) whereby the horizontal momentum satisfies

$$\rho_o f\,(\hat{\boldsymbol{z}} \wedge \boldsymbol{u}) = -\nabla_z p + \partial_z \boldsymbol{\tau}. \tag{49.39}$$

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 $\rho_o f\,(\hat{\boldsymbol{z}} \wedge \boldsymbol{u}^*)$ to both sides of equation (49.39) to obtain

$$\rho_o f\,(\hat{\boldsymbol{z}} \wedge \boldsymbol{u}^\dagger) = -\nabla_z p + \partial_z \boldsymbol{\tau} + \rho_o f\,(\hat{\boldsymbol{z}} \wedge \boldsymbol{u}^*), \tag{49.40}$$

where $\boldsymbol{u}^\dagger = \boldsymbol{u} + \boldsymbol{u}^*$ 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 26.4.5), whereby

$$f\,\frac{\partial \boldsymbol{u}}{\partial z} = -(g/\rho_0)\,\hat{\boldsymbol{z}} \wedge \nabla\rho = -\hat{\boldsymbol{z}} \wedge N^2 \boldsymbol{S}, \tag{49.41}$$

with

$$\boldsymbol{S} = -\frac{\nabla_z \rho}{\partial_z \rho} = \frac{g}{\rho_0}\,\frac{\nabla_z \rho}{N^2} \tag{49.42}$$

the isopycnal slope. We can thus write the Coriolis acceleration from the eddy-induced velocity in the form

$$f\,(\hat{\boldsymbol{z}} \wedge \boldsymbol{u}^*) = -f\,[\hat{\boldsymbol{z}} \wedge \partial_z(\kappa\,\boldsymbol{S})] \tag{49.43a}$$

$$= -\partial_z\,[\hat{\boldsymbol{z}} \wedge (f\,\kappa\,\boldsymbol{S})] \tag{49.43b}$$

$$= \frac{\partial}{\partial z}\left[\frac{\kappa\,f^2}{N^2}\,\frac{\partial \boldsymbol{u}}{\partial z}\right] \tag{49.43c}$$

$$= \frac{\partial}{\partial z}\left[\nu_e\,\frac{\partial \boldsymbol{u}}{\partial z}\right], \tag{49.43d}$$

where the final equality introduced an eddy-induced vertical viscosity

$$\nu_e \equiv \kappa\,(f^2/N^2). \tag{49.44}$$

Making use of this result in the planetary geostrophic equation (49.40) thus leads to

$$\rho_o f\,(\hat{\boldsymbol{z}} \wedge \boldsymbol{u}^\dagger) = -\nabla_z p + \partial_z(\boldsymbol{\tau} + \boldsymbol{\tau}_e), \tag{49.45}$$

where

$$\boldsymbol{\tau}_e = \rho_0 \, \nu_e \, \frac{\partial \boldsymbol{u}}{\partial z} \tag{49.46}$$

defines a horizontal mesoscale eddy stress arising from the thermal wind shears. Equation (49.45) 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 26.5.

49.2.5 Isopycnal thickness diffusion and GM

Recall the ensemble mean thickness equation (48.114) derived in Section 48.4.5

$$\partial_t \, h + \nabla_\rho \cdot (h \, \widehat{\boldsymbol{u}}) = 0, \tag{49.47}$$

where

$$\widehat{\boldsymbol{u}} = \boldsymbol{u} + \boldsymbol{u}^{\text{bolus}} \tag{49.48}$$

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 48.4.5.

Isopycnal correlations of horizontal velocity and layer thickness define the bolus velocity via

$$h \, \boldsymbol{u}^{\text{bolus}} = \overline{h' \, \boldsymbol{u}'} \tag{49.49}$$

Now consider a downgradient diffusive closure for this correlation

$$h \, \boldsymbol{u}^{\text{bolus}} = \overline{h' \, \boldsymbol{u}'}^{(\rho)} = -\mathbb{K}_{\text{thick}} \cdot \nabla_\rho h, \tag{49.50}$$

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

$$\partial_t \, h + \nabla_\rho \cdot (h \, \boldsymbol{u}) = \nabla_\rho \cdot (\mathbb{K}_{\text{thick}} \cdot \nabla_\rho h). \tag{49.51}$$

To make a connection between the thickness diffusion closure (49.50) and the *Gent et al.* (1995) closure discussed in Section 49.2.1, note that the specific thickness is the inverse of the vertical derivative of the potential density

$$h = (\partial_z \, \rho)^{-1}. \tag{49.52}$$

Correspondingly, using the relation between derivative operators, $\nabla_\rho = \nabla_z + \boldsymbol{S} \, \partial_z$, gives

$$h^{-1} \, \nabla_\rho h = -h \, \nabla_\rho (1/h) \tag{49.53a}$$

$$= -(\partial_z \, \rho)^{-1} \, (\nabla_z + \boldsymbol{S} \, \partial_z) \, \partial_z \, \rho \tag{49.53b}$$

$$= -\frac{\partial_z \, (\nabla_z \rho)}{\partial_z \rho} + \frac{\partial_{zz} \, \rho \, \nabla_z \rho}{(\partial_z \rho)^2} \tag{49.53c}$$

$$= \partial_z \boldsymbol{S}. \tag{49.53d}$$

Consequently, the bolus velocity takes the form

$$\boldsymbol{u}^{\text{bolus}} = -h^{-1} \, \mathbb{K}_{\text{thick}} \cdot \nabla_\rho h = -\mathbb{K}_{\text{thick}} \cdot \partial_z \boldsymbol{S}. \tag{49.54}$$

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

$$\boldsymbol{u}^{\text{bolus}} = -\partial_z \left(\kappa \, \boldsymbol{S}\right) = \boldsymbol{u}^*, \tag{49.55}$$

where the horizontal component of the *Gent et al.* (1995) velocity $\boldsymbol{u}^*$ was identified from equation (49.27). 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 *Abernathey 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.

49.3 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 21.4, 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

$$\boldsymbol{J} \cdot \hat{\boldsymbol{\gamma}} = 0 \tag{49.56}$$

where (equation (21.42))

$$\hat{\boldsymbol{\gamma}} = \frac{-\alpha \, \nabla\Theta + \beta \, \nabla S}{|-\alpha \, \nabla\Theta + \beta \, \nabla S|}. \tag{49.57}$$

49.3.1 Redi neutral diffusion

One diffusive flux satisfying the property (49.56) is given by

$$\boldsymbol{J}_{\text{redi}} = -\rho \, A \, [\nabla C - \hat{\boldsymbol{\gamma}} \, (\hat{\boldsymbol{\gamma}} \cdot \nabla C)], \tag{49.58}$$

where $A > 0$ is the neutral diffusivity (dimensions of squared length per time). We confirm that $\boldsymbol{J}_{\text{redi}}$ is oriented down the tracer gradient by noting that

$$\boldsymbol{J}_{\text{redi}} \cdot \nabla C = -\rho \, A \, \left[|\nabla C|^2 - (\hat{\boldsymbol{\gamma}} \cdot \nabla C)^2\right] \leq 0. \tag{49.59}$$

The flux $\boldsymbol{J}_{\text{redi}}$ can be written in a downgradient flux-gradient relation

$$J^m = -\rho \, \mathbb{K}^{mn}_{\text{redi}} \, \partial_n C, \tag{49.60}$$

where the Redi diffusion tensor is given by

$$\mathbb{K}^{mn}_{\text{redi}} = \frac{A}{1 + S_x^2 + S_y^2} \begin{bmatrix} 1 + S_y^2 & -S_x S_y & S_x \\ -S_x S_y & 1 + S_x^2 & S_y \\ S_x & S_y & S_x^2 + S_y^2 \end{bmatrix} \tag{49.61}$$

with the corresponding neutral diffusion operator given by the three-dimensional flux convergence

$$\mathcal{R}_{\text{redi}} = -\nabla \cdot \boldsymbol{J}_{\text{redi}} = \partial_m \left(\rho \, \mathbb{K}^{mn}_{\text{Redi}} \, \partial_n C \right). \tag{49.62}$$

In the Redi tensor (49.61) we introduced the components of the horizontal vector, $\boldsymbol{S} = (S_x, S_y, 0)$, with

$$\boldsymbol{S} = -\left[\frac{-\alpha \, \nabla_z \Theta + \beta \, \nabla_z S}{-\alpha \, \partial_z \Theta + \beta \, \partial_z S} \right] = \frac{g \left(-\alpha \, \nabla_z \Theta + \beta \, \nabla_z S \right)}{N^2} \tag{49.63}$$

the slope of the neutral tangent plane relative to the (x, y) horizontal plane, with

$$N^2 = -g \left[-\alpha \, \partial_z \Theta + \beta \, \partial_z S \right] \tag{49.64}$$

the squared buoyancy frequency (Section 21.4). Notably, it is useful to introduce the slope vector only when the fluid is stratified in the vertical so that the slope magnitude, $|\boldsymbol{S}|$, is finite.

49.3.2 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

$$\mathbb{K}^{mn}_{\text{small}} = A \begin{bmatrix} 1 & 0 & S_x \\ 0 & 1 & S_y \\ S_x & S_y & S_x^2 + S_y^2 \end{bmatrix}. \tag{49.65}$$

The corresponding small slope neutral diffusive flux is

$$\boldsymbol{J}_{\text{small}} = -\rho \, A \left[\nabla_\gamma + \hat{\boldsymbol{z}} \left(\boldsymbol{S} \cdot \nabla_\gamma \right) \right] C \tag{49.66}$$

where

$$\nabla_\gamma = \nabla_z + \boldsymbol{S} \, \partial_z \tag{49.67}$$

is the horizontal derivative operator computed on the neutral tangent plane (see equation (8.61)). To show that $\boldsymbol{J}_{\text{small}} \cdot \hat{\boldsymbol{\gamma}} = 0$, we make use of the identity

$$\hat{\boldsymbol{\gamma}} = \frac{\boldsymbol{S} - \hat{\boldsymbol{z}}}{(1 + \boldsymbol{S} \cdot \boldsymbol{S})^{1/2}}, \tag{49.68}$$

so that

$$\boldsymbol{J}_{\text{small}} \cdot \hat{\boldsymbol{\gamma}} = \frac{\boldsymbol{J}_{\text{small}} \cdot \boldsymbol{S} - \boldsymbol{J}_{\text{small}} \cdot \boldsymbol{S}}{(1 + \boldsymbol{S} \cdot \boldsymbol{S})^{1/2}} = 0. \tag{49.69}$$

Furthermore, we confirm that $\boldsymbol{J}_{\text{small}}$ is oriented down the tracer gradient by noting that

$$\begin{aligned} \boldsymbol{J}_{\text{small}} \cdot \nabla C &= -\rho \, A \left[\nabla_\gamma C \cdot \nabla_z C + \left(\boldsymbol{S} \cdot \nabla_\gamma C \right) \partial_z C \right] && \text{(49.70a)} \\ &= -\rho \, A \left[|\nabla_z C|^2 + 2 \left(\boldsymbol{S} \cdot \nabla_z C \right) \partial_z C + |\boldsymbol{S} \, \partial_z C|^2 \right] && \text{(49.70b)} \\ &= -\rho \, A \, |\nabla_z C + \boldsymbol{S} \, \partial_z C|^2 && \text{(49.70c)} \\ &= -\rho \, A \, |\nabla_\gamma C|^2 && \text{(49.70d)} \\ &\leq 0. && \text{(49.70e)} \end{aligned}$$

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 $\boldsymbol{J}_{\text{small}} \cdot \hat{\boldsymbol{\gamma}} = 0$. The corrected form given by equation (49.66) 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

$$\mathcal{R}_{\text{small}} = -\nabla \cdot \boldsymbol{J}_{\text{small}} = \nabla_z \cdot (\rho\, A\, \nabla_\gamma C) + \partial_z \left(\rho\, A\, \boldsymbol{S} \cdot \nabla_\gamma C\right). \tag{49.71}$$

49.3.3 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 8.

Following the derivations given in Section 8.14, the neutral diffusive flux in the neutral tangent frame is given by the horizontal flux

$$\boldsymbol{J}_{\text{ntp}} = -\rho\, A\, \nabla_\gamma C. \tag{49.72}$$

This flux is oriented down the tracer gradient as oriented along neutral directions

$$\boldsymbol{J}_{\text{ntp}} \cdot \nabla_\gamma C = -\rho\, A\, |\nabla_\gamma C|^2, \tag{49.73}$$

which is the same as equation (49.70d) for the small slope fluxes. However, as a purely horizontal flux, $\boldsymbol{J}_{\text{ntp}}$ is not oriented along neutral directions

$$\boldsymbol{J}_{\text{ntp}} \cdot \hat{\boldsymbol{\gamma}} \neq 0. \tag{49.74}$$

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 $\boldsymbol{J}_{\text{ntp}}$ along the neutral tangent plane as per equation (8.81)

$$\mathcal{R}_{\text{ntp}} = -\frac{1}{h^\gamma}\left[\nabla_\gamma \cdot (h^\gamma\, \boldsymbol{J}_{\text{ntp}})\right] = \frac{1}{h^\gamma}\left[\nabla_\gamma \cdot (h^\gamma\, \rho\, A\, \nabla_\gamma C)\right], \tag{49.75}$$

where

$$h^\gamma = \frac{\partial z}{\partial \gamma}\, \mathrm{d}\gamma = -\left[\frac{g}{\rho_0\, N^2}\right] \mathrm{d}\gamma \tag{49.76}$$

measures the thickness of a layer defined by two neutral tangent planes (see equation (8.78)).

As detailed in Section 8.14, $\mathcal{R}_{\text{ntp}}$ is identical to the small slope neutral diffusion operator (49.71)

$$\mathcal{R}_{\text{ntp}} = \mathcal{R}_{\text{small}}. \tag{49.77}$$

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 $\mathcal{R}_{\text{ntp}}$ 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 47.3). 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.

49.3.4 Neutrality condition

Given the expression (49.57) for the dianeutral unit vector, $\hat{\boldsymbol{\gamma}}$, it is straightforward to show that the neutral diffusive flux for Conservative Temperature balances that for salinity

$$\alpha\, \boldsymbol{J}(\Theta) = \beta\, \boldsymbol{J}(S). \tag{49.78}$$

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 (49.58) of *Redi* (1982), the small slope flux (49.66) of *Gent and McWilliams* (1990), and the neutral tangent frame neutral diffusive flux (49.72). 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).

49.3.5 Symmetry condition

Since the neutral diffusion tensor is symmetric (as are all diffusion tensors; see Section 47.3), we have

$$\boldsymbol{J}(\Theta) \cdot \nabla S = -A\, \rho\, \mathbb{K}^{mn} \partial_n \Theta\, \partial_m S \tag{49.79a}$$
$$= -A\, \rho\, \mathbb{K}^{nm}\, \partial_n S\, \partial_m \Theta \tag{49.79b}$$
$$= -A\, \rho\, \mathbb{K}^{nm}\, \partial_n S\, \partial_m \Theta \tag{49.79c}$$
$$= \boldsymbol{J}(S) \cdot \nabla \Theta. \tag{49.79d}$$

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 49.4.

49.3.6 GM skewsion 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 skewsion (Section 49.2.1) and small slope neutral diffusion (Section 49.3.2). The combined tracer flux takes the form

$$\boldsymbol{F} = -A\, \nabla_z C + (\kappa - A)\, \mathbf{S}\, \partial_z C - \hat{\mathbf{z}}\, [(A + \kappa)\, \mathbf{S} \cdot \nabla_z C + A\, S^2\, \partial_z C], \tag{49.80}$$

which can be written in terms of a subgrid scale transport tensor (*Griffies*, 1998)

$$\begin{bmatrix} F^{(x)} \\ F^{(y)} \\ F^{(z)} \end{bmatrix} = \begin{bmatrix} A & 0 & (A-\kappa)\, S_x \\ 0 & A & (A-\kappa)\, S_y \\ (A+\kappa)\, S_x & (A+\kappa)\, S_y & A\, S^2 \end{bmatrix} \begin{bmatrix} \partial_x C \\ \partial_y C \\ \partial_z C \end{bmatrix}. \tag{49.81}$$

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

$$\mathbf{F} = -\kappa\, \nabla_z C - \hat{\mathbf{z}}\, \kappa\, (2\, \mathbf{S} \cdot \nabla_z C + S^2\, C_{,z}). \tag{49.82}$$

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 skewsion are distinct, thus breaking their symmetry. Furthermore, studies such as *Smith and Marshall* (2009) and *Abernathey et al.* (2013) clearly point to the distinct vertical structure for the two diffusivities. Such distinctions 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.

49.3.7 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 8 and 17. 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 (8.80) for a general diffusion operator written in terms of the generalized vertical coordinate, $\sigma = \sigma(x, y, z, t)$

$$\mathcal{R} = -\frac{1}{h^\sigma}\left[\nabla_\sigma \cdot (h^\sigma \, \boldsymbol{J}^h) + \delta_\sigma(z_\sigma \, \nabla\sigma \cdot \boldsymbol{J})\right], \tag{49.83}$$

where $\delta_\sigma \equiv \mathrm{d}\sigma \, \partial_\sigma$ is the dimensionless derivative operator, and the thickness of a σ-layer is

$$h^\sigma = \mathrm{d}z = z_\sigma \, \mathrm{d}\sigma = \frac{\partial z}{\partial \sigma} \, \mathrm{d}\sigma. \tag{49.84}$$

Now assume the flux, $\boldsymbol{J}$, is given by equation (49.66) for small slope neutral diffusion. Transforming to GVCs leads to the horizontal flux component

$$\begin{aligned}
\boldsymbol{J}^h_{\text{small}} &= -\rho \, A \, \nabla_\gamma C && \text{(49.85a)}\\
&= -\rho \, A \, [\nabla_z + (\nabla_\gamma z) \, \partial_z] \, C && \text{(49.85b)}\\
&= -\rho \, A \, [\nabla_\sigma + (-\nabla_\sigma z + \nabla_\gamma z) \, \partial_z] \, C && \text{(49.85c)}\\
&= -\rho \, A \, [\nabla_\sigma + (-\boldsymbol{S}^{(\sigma/z)} + \boldsymbol{S}^{(\gamma/z)}) \, \partial_z] \, C && \text{(49.85d)}\\
&= -\rho \, A \, (\nabla_\sigma + \boldsymbol{S}^{(\gamma/\sigma)} \, \partial_z) \, C, && \text{(49.85e)}
\end{aligned}$$

where the neutral slopes as shown in Figure 49.2 satisfy the identity

$$\boldsymbol{S}^{(\sigma/z)} = \boldsymbol{S}^{(\gamma/z)} - \boldsymbol{S}^{(\gamma/\sigma)}. \tag{49.86}$$

Furthermore, we made use of the identity (8.61) relating the partial derivative operators

$$\nabla_\gamma = \nabla_z + (\nabla_\gamma z) \, \partial_z \qquad \nabla_z = \nabla_\sigma - (\nabla_\sigma z) \, \partial_z. \tag{49.87}$$

The horizontal flux (49.85e) has the same form as when written using geopotential coordinates, only now with the derivative operator ∇_σ and the slope $\boldsymbol{S}^{(\gamma/\sigma)}$. Correspondingly, the vertical flux component

$$J^z_{\text{small}} = \boldsymbol{J}^h_{\text{small}} \cdot \boldsymbol{S}^{(\gamma/z)} \tag{49.88}$$

takes the form

$$z_\sigma \, \nabla\sigma \cdot \boldsymbol{J}_{\text{small}} = -\boldsymbol{S}^{(\sigma/z)} \cdot \boldsymbol{J}^h_{\text{small}} + J^z_{\text{small}} = \boldsymbol{J}^h_{\text{small}} \cdot \boldsymbol{S}^{(\gamma/\sigma)}, \tag{49.89}$$

which in turn yields the diffusion operator (49.83)

$$\mathcal{R} = -\frac{1}{h^\sigma}\left[\nabla_\sigma \cdot (h^\sigma \, \boldsymbol{J}^h_{\text{small}}) + \delta_\sigma(\boldsymbol{J}^h_{\text{small}} \cdot \boldsymbol{S}^{(\gamma/\sigma)})\right]. \tag{49.90}$$

In the special case when σ is parallel to the neutral direction so that $\boldsymbol{S}^{(\gamma/\sigma)} = 0$, the diffusion operator (49.90) reduces to the neutral tangent plane version given by equation (49.75).

Figure 49.2: *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 $\boldsymbol{S}^{(\sigma/z)} = \boldsymbol{S}^{(\gamma/z)} - \boldsymbol{S}^{(\gamma/\sigma)}$, where $|\boldsymbol{S}^{(\gamma/z)}| = |\tan\theta^{(\gamma/z)}|$ and $|\boldsymbol{S}^{(\gamma/\sigma)}| = |\tan\theta^{(\gamma/\sigma)}|$. Note that this relation between slope vectors also holds for arbitrary orientations of the σ and γ isolines.*

49.3.8 Comments

As noted in Section 48.6.3, 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 48.3.2 and *Middleton and Loder* (1989).

49.4 Cabbeling and thermobaricity

We now return to the density equation (50.30)

$$\frac{\mathrm{D}\ln\rho}{\mathrm{D}t} = \nabla \cdot [\nu_\Theta \, \boldsymbol{J}(\Theta) + \nu_S \, \boldsymbol{J}(S)] - (\boldsymbol{J}(\Theta) \cdot \nabla\nu_\Theta + \boldsymbol{J}(S) \cdot \nabla\nu_S) + \frac{\omega}{\rho \, c^2}. \tag{49.91}$$

We focus here on temperature and salinity fluxes just from neutral diffusion. The neutrality condition (49.78) takes the following form in terms of specific volume

$$\nu_\Theta \, \boldsymbol{J}(\Theta) + \nu_S \, \boldsymbol{J}(S) = 0. \tag{49.92}$$

Consequently, neutral diffusion affects density evolution only through the source term

$$\left[\frac{\mathrm{D}\ln\rho}{\mathrm{D}t}\right]_{\text{ntrl diff}} = -\boldsymbol{J}(\Theta) \cdot \nabla\nu_\Theta - \boldsymbol{J}(S) \cdot \nabla\nu_S. \tag{49.93}$$

49.4.1 Basic manipulations

As a first step, eliminate the salt flux by using the neutrality condition (49.92) so that

$$\boldsymbol{J}(\Theta)\cdot\nabla\nu_\Theta + \boldsymbol{J}(S)\cdot\nabla\nu_S = \boldsymbol{J}(\Theta)\cdot[\nu_S\,\nabla\nu_\Theta - \nu_\Theta\,\nabla\nu_S]/\nu_S. \tag{49.94}$$

Next, expand the gradients of the specific volume to write

$$\nabla\nu_\Theta = \nu_{\Theta\Theta}\,\nabla\Theta + \nu_{\Theta S}\,\nabla S + \nu_{\Theta p}\,\nabla p \tag{49.95a}$$

$$\nabla\nu_S = \nu_{SS}\,\nabla S + \nu_{\Theta S}\,\nabla\Theta + \nu_{Sp}\,\nabla p, \tag{49.95b}$$

so that

$$\nu_S\,\nabla\nu_\Theta - \nu_\Theta\,\nabla\nu_S = \nabla\Theta\,(\nu_S\,\nu_{\Theta\Theta} - \nu_\Theta\,\nu_{\Theta S}) + \nabla S\,(\nu_S\,\nu_{\Theta S} - \nu_\Theta\,\nu_{SS}) + \nabla p\,(\nu_S\,\nu_{\Theta p} - \nu_\Theta\,\nu_{Sp}). \tag{49.96}$$

We again make use of the neutrality condition (49.92), as well as the symmetry condition (49.79d) to write

$$\boldsymbol{J}(\Theta)\cdot\nabla S\,(\nu_S\,\nu_{\Theta S} - \nu_\Theta\,\nu_{SS}) = -\boldsymbol{J}(\Theta)\cdot\nabla\Theta\left(\nu_\Theta\,\nu_{\Theta S} - \nu_{SS}\,\frac{(\nu_\Theta)^2}{\nu_S}\right). \tag{49.97}$$

Bringing these results together leads to

$$\boldsymbol{J}(\Theta)\cdot\nabla\nu_\Theta + \boldsymbol{J}(S)\cdot\nabla\nu_S = \boldsymbol{J}(\Theta)\cdot\nabla p\left[\nu_{\Theta p} - \nu_{pS}\,\frac{\nu_\Theta}{\nu_S}\right] + \boldsymbol{J}(\Theta)\cdot\nabla\Theta\left[\nu_{\Theta\Theta} - 2\,\nu_{\Theta S}\,\frac{\nu_\Theta}{\nu_S} + \nu_{SS}\left(\frac{\nu_\Theta}{\nu_S}\right)^2\right], \tag{49.98}$$

which can be written in terms of density partial derivatives as

$$\boldsymbol{J}(\Theta)\cdot\nabla\nu_\Theta + \boldsymbol{J}(S)\cdot\nabla\nu_S = -\rho^{-2}\,\boldsymbol{J}(\Theta)\cdot\nabla p\left[\rho_{\Theta p} - \rho_{pS}\,\frac{\rho_\Theta}{\rho_S}\right] - \rho^{-2}\,\boldsymbol{J}(\Theta)\cdot\nabla\Theta\left[\rho_{\Theta\Theta} - 2\,\rho_{\Theta S}\,\frac{\rho_\Theta}{\rho_S} + \rho_{SS}\left(\frac{\rho_\Theta}{\rho_S}\right)^2\right]. \tag{49.99}$$

49.4.2 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

$$\mathcal{T} = \beta\,\partial_p\left[\frac{\alpha}{\beta}\right] \tag{49.100a}$$

$$= \frac{\partial\alpha}{\partial p} - \frac{\alpha}{\beta}\,\frac{\partial\beta}{\partial p} \tag{49.100b}$$

$$= \rho\,\nu_S\,\partial_p\left[\frac{\nu_\Theta}{\nu_S}\right] \tag{49.100c}$$

$$= -\rho^{-1}\,\rho_S\,\partial_p\left[\frac{\rho_\Theta}{\rho_S}\right] \tag{49.100d}$$

$$= -\rho^{-1}\left[\rho_{\Theta p} - \rho_{pS}\left[\frac{\rho_\Theta}{\rho_S}\right]\right], \tag{49.100e}$$

and the *cabbeling* parameter (dimensions of squared inverse temperature)

$$\mathcal{C} = \frac{\partial \alpha}{\partial \Theta} + 2\frac{\alpha}{\beta}\frac{\partial \alpha}{\partial S} - \left(\frac{\alpha}{\beta}\right)^2 \frac{\partial \beta}{\partial S} \tag{49.101a}$$

$$= -\rho^{-1}\left[\rho_{\Theta\Theta} - 2\,\rho_{\Theta S}\left[\frac{\rho_\Theta}{\rho_S}\right] + \rho_{SS}\left[\frac{\rho_\Theta}{\rho_S}\right]^2\right] \tag{49.101b}$$

$$= \rho\left[\nu_{\Theta\Theta} - 2\,\nu_{\Theta S}\left[\frac{\nu_\Theta}{\nu_S}\right] + \nu_{SS}\left[\frac{\nu_\Theta}{\nu_S}\right]^2\right] \tag{49.101c}$$

to render the very compact result

$$\boldsymbol{J}(\Theta)\cdot\nabla\nu_\Theta + \boldsymbol{J}(S)\cdot\nabla\nu_S = \rho^{-1}\,\boldsymbol{J}(\Theta)\cdot(\mathcal{T}\,\nabla p + \mathcal{C}\,\nabla\Theta) \tag{49.102}$$

which in turn yields the material evolution of *in situ* density due to neutral diffusion

$$\left[\frac{\mathrm{D}\rho}{\mathrm{D}t}\right]_{\text{ntrl diff}} = -\boldsymbol{J}(\Theta)\cdot(\mathcal{T}\,\nabla p + \mathcal{C}\,\nabla\Theta). \tag{49.103}$$

49.4.3 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 elementl 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

$$\mathcal{C} = \frac{\partial \alpha}{\partial \Theta} + 2\frac{\alpha}{\beta}\frac{\partial \alpha}{\partial S} - \left[\frac{\alpha}{\beta}\right]^2 \frac{\partial \beta}{\partial S} \geq 0. \tag{49.104}$$

Given the downgradient nature of the neutral diffusive fluxes, we have

$$\text{Cabbeling} \equiv -\mathcal{C}\,\boldsymbol{J}(\Theta)\cdot\nabla\Theta \geq 0, \tag{49.105}$$

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.

49.4.4 Thermobaricity

The thermobaricity parameter

$$\mathcal{T} = \beta\,\frac{\partial}{\partial p}\left[\frac{\alpha}{\beta}\right] \tag{49.106}$$

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

$$\begin{aligned} \mathcal{T} &= \frac{\partial \alpha}{\partial p} - \frac{\alpha}{\beta}\frac{\partial \beta}{\partial p} \\ &= -\frac{\rho_{\Theta p}}{\rho} + \frac{\rho_\Theta}{\rho_S}\frac{\rho_{pS}}{\rho} \end{aligned} \tag{49.107}$$

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)

$$\text{Thermobaricity} \equiv -\mathcal{T}\, \boldsymbol{J}(\Theta) \cdot \nabla p \tag{49.108}$$

appearing in equation (49.102). 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.

49.4.5 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 densitification, 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 49.1). Namely, cabbeling and thermobarocity 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.

50

Evolution of ocean buoyancy and global 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 establish the evolution equations for Θ and S and in turn determine their affects on 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 occuring 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 we measure global mean sea level, with rudiments of this analysis a topic for this chapter.

READER'S GUIDE FOR THIS CHAPTER

Development in Section 50.1 presumes an understanding of how the tracer equation is derived from the kinematics of mass conserving fluid elements detailed in Section 15.1. Basic notions of thermodynamics, such as Section 23.5, 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 buouyancy fluxes and analyze changes to global mean sea level. Both of these topics are increasingly important in our anthropogenically warming world.

50.1 Salt and freshwater budgets **900**
50.1.1 Mass budgets 901
50.1.2 Kinematic boundary conditons 902
50.1.3 Further reading 903
50.2 Material evolution of *in situ* density **903**
50.2.1 Material changes to pressure 903
50.2.2 Material changes to Θ and S 904
50.2.3 Summary of density changes 904
50.2.4 A synopsis of ocean mixing processes 905
50.3 Boundary fluxes of buoyancy **905**
50.3.1 Outlining the boundary fluxes of enthalpy and salt 906
50.3.2 Transport of scalars across the ocean surface 906

50.3.3 Scalar budgets for a surface ocean model grid cell 907
50.3.4 Salt fluxes from sea ice melt and formation 908
50.3.5 Salt and enthalpy fluxes associated with water transport 908
50.3.6 Non-penetrative surface enthalpy fluxes 908
50.3.7 The case of frazil 909
50.3.8 Penetrative shortwave radiation 910
50.3.9 Buoyancy budget for a surface ocean model grid cell 910
50.3.10 Comments 911
50.4 Global mean sea level **911**
50.4.1 Definitions and assumptions 911
50.4.2 Budget for global mean sea level 912
50.4.3 Changes due to mass input 913
50.4.4 Steric changes due to changes in density 913
50.4.5 Enthalpy flux imbalances giving rise to thermosteric sea level 914
50.4.6 Global sea level in a Boussinesq fluid 914
50.4.7 Global halosteric sea level changes are negligible 915
50.4.8 Further study 915

50.1 Salt and freshwater budgets

Seawater is comprised of freshwater 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 of a single effective mass concentration known as the *salt concentration*

$$C^{(\mathrm{s})} = \frac{\text{mass of salt}}{\text{mass of seawater}} = \frac{\text{mass of salt}}{\text{mass of freshwater} + \text{mass of salt}} \tag{50.1}$$

to specify the amount of salt within an element of seawater. In practice oceanographers often prefer to work with *salinity*,

$$S = 1000\, C^{(\mathrm{s})}, \tag{50.2}$$

which converts from typical salt concentrations of $C^{(\mathrm{s})} = 0.035$ to a salinity of $S = 35$. For brevity in the remainder of this section, we write

$$C = C^{(\mathrm{s})}. \tag{50.3}$$

The complement to salt concentration is the freshwater concentration or mass fraction for an element of seawater

$$F = \frac{\text{mass of freshwater}}{\text{mass of seawater}} = \frac{\text{mass of freshwater}}{\text{mass of freshwater} + \text{mass of salt}} = 1 - C. \tag{50.4}$$

Other trace matter occurs at very low concentrations so as to make seawater a two-component fluid consisting of freshwater plus dissolved salt. We here derive the mass budget for salt and freshwater as well as the associated kinematic boundary conditions.

50.1.1 Mass budgets

Following our discussion of the tracer equation in Section 15.1, the mass budget equations for an element of seawater take the form

$$\frac{\partial \rho}{\partial t} + \nabla \cdot (\rho \, \boldsymbol{v}) = 0 \qquad \text{seawater} \tag{50.5}$$

$$\frac{\partial (\rho \, C)}{\partial t} + \nabla \cdot (\rho \, \boldsymbol{v} \, C + \boldsymbol{J}^{(\mathrm{C})}) = 0 \qquad \text{salt} \tag{50.6}$$

$$\frac{\partial (\rho \, F)}{\partial t} + \nabla \cdot (\rho \, \boldsymbol{v} \, F + \boldsymbol{J}^{(\mathrm{F})}) = 0 \qquad \text{freshwater,} \tag{50.7}$$

where ρ is the seawater mass density. Equation (50.5) is the mass budget for seawater and equation (50.6) is the mass budget for salt. The freshwater budget (50.7) is derived by subtracting the salt budget (50.6) from the seawater mass budget (50.5). Hence, only two of the three mass budget equations (50.5)-(50.7) are independent.

We make use of the barycentric velocity in the above conservation laws, where the barycentric velocity is given by

$$\boldsymbol{v} = C \, \boldsymbol{v}^{(\mathrm{C})} + F \, \boldsymbol{v}^{(\mathrm{F})}. \tag{50.8}$$

The velocities $\boldsymbol{v}^{(\mathrm{C})}$ and $\boldsymbol{v}^{(\mathrm{F})}$ are, respectively, the molecular center of mass velocities for salt and freshwater contained in a fluid element, in which case

$$\rho \left[\frac{\partial C}{\partial t} + \boldsymbol{v}^{(\mathrm{C})} \cdot \nabla C \right] = \frac{\partial (\rho \, C)}{\partial t} + \nabla \cdot (\rho \, \boldsymbol{v}^{(\mathrm{C})} \, C) = 0 \tag{50.9a}$$

$$\rho \left[\frac{\partial F}{\partial t} + \boldsymbol{v}^{(\mathrm{F})} \cdot \nabla F \right] = \frac{\partial (\rho \, F)}{\partial t} + \nabla \cdot (\rho \, \boldsymbol{v}^{(\mathrm{F})} \, F) = 0. \tag{50.9b}$$

Furthermore, the fluxes $\boldsymbol{J}^{(\mathrm{C})}$ and $\boldsymbol{J}^{(\mathrm{F})}$ arise from the difference between the salt and freshwater velocities from the barycentric velocity

$$\boldsymbol{J}^{(\mathrm{C})} = \rho \, C \, (\boldsymbol{v}^{(\mathrm{C})} - \boldsymbol{v}) \qquad \text{and} \qquad \boldsymbol{J}^{(\mathrm{F})} = \rho \, F \, (\boldsymbol{v}^{(\mathrm{F})} - \boldsymbol{v}). \tag{50.10}$$

These fluxes are generally parameterized by downgradient diffusive fluxes

$$\boldsymbol{J}^{(\mathrm{C})} = -\rho \, \kappa_{\mathrm{C}} \, \nabla C \qquad \text{and} \qquad \boldsymbol{J}^{(\mathrm{F})} = -\rho \, \kappa_{\mathrm{C}} \, \nabla F, \tag{50.11}$$

where $\kappa_{\mathrm{C}} > 0$ is the kinematic diffusivity for salt in seawater (*Gill*, 1982). We use the same diffusivity for salt and freshwater, as the diffusion of one is balanced by the other. As discussed in Section 49.1, the effective diffusivity is enhanced beyond the molecular value in the presence of subgrid scale eddy effects.

The advective flux of seawater is comprised of a salt flux plus a freshwater flux

$$\rho \, \boldsymbol{v} = \rho \, C \, \boldsymbol{v}^{(\mathrm{C})} + \rho \, F \, \boldsymbol{v}^{(\mathrm{F})}. \tag{50.12}$$

Conversely, the salt flux and freshwater flux can be represented as a diffusive flux plus an advective flux where advection is via the barycentric velocity

$$\rho \, C \, \boldsymbol{v}^{(\mathrm{C})} = \rho \, C \, (\boldsymbol{v}^{(\mathrm{C})} - \boldsymbol{v}) + \rho \, C \, \boldsymbol{v} = \boldsymbol{J}^{(\mathrm{C})} + \rho \, C \, \boldsymbol{v} \tag{50.13a}$$

$$\rho \, F \, \boldsymbol{v}^{(\mathrm{F})} = \rho \, F \, (\boldsymbol{v}^{(\mathrm{F})} - \boldsymbol{v}) + \rho \, F \, \boldsymbol{v} = \boldsymbol{J}^{(\mathrm{F})} + \rho \, F \, \boldsymbol{v}. \tag{50.13b}$$

The center of mass velocities offer a conceptual framework of use for a formulation of kinematic boundary conditions in Section 50.1.2. Even so, they offer no new information beyond the parameterized fluxes $\boldsymbol{J}^{(\mathrm{C})}$ and $\boldsymbol{J}^{(\mathrm{F})}$: knowledge of one is sufficient for determining the other.

50.1.2 Kinematic boundary conditons

Recall the boundary condition (14.60) derived in Section 14.4.3 for matter crossing the ocean surface

$$\rho\,(\boldsymbol{v} - \boldsymbol{v}^{(\eta)}) \cdot \hat{\boldsymbol{n}} = -\mathcal{Q}_{\mathrm{m}}, \tag{50.14}$$

where $\mathcal{Q}_{\mathrm{m}}$ is the net mass flux of freshwater plus salt crossing the surface boundary, and $\boldsymbol{v}^{(\eta)}$ is the velocity of a point on the free surface.

Boundary conditions in terms of salt and freshwater velocities

In deriving the boundary condition (50.14), we made use of the barycentric velocity $\boldsymbol{v}$ for an element of seawater. A directly analogous procedure can be applied to the salt and freshwater crossing the surface to render

$$\rho\,(\boldsymbol{v}^{(\mathrm{C})} - \boldsymbol{v}^{(\eta)}) \cdot \hat{\boldsymbol{n}} = -\mathcal{Q}_{\mathrm{C}} \tag{50.15a}$$

$$\rho\,(\boldsymbol{v}^{(\mathrm{F})} - \boldsymbol{v}^{(\eta)}) \cdot \hat{\boldsymbol{n}} = -\mathcal{Q}_{\mathrm{F}}, \tag{50.15b}$$

where

$$\mathcal{Q}_{\mathrm{m}} = \mathcal{Q}_{\mathrm{C}} + \mathcal{Q}_{\mathrm{F}} \tag{50.16}$$

relates the mass fluxes of salt and freshwater to the total mass flux crossing the boundary (mass per time per surface normalized area). Adding the boundary conditions (50.15a) and (50.15b), and using the relation $C + F = 1$, recovers the boundary condition (50.14) written in terms of the barycentric velocity.

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

$$\rho\,(\boldsymbol{v}^{(\mathrm{C})} - \boldsymbol{v}^{(\eta)}) \cdot \hat{\boldsymbol{n}} = 0 \qquad \text{zero surface salt flux.} \tag{50.17}$$

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.

Diffusive flux boundary condition for salt

We find it useful to make use of relation (50.13a) to eliminate the salt velocity $\boldsymbol{v}^{(\mathrm{C})}$ in favor of the diffusive flux $\boldsymbol{J}^{(\mathrm{C})} = \rho\, C\,(\boldsymbol{v}^{(\mathrm{C})} - \boldsymbol{v})$, in which case the kinematic boundary condition (50.15a) takes the form

$$\boldsymbol{J}^{(\mathrm{C})} \cdot \hat{\boldsymbol{n}} = C\,\mathcal{Q}_{\mathrm{F}}. \tag{50.18}$$

This kinematic boundary condition relates the surface freshwater flux crossing the ocean surface (right hand side) to the normal component of the subgrid scale salt flux (left hand side). Notably, this result holds whether or not there is a surface salt flux.

Now assume $\boldsymbol{J}^{(\mathrm{C})}$ is in the form of a diffusive flux

$$\boldsymbol{J}^{(\mathrm{C})} = -\rho\,\kappa_{\mathrm{C}}\,\nabla C, \tag{50.19}$$

in which case the boundary condition (50.18) takes the form

$$\rho\,\kappa_{\mathrm{C}}\,\nabla C \cdot \hat{\boldsymbol{n}} = -C\,\mathcal{Q}_{\mathrm{F}}. \tag{50.20}$$

Diffusive mixing of salinity within the ocean thus mediates the incorporation of boundary freshwater fluxes into the ocean. Since it is the mass of a fluid element that is constant, any transfer of

freshwater into that element must be compensated by a removal of salt, and vice versa. Through the act of salt diffusion in one direction, freshwater diffuses in the opposite. For example, when adding freshwater to the ocean, $P - E > 0$, it enters the ocean (moves downward) so long as salt diffuses upward toward the surface. Correspondingly, in the absence of diffusive mixing, boundary freshwater is not incorporated into the ambient ocean fluid. Rather, it remains an unmixed lens sitting on top of the seawater.

50.1.3 Further reading

Nurser and Griffies (2019) offer a more detailed discussion of this material.

50.2 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 21.3.2, 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

$$\rho = \rho(\Theta, S, p). \tag{50.21}$$

We formulate the material evolution of density as weighted by the specific volume

$$\nu = \rho^{-1}, \tag{50.22}$$

so that we consider

$$\frac{\mathrm{D}\ln\rho}{\mathrm{D}t} = \frac{\partial\ln\rho}{\partial\Theta}\frac{\mathrm{D}\Theta}{\mathrm{D}t} + \frac{\partial\ln\rho}{\partial S}\frac{\mathrm{D}S}{\mathrm{D}t} + \frac{\partial\ln\rho}{\partial p}\frac{\mathrm{D}p}{\mathrm{D}t} \tag{50.23a}$$

$$= -\alpha_\Theta\frac{\mathrm{D}\Theta}{\mathrm{D}t} + \beta_\mathrm{S}\frac{\mathrm{D}S}{\mathrm{D}t} + \frac{\dot{p}}{\rho\, c_\mathrm{s}^2}. \tag{50.23b}$$

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

$$\alpha_\Theta = -\left[\frac{\partial\ln\rho}{\partial\Theta}\right]_{p,S} \qquad \beta_\mathrm{S} = \left[\frac{\partial\ln\rho}{\partial S}\right]_{p,\Theta} \qquad c_\mathrm{s}^2 = \left[\frac{\partial p}{\partial\rho}\right]_{S,\Theta} \qquad \dot{p} = \frac{\mathrm{D}p}{\mathrm{D}t}. \tag{50.24}$$

50.2.1 Material changes to pressure

To garner some exposure to the physics of $\dot{p}$ as it appears in equation (50.23), 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 17.3.6)

$$\begin{aligned} w^{(p)} &= \frac{\partial z}{\partial p}\frac{\mathrm{D}p}{\mathrm{D}t} \\ &= -(\rho\, g)^{-1}\,\dot{p}. \end{aligned} \tag{50.25}$$

The transport measured by $w^{(p)}$ is the pressure-coordinate analog of the vertical velocity component $w = \mathrm{D}z/\mathrm{D}t$ 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.

50.2.2 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

$$\rho \frac{\mathrm{D}\Theta}{\mathrm{D}t} = -\nabla \cdot \boldsymbol{J}(\Theta) \tag{50.26a}$$

$$\rho \frac{\mathrm{D}S}{\mathrm{D}t} = -\nabla \cdot \boldsymbol{J}(S). \tag{50.26b}$$

This form for material changes in temperature and salinity then lead to

$$-\alpha_\Theta \frac{\mathrm{D}\Theta}{\mathrm{D}t} + \beta_\mathrm{s} \frac{\mathrm{D}S}{\mathrm{D}t} = \nu_\Theta \, \nabla \cdot \boldsymbol{J}(\Theta) + \nu_S \, \nabla \cdot \boldsymbol{J}(S) \tag{50.27a}$$

$$= \nabla \cdot [\nu_\Theta \, \boldsymbol{J}(\Theta) + \nu_S \, \boldsymbol{J}(S)] - [\boldsymbol{J}(\Theta) \cdot \nabla \nu_\Theta + \boldsymbol{J}(S) \cdot \nabla \nu_S] \tag{50.27b}$$

where again $\nu = \rho^{-1}$ is the specific volume and its partial derivatives are

$$\nu_\Theta = \frac{\partial \nu}{\partial \Theta} = \frac{\alpha_\Theta}{\rho} \qquad \text{and} \qquad \nu_S = \frac{\partial \nu}{\partial S} = -\frac{\beta_\mathrm{s}}{\rho}. \tag{50.28}$$

50.2.3 Summary of density changes

Bringing the above results together leads to the density equation

$$\frac{\mathrm{D} \ln \rho}{\mathrm{D}t} - \frac{\dot{p}}{\rho \, c_\mathrm{s}^2} = \nabla \cdot [\nu_\Theta \, \boldsymbol{J}(\Theta) + \nu_S \, \boldsymbol{J}(S)] - [\boldsymbol{J}(\Theta) \cdot \nabla \nu_\Theta + \boldsymbol{J}(S) \cdot \nabla \nu_S] \,, \tag{50.29}$$

which has the equivalent form

$$\frac{\mathrm{D}\rho}{\mathrm{D}t} - \frac{\dot{p}}{c_\mathrm{s}^2} = \nabla \cdot [\alpha_\Theta \, \boldsymbol{J}(\Theta) - \beta_\mathrm{s} \, \boldsymbol{J}(S)] - [\boldsymbol{J}(\Theta) \cdot \nabla \alpha_\Theta - \boldsymbol{J}(S) \cdot \nabla \beta_\mathrm{s}] \,. \tag{50.30}$$

We brought the adiabatic source term from motion across pressure surfaces (Section 50.2.1) 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., temperatureΘ reducing and S increasing). These fluxes arise from a variety of mixing processes, some of which are surveyed in Section 50.2.4. The second term on the right hand side of equations (50.29) and (50.30) relates to properties of the locally referenced potential density surface. We study this source term arising from neutral diffusion in Section 49.4, 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 49.1.4).

50.2.4 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} m^2 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 $\mathcal{O}$(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 $\mathcal{O}(10^{-3} - 10^{-2}$ m^2 s$^{-1})$ in boundary layers, and $\mathcal{O}(10^{-5}$ m^2 s$^{-1})$ in the quiescent ocean interior (*Polzin et al.*, 1997; *Whalen et al.*, 2012; *Waterhouse et al.*, 2014).

Mesoscale eddies, with size $\mathcal{O}(10 - 100)$ 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 49.3). 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 $\mathcal{O}(10^2 - 10^3$ m^2 s$^{-1})$ in the ocean interior and can rise to $\mathcal{O}(10^4$ m^2 s$^{-1})$ in the ocean surface layer (*Abernathey et al.*, 2013; *Klocker and Abernathey*, 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).

50.3 Boundary fluxes of buoyancy

As introduced in Chapter 21, 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.

50.3.1 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.[1]
- 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. More generally, we simply consider this process to be associated with a salt flux between sea ice and ocean.
- Advective processes transfer enthalpy and salt across the ocean surface through the transfer of water mass across the interface.

50.3.2 Transport of scalars across the ocean surface

To develop a quantative understanding of how buoyancy is impacted by surface boundary fluxes, we develop the evolution equations for temperature, salinity, and mass in an arbitrary top model grid cell, and focus on evolution arising from surface boundary fluxes. For this purpose, recall the derivation of scalar budgets for an arbitrary grid cell depicted in Figure 15.2, and specialize to the case of a grid cell with upper boundary the air-sea interface. From equation (15.35), the transport of an abitrary scalar, ψ, crossing the upper ocean interface is written

$$Q_\psi \, \mathrm{d}A = -[\rho \, \psi \, (\boldsymbol{v} - \boldsymbol{v}^{(b)}) + \boldsymbol{J}] \cdot \hat{\boldsymbol{n}} \, \mathrm{d}\mathcal{S}, \tag{50.31}$$

where $\boldsymbol{J}$ is a subgrid scale flux, $\mathrm{d}\mathcal{S}$ is the area along the surface interface, and $\mathrm{d}A$ is the horizontal projection of that area. Equation (14.70c) renders the surface kinematic boundary condition

$$\rho \, (\boldsymbol{v} - \boldsymbol{v}^{(s)}) \cdot \hat{\boldsymbol{n}} \, \mathrm{d}\mathcal{S} = -Q_{\mathrm{m}} \, \mathrm{d}A \tag{50.32}$$

for the mass per time of material crossing the surface through such processes as precipitation, evaporation, and river runoff. We introduce the analogous expression for the subgrid scalar transport

$$\boldsymbol{J} \cdot \hat{\boldsymbol{n}} \, \mathrm{d}\mathcal{S} = -Q_{\mathrm{sgs}} \, \mathrm{d}A, \tag{50.33}$$

where Q_{sgs} encapsulates the turbulent fluxes of scalar, such as those for enthalpy and salt listed above. We are thus left with scalar flux across the surface ocean in the form

$$Q_\psi = Q_{\mathrm{m}} \, \psi + Q_{\mathrm{sgs}}. \tag{50.34}$$

By construction, a positive contribution to Q_ψ increases the amount of ψ within the ocean domain.

[1] The 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.

50.3.3 Scalar budgets for a surface ocean model grid cell

We now make use of the surface flux expression (50.34) within the finite volume budget equation (15.35) to develop budget equations for temperature, salinity, and mass for a grid cell. Again, we focus exclusively on contributions from surface boundary fluxes, with lateral transport important but not our present concern. The budget equations are thus given by

$$\frac{\partial\,(\rho\,\mathrm{d}z\,\Theta)}{\partial t} = Q_{\mathrm{m}}\,\Theta_{\mathrm{m}} + Q_{\Theta}^{\text{non-pen}} + \left[Q_{\Theta}^{\text{pen}}(z=\eta) - Q_{\Theta}^{\text{pen}}(z=-\Delta z)\right] \tag{50.35}$$

$$\frac{\partial\,(\rho\,\mathrm{d}z\,S)}{\partial t} = Q_{\mathrm{m}}\,S_{\mathrm{m}} + Q_S \tag{50.36}$$

$$\frac{\partial\,(\rho\,\mathrm{d}z)}{\partial t} = Q_{\mathrm{m}}. \tag{50.37}$$

We now detail the terms appearing in these equations.

- $\rho\,\mathrm{d}z$ is the mass per horizontal area of seawater in the grid cell. For a volume conserving Boussinesq fluid discussed in Chapter 25, the *in situ* density, ρ, is set to the constant reference density ρ_0

- Θ_{m} is the temperature of water crossing the ocean surface, and $C_{\mathrm{p}}^{o}\,Q_{\mathrm{m}}\,\Theta_{\mathrm{m}}$ is the associated enthalpy flux ($\mathrm{W\,m^{-2}}$). We further discuss this flux in Section 50.3.5.

- S_{m} is the salinity of water crossing the ocean surface, and $Q_{\mathrm{m}}\,S_{\mathrm{m}}$ is the associated mass flux of salt. S_{m} is typically taken to be zero, as for precipitation and evaporation. However, rivers can contain a nonzero salt concentration, as can sea ice melt. So we keep S_{m} for the following formulation. We further discuss this salt flux in Section 50.3.5.

- C_{p}^{o} is the seawater specific heat capacity at constant pressure ($\mathrm{J\,kg^{-1}\,{}^{\circ}C^{-1}}$). *IOC et al.* (2010) provides the most precise value appropriate for an ocean with potential enthalpy measured through Conservative Temperature.

- $Q_S = 1000\,Q_{C^{(S)}}$, where $Q_{C^{(S)}}$ is the mass flux of salt in units of $\mathrm{kg\,m^{-2}\,sec^{-1}}$. This flux measures the amount of salt per area per time that crosses the ocean surface, with $Q_{C^{(S)}} > 0$ when salt enters the ocean. It arises in the turbulent transfer of salt when sea ice forms and melts. When working with the equation (50.36) for salinity, $S = 1000\,C^{(S)}$, rather than the salt concentration equation for $C^{(S)}$, one must use $Q_S = 1000\,Q_{C^{(S)}}$. We further discuss this salt flux in Section 50.3.4.

- $C_{\mathrm{p}}^{o}\,Q_{\Theta}^{\text{non-pen}}$ is the non-penetrative surface enthalpy flux associated with turbulent processes (latent and sensible) and radiative longwave cooling ($\mathrm{W\,m^{-2}}$) localized to the upper ocean interface. $Q_{\Theta}^{\text{non-pen}} > 0$ for enthalpy entering the ocean surface (i.e., ocean warming). We further discuss this flux in Section 50.3.6.

- $C_{\mathrm{p}}^{o}\,Q_{\Theta}^{\text{pen}}(z=\eta)$ is the radiative shortwave enthalpy flux ($\mathrm{W\,m^{-2}}$) entering the ocean through its surface at $z=\eta$, with $Q_{\Theta}^{\text{pen}}(\eta) > 0$ warming the ocean surface. Likewise, $C_{\mathrm{p}}^{o}\,Q_{\Theta}^{\text{pen}}(z=-\Delta z)$ is the radiative shortwave heat flux leaving the top cell through its bottom interface. It is the difference $\left[Q_{\Theta}^{\text{pen}}(z=\eta) - Q_{\Theta}^{\text{pen}}(z=-\Delta z)\right]$ that contributes to the net warming of the grid cell from shortwave radiation. We further discuss the shortwave radiation flux in Section 50.3.8.

50.3.4 Salt fluxes from sea ice melt and formation

The turbulent mass flux of salt Q_S is positive for salt entering the ocean. There is transport of salt across the ocean surface when sea ice forms and melts, due to the nonzero salt content in sea ice. Otherwise, the surface salt flux is generally zero for the large scale ocean. For ocean models, however, the salt flux can be nonzero when formulating the surface boundary in terms of virtual salt fluxes rather than real water fluxes (*Huang*, 1993; *Griffies et al.*, 2001). It can also be nonzero when using an ocean-ice model that is not coupled to an atmosphere or land model, in which case salt restoring is required to maintain stability of the overturning circulation (see Section 3 of *Griffies et al.* (2009)).

50.3.5 Salt and enthalpy fluxes associated with water transport

In most cases, salinity in the water fluxed across the ocean surface is zero, so that $S_{\mathrm{m}} = 0$. However, there are some cases where rivers have a nonzero salinity so that $S_{\mathrm{m}} \neq 0$ and the product $Q_{\mathrm{m}}\, S_{\mathrm{m}}$ leads to an advective transport of salt across the ocean surface. Also, in some contexts the advective flux of salt due to sea ice melt and formation is encapsulated in the $Q_{\mathrm{m}}\, S_{\mathrm{m}}$ term rather than the turbulent salt flux Q_S.

Since water transported across the ocean has a nonzero temperature, this transport in turn affects the ocean enthalpy budget. One can either prescribe the temperature of this water, Θ_{m}, or the product $Q_{\mathrm{m}}\, \Theta_{\mathrm{m}}$. Consider the case where the product is specified for river water entering the ocean, which is the case with certain river models such as in *Dunne et al.* (2012). In this case, the enthalpy flux with respect to $0°C$ (in units of W m^{-2}) of liquid river runoff $\mathcal{H}^{\text{liquid runoff}}$ is given to the ocean from the river model, so that

$$Q_{\mathrm{m}}\, \Theta_{\mathrm{m}} = \frac{\mathcal{H}^{\text{liquid runoff}}}{C_p^{\text{liquid runoff}}}, \tag{50.38}$$

with $C_p^{\text{liquid runoff}}$ the specific heat capacity of the water coming in from the river runoff. Likewise, if the enthalpy associated with frozen runoff (e.g., calving land ice) is provided by the land model, then we have

$$Q_{\mathrm{m}}\, \Theta_{\mathrm{m}} = \frac{\mathcal{H}^{\text{solid runoff}}}{C_p^{\text{solid runoff}}}, \tag{50.39}$$

with $C_p^{\text{solid runoff}}$ the specific heat capacity of the solid runoff. These two heat capacities are typically provided by the component model (i.e., the river model) used to compute the runoff fields. Similar considerations hold for transfer of water between sea ice models and the ocean.

50.3.6 Non-penetrative surface enthalpy fluxes

The enthalpy flux $C_{\mathrm{p}}^{o}\, Q_{\Theta}^{\text{non-pen}}$ (W m^{-2}) is positive for enthalpy entering the ocean. This flux is comprised of the following contributions (e.g., see page 34 of *Gill*, 1982)

$$C_{\mathrm{p}}^{o}\, Q_{\Theta}^{\text{non-pen}} = Q_{\text{long}} + Q_{\text{latent}} + Q_{\text{sens}}. \tag{50.40}$$

These fluxes are termed non-penetrative since they are deposited or withdrawn from a small region near the ocean surface interface.

Longwave radiation

Longwave radiation leaves the ocean in the form of the $\sigma_{\rm SB}\, T^4$ Stefan-Boltzmann Law, with T the ocean skin temperature and

$$\sigma_{\rm SB} = 5.6734 \times 10^{-8} \text{ W m}^{-2} \text{ °K}^{-4} \tag{50.41}$$

the Stefan-Boltzmann constant. $Q_{\rm long} < 0$ since the longwave flux removes enthalpy from the ocean surface and sends it back to the atmosphere.

Latent enthalpy fluxes

$Q_{\rm latent}$ arises from phase changes whereby liquid seawater either evaporates, or it acts to melt frozen precipitation. In either case, $Q_{\rm latent} < 0$ since the liquid ocean loses enthalpy to energize the phase changes.

When seawater evaporates, the latent enthalpy lost by the ocean is determined by the latent enthalpy of vaporization for fresh water

$$H^{\rm vapor} = 2.5 \times 10^6 \text{ J kg}^{-1}, \tag{50.42}$$

so that

$$Q_{\rm evap} = H^{\rm vapor}\, Q_{\rm m}^{\rm evap} \tag{50.43}$$

where $Q_{\rm m}^{\rm evap} < 0$ is the mass flux (kg m^{-2} sec^{-1}) of fresh water leaving the ocean due to evaporation. A similar expression holds when seawater melts frozen precipitation (e.g., snow), in which case

$$H^{\rm fusion} = 3.34 \times 10^5 \text{ J kg}^{-1}, \tag{50.44}$$

so that

$$Q_{\rm melt} = -H^{\rm fusion}\, Q_{\rm m}^{\rm frozen\ precip}, \tag{50.45}$$

where $Q_{\rm m}^{\rm frozen\ precip} > 0$ is the mass flux (kg m^{-2} sec^{-1}) of frozen precipitation falling onto the ocean surface. Again, both $Q_{\rm evap}$ and $Q_{\rm melt}$ are negative since latent heating extracts heat from the ocean.

Sensible fluxes

$Q_{\rm sens}$ is the sensible enthalpy transfer proportional to the difference between the ocean temperature and that of the atmosphere, sea ice, or land ice. Sensible heating generally acts to cool the ocean ($Q_{\rm sens} < 0$), particularly near western boundary currents such as the Gulf Stream, Kuroshio, and Agulhas.

50.3.7 The case of frazil

As the temperature of seawater cools to the freezing point, sea ice is formed, initially through the production of frazil ice. Frazil can generally form at various levels in the upper ocean, though many ocean models assume frazil production occurs just in the top grid cell. Operationally in an ocean model, liquid water can be supercooled at any particular time step through surface fluxes and transport. An adjustment process is used to bring the liquid water temperature back to the freezing point, with the associated positive enthalpy flux $Q_{\rm frazil} > 0$ extracted from the ice model as frazil sea ice is formed.

50.3.8 Penetrative shortwave radiation

Momentum and buoyancy are transferred across the upper ocean surface boundary, with ocean processes such as advection and mixing then transporting the boundary momentum and buoyancy laterally as well as into the ocean interior. In contrast, penetrative shortwave radiation is absorbed into the ocean absent ocean transport processes, with such absorption a function of ocean optical properties. Optical properties are affected by flow (e.g., strongly turbulent flow is generally more opaque than laminar flow) as well as biological activity. Much of the shortwave radiation is absorbed in the upper ocean turbulent boundary layer, though a fraction leaks through to the interior. In general, such non-turbulent and non-advective transport of buoyancy via penetrative radiation represents a fundamentally novel aspect of ocean boundary layer physics relative to the atmosphere. Namely, for the atmosphere, radiative absorption is far less relevant than in the upper ocean, since the atmosphere is largely transparent to radiation. It is therefore useful to distinguish penetrative shortwave radiation from other buoyancy fluxes when formulating how boundary fluxes impact the ocean.

The penetrative shortwave radiative flux $C_{\mathrm{p}}^{o}\, Q_{\Theta}^{\mathrm{pen}} > 0$ arises from the net shortwave radiation entering through the ocean surface and absorbed by seawater. As noted above, a fraction of this radiation generally penetrates beneath the surface ocean grid cell, with the fraction depending on the optical properties of seawater and thickness of the grid cell. Hence, we subtract a flux $C_{\mathrm{p}}^{o}\, Q_{\Theta}^{\mathrm{pen}}(z = -\Delta z)$, which represents the radiative shortwave flux passing through the bottom of the surface ocean cell at $z = -\Delta z$. It is the difference

$$\text{net shortwave heating of surface grid cell} = C_{\mathrm{p}}^{o}\, [Q_{\Theta}^{\mathrm{pen}}(z = \eta) - Q_{\Theta}^{\mathrm{pen}}(z = -\Delta z)] \tag{50.46}$$

that stays in the surface grid cell.

50.3.9 Buoyancy budget for a surface ocean model grid cell

We now bring the previous fluxes together to form the budget for buoyancy in a surface grid cell due to the impacts of surface fluxes. The resulting expression is then used to derive an expression for the buoyancy forcing that acts on the ocean surface boundary layer. Buoyancy has a time tendency given by

$$-\left[\frac{\rho_o}{g}\right] \frac{\partial b}{\partial t} = \rho_\Theta \frac{\partial \Theta}{\partial t} + \rho_S \frac{\partial S}{\partial t}, \tag{50.47}$$

where we introduced the shorthand

$$\rho_\Theta = \left[\frac{\partial \rho}{\partial \Theta}\right]_{S,p} \qquad \text{and} \qquad \rho_S = \left[\frac{\partial \rho}{\partial S}\right]_{\Theta,p} \tag{50.48}$$

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 (50.35) by ρ_Θ and add to the surface salinity equation (50.36) multiplied by ρ_S

$$\rho_\Theta \left[\frac{\partial(\rho\, \mathrm{d}z\, \Theta)}{\partial t}\right] + \rho_S \left[\frac{\partial(\rho\, \mathrm{d}z\, S)}{\partial t}\right] = Q_{\mathrm{m}}\, (\rho_\Theta\, \Theta_{\mathrm{m}} + \rho_S\, S_{\mathrm{m}}) + \rho_\Theta \left[Q_{\Theta}^{\text{non-pen}} + \delta_k Q_{\Theta}^{\mathrm{pen}}\right] + \rho_S\, Q_S, \tag{50.49}$$

where we introduced the shorthand

$$\delta_k Q_{\Theta}^{\mathrm{pen}} = Q_{\Theta}^{\mathrm{pen}}(z = \eta) - Q_{\Theta}^{\mathrm{pen}}(z = -\Delta z). \tag{50.50}$$

We now use the mass budget (50.37) and introduce the buoyancy tendency according to equation (50.47) to realize an expression for the time tendency of the surface ocean buoyancy

$$(\rho_o/g)\,\rho\,\mathrm{d}z\,\frac{\partial b}{\partial t} = Q_{\mathrm{m}}\left[\rho_\Theta\,(\Theta-\Theta_{\mathrm{m}}) + \rho_S\,(S-S_{\mathrm{m}})\right] + \rho_\Theta\left[Q_\Theta^{\text{non-pen}} + \delta_k Q_\Theta^{\text{pen}}\right] - \rho_S\,Q_S. \tag{50.51}$$

Introducing the thermal expansion and saline contraction coefficients

$$\alpha_\Theta = -\frac{1}{\rho}\left[\frac{\partial\rho}{\partial\Theta}\right]_{S,p} \qquad \text{and} \qquad \beta_{\mathrm{S}} = \frac{1}{\rho}\left[\frac{\partial\rho}{\partial S}\right]_{\Theta,p} \tag{50.52}$$

yields

$$\mathrm{d}z\,\frac{\partial b}{\partial t} = \frac{g}{\rho_o}\left(Q_{\mathrm{m}}\left[-\alpha_\Theta\,(\Theta-\Theta_{\mathrm{m}}) + \beta_{\mathrm{S}}\,(S-S_{\mathrm{m}})\right] + \alpha_\Theta\,(\delta_k Q_\Theta^{\text{pen}} + Q_\Theta^{\text{non-pen}}) - \beta_{\mathrm{S}}\,Q_S\right). \tag{50.53}$$

50.3.10 Comments

The buoyancy flux expression (50.53) 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 51.

50.4 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.

50.4.1 Definitions and assumptions

Seawater mass is a conserved quantity so that the total liquid seawater mass, $\mathcal{M}$, changes only via boundary mass fluxes

$$\frac{\mathrm{d}\mathcal{M}}{\mathrm{d}t} = \mathcal{A}\,\overline{Q_{\mathrm{m}}}, \tag{50.54}$$

where $\mathcal{A}$ is the ocean surface area and $\overline{Q_{\mathrm{m}}}$ is the area averaged surface mass flux. The global volume of liquid seawater

$$\mathcal{V} = \frac{\mathcal{M}}{\langle\rho\rangle} \tag{50.55}$$

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 associated with volume changes along sloping beaches. We also assume constant area averaged ocean bottom depth, $\overline{H}$. These two assumptions mean that changes in ocean volume arise just from changes in global mean sea level, $\overline{\eta}$. Since around the year 2000, measurements estimate that global area mean sea level has increased at a rate of

$$\left[\frac{\mathrm{d}\overline{\eta}}{\mathrm{d}t}\right]_{\text{observed}} \approx 3\ \mathrm{mm\ yr^{-1}}, \tag{50.56}$$

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 $\mathcal{V} \approx 1.3 \times 10^{18}$ m^3
- global ocean surface area $\mathcal{A} \approx 3.6 \times 10^{14}$ m^2
- global ocean mean density $\langle \rho \rangle \approx 1035$ kg m^{-3}
- specific heat capacity for seawater $c_{\mathrm{p}} \approx 3992$ J kg^{-1}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

$$\alpha_\Theta = -\frac{1}{\rho}\left[\frac{\partial \rho}{\partial \Theta}\right]_{S,p} \approx 2 \times 10^{-4}\ \mathrm{K}^{-1}. \tag{50.57}$$

  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.

50.4.2 Budget for global mean sea level

Expression (50.55) for ocean volume leads to its time derivative

$$\frac{\mathrm{d}\mathcal{V}}{\mathrm{d}t} = \frac{1}{\langle \rho \rangle}\frac{\mathrm{d}\mathcal{M}}{\mathrm{d}t} - \frac{\mathcal{M}}{\langle \rho \rangle^2}\frac{\mathrm{d}\langle \rho \rangle}{\mathrm{d}t} \tag{50.58a}$$

$$= \frac{\mathcal{A}\,\overline{Q_{\mathrm{m}}}}{\langle \rho \rangle} - \frac{\mathcal{V}}{\langle \rho \rangle}\frac{\mathrm{d}\langle \rho \rangle}{\mathrm{d}t}, \tag{50.58b}$$

where we used equation (50.54) to express mass changes in terms of the surface mass flux. Additionally, the ocean volume is given by

$$\mathcal{V} = \int \mathrm{d}A \int_{-H}^{\eta} \mathrm{d}z = \mathcal{A}\,(\overline{H} + \overline{\eta}), \tag{50.59}$$

so that its time changes arise from changes in the global mean sea level

$$\frac{\mathrm{d}\mathcal{V}}{\mathrm{d}t} = \mathcal{A}\,\frac{\mathrm{d}\overline{\eta}}{\mathrm{d}t}. \tag{50.60}$$

Combining the two volume equations (50.58b) and (50.60) yields the budget equation for global mean sea level

$$\frac{\mathrm{d}\overline{\eta}}{\mathrm{d}t} = \frac{\overline{Q_{\mathrm{m}}}}{\langle \rho \rangle} - \frac{\mathcal{V}}{\mathcal{A}\,\langle \rho \rangle}\frac{\mathrm{d}\langle \rho \rangle}{\mathrm{d}t}. \tag{50.61}$$

The first term arises from changes in ocean mass whereas the second term arises from changes in global mean seawater density.

50.4.3 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

$$\frac{1}{2}\left[\frac{\mathrm{d}\overline{\eta}}{\mathrm{d}t}\right]_{\text{observed}} = \frac{\overline{Q_{\mathrm{m}}}}{\langle\rho\rangle}. \tag{50.62}$$

This distribution of measured sea level is roughly correct. With $\langle\rho\rangle = 1035$ kg m^{-3} and $\mathrm{d}\overline{\eta}/\mathrm{d}t \approx 3$ mm yr^{-1}, we need an area averaged mass flux across the ocean surface

$$\overline{Q_{\mathrm{m}}} \approx 5 \times 10^{-8} \text{ kg m}^{-2} \text{ s}^{-1}. \tag{50.63}$$

Integrated over the global ocean area, this flux leads to a mass transport of

$$\mathcal{T} = \mathcal{A}\ \overline{Q_{\mathrm{m}}} \approx 1.8 \times 10^{7} \text{ kg s}^{-1} \approx 0.015 \times \mathcal{T}^{\text{river}}. \tag{50.64}$$

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.

50.4.4 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 (50.61) we know that steric changes are written mathematically as

$$\left[\frac{\mathrm{d}\overline{\eta}}{\mathrm{d}t}\right]_{\text{steric}} \equiv -\frac{\mathcal{V}}{\mathcal{A}\,\langle\rho\rangle}\,\frac{\mathrm{d}\langle\rho\rangle}{\mathrm{d}t}. \tag{50.65}$$

Global mean density will change primarily from changes in global mean Conservative Temperature. If we assume the ocean thermal expansion is constant, then

$$\frac{1}{\langle\rho\rangle}\frac{\mathrm{d}\langle\rho\rangle}{\mathrm{d}t} = -\alpha_{\Theta}\frac{\mathrm{d}\langle\Theta\rangle}{\mathrm{d}t}, \tag{50.66}$$

so that steric sea level changes are primarily driven by *thermosteric* effects

$$\left[\frac{\mathrm{d}\overline{\eta}}{\mathrm{d}t}\right]_{\text{thermosteric}} \equiv \frac{\alpha_{\Theta}\,\mathcal{V}}{\mathcal{A}}\,\frac{\mathrm{d}\langle\Theta\rangle}{\mathrm{d}t}. \tag{50.67}$$

With $\alpha_{\Theta} \approx 2 \times 10^{-4}$ K^{-1} and $\mathrm{d}\overline{\eta}/\mathrm{d}t \approx 1.5$ mm yr^{-1}, we have

$$\frac{\mathrm{d}\langle\Theta\rangle}{\mathrm{d}t} \approx 0.2 \text{ K century}^{-1}. \tag{50.68}$$

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 K century^{-1}.

50.4.5 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

$$\overline{Q_{\mathrm{H}}} \approx \langle \rho \rangle \, c_{\mathrm{p}} \, \overline{H} \, \frac{\mathrm{d}\langle \Theta \rangle}{\mathrm{d}t}. \tag{50.69}$$

Plugging in numbers leads to the enthalpy flux

$$\overline{Q_{\mathrm{H}}} \approx 1 \text{ W m}^{-2}. \tag{50.70}$$

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,[2] 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 ($\Delta t = 1$ s) and uniformly distributing its released energy over the ocean surface area every second. $\mathcal{E}_{\mathrm{bomb}} \approx 6.3 \times 10^{13}$ J distributed over the ocean area each second corresponds to a surface ocean enthalpy flux of

$$Q_{\mathrm{bomb}} = \frac{\mathcal{E}_{\mathrm{bomb}}}{\mathcal{A} \, \Delta t} \approx 0.17 \text{ W m}^{-2}. \tag{50.71}$$

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.

50.4.6 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 (16.8))

$$\frac{\partial \eta^{\mathrm{bouss}}}{\partial t} = -\nabla \cdot \boldsymbol{U} + Q_{\mathrm{m}}/\rho_o, \tag{50.72}$$

where

$$\boldsymbol{U} = \int_{-H}^{\eta} \boldsymbol{u} \, \mathrm{d}z \tag{50.73}$$

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 (50.72) over the surface area of the global ocean reveals that the global mean Boussinesq sea level evolves according to

$$\frac{\mathrm{d}\overline{\eta}^{\mathrm{bouss}}}{\mathrm{d}t} = \frac{\overline{Q_{\mathrm{m}}}}{\rho_o}. \tag{50.74}$$

[2] The ocean covers about 70% of the earth surface.

Hence, $\overline{\eta}^{\text{bouss}}$ 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, $\overline{Q_{\text{H}}} \neq 0$. 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).

50.4.7 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 is volume is almost entirely barystatic, with the global halosteric effect entirely negligible when considering global sea level changes.

50.4.8 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.

51

Water mass and tracer mass analysis

As commonly used in oceanography, a *water mass* refers to a finite region of seawater characterized by a suite of scalar properties that distinguish it from other regions. Water masses are typically formed through 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 *transformed* by irreversible mixing processes. Water masses offer a conceptual means to partition the ocean into distinct classes whose origin, movement, and modification can be measured, modeled, and studied. Furthermore, scalar properties are generally simpler to measure than vector properties such as velocity, with a water mass perspective offering the means to infer properties of the ocean circulation without directly measuring the velocity field. 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 defined by scalar properties such as buoyancy, Conservative Temperature, salinity, or biogeochemical tracers.

READER'S GUIDE FOR THIS CHAPTER

The theoretical development in this chapter requires a heavy dose of the vector calculus encountered in Chapter 2; elements of the generalized vertical coordinates from Chapters 8 and 17; kinematics from Chapter 14; tracer budgets from Chapter 15; advection and diffusion maths in Chapter 47, and features of parameterized tracer transport and mixing discussed in Chapter 49. Although some of this prior material is reviewed here to keep the discussion reasonably self-contained, the reader may find it necessary to revisit some of the earlier material to garner a full understanding of the concepts and manipulations.

The formalism of water mass and tracer mass analysis can appear mysterious to the novice. To dispel the mystery we attempt to couple the mathematical equations with schematics and specific applications. Even so, the novice should be prepared to devote a nontrivial level of time and energy to realize a level of proficiency with the tools and understanding of the concepts.

51.1 Conceptual introduction **918**
51.2 Transformation and formation of buoyancy **920**
51.2.1 A three-layer thought experiment 920
51.2.2 Characterizing how processes lead to transformations 921
51.3 Seawater mass budget within layers **922**

51.3.1 Dia-surface transport and water mass transformation 922
51.3.2 The volume of a λ-layer 923
51.3.3 Practical expression for the water mass transformation 925
51.3.4 Transport crossing open boundaries and ocean surface 926
51.3.5 The area between two λ-interfaces intersecting a surface 926
51.3.6 The layer mass budget 927
51.3.7 Circulation in λ-space 928
51.3.8 Further study 929
51.4 Tracer mass analysis **930**
51.4.1 General form of the mass budget 931
51.4.2 Tracer processes 931
51.4.3 Transport across an interior layer interface 932
51.4.4 Transport across interior and surface boundaries 933
51.4.5 The layer tracer budget 933
51.4.6 Further study 933
51.5 Regions bounded by a tracer contour/surface **933**
51.5.1 Closed region bounded by a tracer surface/contour 934
51.5.2 Region with $C \geq \tilde{C}$ 936
51.5.3 Further reading 937
51.6 Tracer mass distributions **937**
51.6.1 Tracer mass density function 938
51.6.2 Seawater mass and tracer mass 938
51.6.3 Internal and external tracer masses 940
51.6.4 Further reading 940
51.7 $\Theta - S$ water mass analysis **940**

51.1 Conceptual introduction

Water mass analysis is a mathematical formalism aimed at studying the budgets for seawater mass and tracer mass within layers defined by scalar properties such as buoyancy, Conservative Temperature, salinity, or biogeochemical tracers. We are concerned with processes affecting the evolution of these layers and in the characterization of circulation inferred from this evolution. We speak of a *water mass configuration space*, which is three dimensional just like the geographic/depth space (x, y, z) and yet with some or all of the "spatial" coordinates determined by properties other than geographic/depth position. For example, Figure 51.1 provides an example of (Θ, S, p) space, whereby a fluid element is positioned in water mass configuration space as determined by its value for Conservative Temperature, Θ, salinity, S, and pressure, p. Other examples include the $(x, y, \rho_{\text{potential}})$ of an isopycnal model (Chapters 8 and 17), and yet here without the constraint that the density surfaces are all stably stratified. We prefer the term "configuration space" over the alternative "phase space", since phase space in Hamiltonian dynamics refers to position and momentum coordinates. "Configuration space" is somewhat more generic and allows us to follow a mathematical approach akin to that in equilibrium thermodynamics (Chapter 22).

A water mass configuration space views the ocean circulation through the lens afforded by layers rather than the fixed Eulerian points or the moving Lagrangian particles. 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 12 for a reminder of Eulerian and Lagrangian kinematics).

It is critical to note that there is no assumption of a one-to-one mapping between geographic/depth space and water mass configuration space. Namely, a point in water mass configuration

Figure 51.1: *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 (Θ, S, p) with Conservative Temperature, Θ, salinity, S, and pressure, p. Mapping between the two spaces is generally not one-to-one. 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 (Θ, S, p).*

space can be comprised of many points in geographic/depth space, which is familiar to descriptive oceanographers who well know that there can be more than one fluid element in the ocean that has the same (Θ, S, p) values. This lack of one-to-one mapping is both liberating and frustrating. It is liberating since it offers a framework to infer ocean circulation in water mass configuration space even without measuring velocity. It is frustrating since that circulation is viewed with incomplete geographic/depth information.

Determining ocean circulation in a water mass configuration space requires use to determine how the layers defining water masses are modified or *transformed* by irreversible boundary and interior ocean processes. Such processes fundamentally lead to material changes to fluid elements as arise from 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. Notably, and quite trivially, there is 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 waves and laminar flow generally render nontrivial motion in geographical/depth space and yet lead to no motion in (Θ, S) space.

So if one is interested in reversible processes, then water mass configuration space may not be appropriate. Indeed, given the possible non-local geographic aspects of water mass configuration space, it is an unnatural space to study forces and stresses acting between adjacent fluid elements, so that momentum dynamics is better handled via Eulerian or Lagrangian kinematics. Instead, water mass configuration space more naturally reveals the dynamics associated with irreversible processes that affect the scalar fields defining the layers. Consequently, a water mass configuration space perspective has found use in framing many questions of primary interest in the Anthropocene, such as ocean buoyancy and its transformation through interior and boundary mixing, ocean heat uptake and transport, steric sea level rise, and irreversible changes to biogeochemical properties.

51.2 Transformation and formation of buoyancy

The most common scalar for partitioning the ocean is buoyancy, or its approximate form given by potential density (Chapter 21). In this section we introduce the notions of *transformation* and *formation* when partitioning the ocean according to density (γ) classes that locally measure buoyancy. The ideas presented here extend to any scalar field used to bin the ocean fluid.

Water mass *transformation* measures the mass per time of water that moves across the boundaries of a layer. By convention, the transformation is positive if water moves into a region of larger density (more generally into a region of larger value for the scalar field) and negative if it enters a lighter density layer (lower value for the scalar field). Water mass *formation* refers to the difference in transformation across a layer, so that it measures the change in mass of the layer. That is, *formation is the convergence of transformation in layer space.* Both transformation and formation have dimensions of mass per time (or volume per time when considering Boussinesq fluids). They are typically measured in Sverdrup units, which are given by the following:

$$1 \text{ Sv} = 10^6 \text{ m}^3 \text{ s}^{-1} \qquad \text{volume-Sverdrup} \tag{51.1a}$$

$$1 \text{ Sv} = 10^9 \text{ kg s}^{-1} \qquad \text{mass-Sverdrup.} \tag{51.1b}$$

The volume-Sverdrup is commonly used in Boussinesq fluids with kinematics based on volume conservation, whereas the mass-Sverdrup is commonly used in non-Boussinesq fluids with kinematics based on mass conservation.

51.2.1 A three-layer thought experiment

To illustrate the concepts of transformation and formation, consider an ocean that is binned into three density layers (or density "classes") bounded by four density interfaces:

$$\text{light density layer} = [\gamma - \delta\gamma/2, \gamma + \delta\gamma/2] \tag{51.2a}$$

$$\text{middle density layer} = [\gamma + \delta\gamma/2, \gamma + 3\,\delta\gamma/2] \tag{51.2b}$$

$$\text{heavy density layer} = [\gamma + 3\,\delta\gamma/2, \gamma + 5\,\delta\gamma/2]. \tag{51.2c}$$

Figure 51.2 depicts a sample mass distribution for this ocean; i.e., the mass census for water binned into these three density layers. 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 interfaces; i.e., $G(\sigma)$ is the transformation. This particular thought experiment has the following transformations across the various layer interfaces

$$G(\sigma) = \left[\begin{array}{lll} 0 & \sigma = \gamma - \delta\gamma/2 & \text{closed boundary} \\ < 0 & \sigma = \gamma + \delta\gamma/2 & \text{mass moves to light density from middle density} \\ > 0 & \sigma = \gamma + 3\,\delta\gamma/2 & \text{mass moves from middle density to heavy density} \\ 0 & \sigma = \gamma + 5\,\delta\gamma/2 & \text{closed boundary.} \end{array} \right. \tag{51.3}$$

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:

$$\text{light-formation} = -[G(\gamma + \delta\gamma/2) - G(\gamma - \delta\gamma/2)] > 0 \tag{51.4a}$$

$$\text{middle-formation} = -[G(\gamma + 3\,\delta\gamma/2) - G(\gamma + \delta\gamma/2)] < 0 \tag{51.4b}$$

$$\text{heavy-formation} = -[G(\gamma + 5\,\delta\gamma/2) - G(\gamma + 3\,\delta\gamma/2)] > 0. \tag{51.4c}$$

Figure 51.2: *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.*

51.2.2 Characterizing how processes lead to transformations

The key focus of water mass analysis is the movement of water between layers, with this water movement modifying the distribution of mass within the layers. There are a number of processes that affect the movement of mass across layer boundaries. One involves the passage of mass across domain boundaries such as the surface ocean. In particular, rain and evaporation alter the mass of the ocean, with these mass changes appearing in the mass budget for those density layers that outcrop at the ocean surface.

Mixing provides a second means to move water across layer boundaries, with scalar fields (of non-uniform distribution) materially modified in the presence of mixing. As discussed in Chapter 15, mixing causes trace matter to move between fluid elements even as mixing does not alter the net mass of fluid elements (recall our discussion of barycentric velocity in Section 15.1.2). Hence, seawater fluid elements retain a fixed mass and yet this mass is redistributed among layers since the layer boundaries move. 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 is commonly faster than the movement of fluid elements. Hence, through the seasonal cycle a density bin can inflate or deflate 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, 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 15.6). We generally ignore non-conservative processes in this chapter. However, they are important for treatments that consider biology and chemistry, and they can be incorporated into the formalism considered here through the use of source terms (see *Groeskamp et al.* (2019) for a review of such applications).

51.3 Seawater mass budget within layers

In this section we put some mathematical flesh onto the conceptual bones from Section 51.2 by developing a formalism to quantify water mass budgets within a layer such as in Figure 51.3. For generality, we formulate water mass transformation analysis in terms of an arbitrary smooth scalar field, $\lambda = \lambda(\boldsymbol{x}, t)$, with seawater distributed into discrete bins or classes according to the value of λ. We thus seek a formalism to quantify the evolution of seawater distributions within the λ-classes.

Figure 51.3: *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 $\Omega(\lambda \pm \Delta\lambda/2)$. The water mass transformations crossing the layer interfaces $\partial\Omega(\lambda \pm \Delta\lambda/2)$ are written $G(\lambda \pm \Delta\lambda/2)$, 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. 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 unstable 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 $\partial\Omega(\lambda \pm \Delta\lambda/2) = \partial\Omega_{in}(\lambda \pm \Delta\lambda/2) + \partial\Omega_{out}(\lambda \pm \Delta\lambda/2) + \partial\Omega(\lambda + \Delta\lambda/2) + \partial\Omega(\lambda - \Delta\lambda/2)$. This figure is based on Figure 3 of Groeskamp et al. (2019) and is oriented according to the Southern Ocean with Antarctica to the left.*

In Section 51.2 we assumed λ is a density field, $\lambda = \gamma$, whereas here it is a generic scalar field. Furthermore, in contrast to the case of a generalized vertical coordinate (Chapters 8 and 17), we make no assumption regarding the stratification of λ. Rather, λ-isosurfaces are here free to overturn (e.g., see Figure 51.3) and to disconnect. This freedom is motivated by the behavior of most oceanographic scalar property fields, even buoyancy within boundary layers. As noted in the chapter introduction, this added freedom comes at a cost. Namely, we no longer have a one-to-one relation between a point in water mass configuration space and a point in geographic/depth space. Yet we choose to pay that price given the novel insights into ocean circulation when viewed within water mass configuration space.

51.3.1 Dia-surface transport and water mass transformation

A fundamental element in the analysis of water mass transformation is the dia-surface transport velocity from Section 17.3.7

$$w^{\text{dia}} = \hat{\boldsymbol{n}} \cdot (\boldsymbol{v} - \boldsymbol{v}^{(\lambda)}) = \frac{\dot{\lambda}}{|\nabla\lambda|} \qquad \text{with} \qquad \hat{\boldsymbol{n}} = \frac{\nabla\lambda}{|\nabla\lambda|} \qquad \text{and} \qquad \dot{\lambda} = \frac{\mathrm{D}\lambda}{\mathrm{D}t}. \tag{51.5}$$

The transport $w^{\rm dia}$ locally measures the amount of seawater that penetrates a λ-interface in the direction of increasing λ, with the λ-interface moving with a non-zero velocity, $\boldsymbol{v}^{(\lambda)}$, defined according to

$$\frac{\partial \lambda}{\partial t} + \boldsymbol{v}^{(\lambda)} \cdot \nabla \lambda = 0. \tag{51.6}$$

For a particular λ-interface labelled by $\lambda(\boldsymbol{x}, t) = \lambda^*$ we define the transformation, $G(\lambda^*)$, as the area integral of $\rho\, w^{\rm dia}$ over the λ^*-interface

$$G(\lambda^*) = \int_{\partial\Omega(\lambda^*)} \rho\, w^{\rm dia}\, {\rm d}\mathcal{S} = \int_{\partial\Omega(\lambda^*)} \frac{\rho\, \dot{\lambda}}{|\nabla \lambda|}\, {\rm d}\mathcal{S} \Longrightarrow \text{dimensions} = \text{M T}^{-1}. \tag{51.7}$$

We write $\partial\Omega(\lambda^*)$ for the geographical area occupied by the interface (see Figure 51.3).

Recall our conceptual description of water mass transformation in Section 51.2.2. Based on the definition (51.7), we see that water mass transformation across a λ-interface occurs when there is a material change, $\dot{\lambda} \neq 0$, in the surface interface. Such changes arise from mixing and boundary processes thus causing irreversible changes to the matter content within λ-layers by driving seawater to cross λ-interfaces. *Formation* occurs when more water is transformed into a layer than that leaving the layer, thus leading to a modified distribution of mass within the λ-layers.

Equation (51.7) for the water mass transformation is written as a surface integral of the dia-surface transport across the λ interface. In practice, the surface integral is not easy to compute diagnostically since the λ-surface can be highly corrugated and disconnected, and discrete estimates of $|\nabla\lambda|^{-1}$ are prone to nontrivial errors and noise. A practical evaluation of the water mass transformation thus requires an alternative method. In the next few sections we derive mathematical results that allow us to develop the practical means for computing the transformation.

51.3.2 The volume of a λ-layer

Consider how we measure the volume within an infinitesimal λ-layer bounded by two interfaces $[\lambda - \delta\lambda/2, \lambda + \delta\lambda/2]$, with Figure 51.4 providing an example. The volume within a tiny cylinder extending from one interface to the other is given by

$$\delta V = \delta h\, \delta\mathcal{S} \tag{51.8}$$

where ${\rm d}\mathcal{S}$ is the area element and δh is the layer thickness. The differential λ-increment separating the two interfaces is given by

$$\delta\lambda = \nabla\lambda \cdot \delta\boldsymbol{x} = |\nabla\lambda|\, \hat{\boldsymbol{n}} \cdot \delta\boldsymbol{x} = |\nabla\lambda|\, \delta h, \tag{51.9}$$

so that the layer thickness is

$$\delta h = \frac{\delta\lambda}{|\nabla\lambda|}. \tag{51.10}$$

Hence, for a given increment, $\delta\lambda > 0$, the layer thickness is inversely proportional to the modulus of the λ-gradient. The thickness is oriented according to the normal direction, $\hat{\boldsymbol{n}} = \nabla\lambda\, |\nabla\lambda|^{-1}$, so that it measures the distance between the interfaces in the direction of the normal.

Given the expression (51.10) for the layer thickness, the volume for the cylinder between two infinitesimally close λ-interfaces is

$$\delta V = \delta h\, \delta\mathcal{S} = \frac{\delta\lambda\, \delta\mathcal{S}}{|\nabla\lambda|}. \tag{51.11}$$

Figure 51.4: *This schematic shows an infinitesimally thin λ-layer bounded by two interfaces $[\lambda - \delta\lambda/2, \lambda + \delta\lambda/2]$. We here assume the λ-interfaces are monotonically layered in one of the three directions (e.g., stably stratified density in the vertical). However, this assumption is not generally needed for water mass analysis; we make it here only for illustrative purposes. The cylinder region extends between the two iso-surfaces and it has thickness $\delta h = \delta\lambda/|\nabla\lambda|$ and cross-sectional area $\delta\mathcal{S}$. The cylinder is oriented according to the normal direction $\hat{\boldsymbol{n}} = |\nabla\lambda|^{-1}\,\nabla\lambda$. Note that we assume that $|\nabla\lambda| \neq 0$, 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.*

It follows that the volume of a finite layer bounded by two interfaces is given by the integral over the layer

$$\Delta V(\lambda) \equiv \int_{\Omega(\lambda\pm\Delta\lambda/2)} \mathrm{d}V = \int_{\lambda-\Delta\lambda/2}^{\lambda+\Delta\lambda/2} \left[\int_{\partial\Omega(\lambda')} \frac{\mathrm{d}\mathcal{S}}{|\nabla\lambda'|}\right] \mathrm{d}\lambda'. \tag{51.12}$$

In addition to determining the volume within a layer, we find it useful to measure the volume of seawater with values of λ greater than a given value; e.g., for $\lambda_0 \leq \lambda$ where λ_0 is an arbitrary reference value. We depict this situation in Figure 51.5. The volume of water in this region is written

$$V(\lambda) \equiv \int_{\Omega(\lambda_0\leq\lambda)} \mathrm{d}V = \int_{\lambda_0}^{\lambda} \left[\int_{\partial\Omega(\lambda')} \frac{\mathrm{d}\mathcal{S}}{|\nabla\lambda'|}\right] \mathrm{d}\lambda', \tag{51.13}$$

where the integral excludes regions that are not in the ocean. Correspondingly, the volume within a λ-layer $[\lambda - \Delta\lambda/2, \lambda + \Delta\lambda/2]$ is given by the difference

$$\Delta V(\lambda) = V(\lambda + \Delta\lambda/2) - V(\lambda - \Delta\lambda/2). \tag{51.14}$$

Notably, the arbitrary reference value, λ_0, drops out from the difference. Taking the limit $\Delta\lambda \to 0$ renders the derivative

$$\frac{\partial V(\lambda)}{\partial\lambda} = \lim_{\Delta\lambda\to 0} \frac{V(\lambda + \Delta\lambda/2) - V(\lambda - \Delta\lambda/2)}{\Delta\lambda} = \int_{\partial\Omega(\lambda)} \frac{\mathrm{d}\mathcal{S}}{|\nabla\lambda|}, \tag{51.15}$$

which provides a measure for the volume per unit λ within a layer. This is result is a finite volume extension of the infinitesimal result (51.11). We will make use of it in the following.

Figure 51.5: *Depicting the volume of fluid with $\lambda_0 \leq \lambda$, where λ_0 is an arbitrary reference value, with the expression for the volume given by equation (51.13). For the special case of λ_0 less than any value of λ in the ocean, then the region with $\lambda_0 \leq \lambda$ encompasses full ocean domain.*

51.3.3 Practical expression for the water mass transformation

We now extend the identities given by equation (51.15) to include the volume integral of an arbitrary function, $\Phi(\boldsymbol{x}, t)$

$$\mathcal{I}_\Phi(\lambda) = \int_{\Omega(\lambda_0 \leq \lambda)} \Phi \, \mathrm{d}V = \int_{\lambda_0}^{\lambda} \left[\int_{\partial\Omega(\lambda')} \frac{\Phi \, \mathrm{d}\mathcal{S}}{|\nabla\lambda'|} \right] \mathrm{d}\lambda'. \tag{51.16}$$

Taking the λ-derivative of the volume integral and using the fundamental theorem of calculus picks out the surface integral

$$\frac{\partial \mathcal{I}_\Phi(\lambda)}{\partial\lambda} = \frac{\partial}{\partial\lambda} \left(\int_{\lambda_0}^{\lambda} \left[\int_{\partial\Omega(\lambda')} \frac{\Phi \, \mathrm{d}\mathcal{S}}{|\nabla\lambda'|} \right] \mathrm{d}\lambda' \right) = \int_{\partial\Omega(\lambda)} \frac{\Phi \, \mathrm{d}\mathcal{S}}{|\nabla\lambda|}. \tag{51.17}$$

The identity (51.17) is the key to a practical discrete approximation of the surface integral

$$\int_{\partial\Omega(\lambda)} \frac{\Phi \, \mathrm{d}\mathcal{S}}{|\nabla\lambda|} = \frac{\partial \mathcal{I}_\Phi(\lambda)}{\partial\lambda} \tag{51.18a}$$

$$= \lim_{\Delta\lambda \to 0} \frac{\mathcal{I}_\Phi(\lambda + \Delta\lambda/2) - \mathcal{I}_\Phi(\lambda - \Delta\lambda/2)}{\Delta\lambda} \tag{51.18b}$$

$$= \lim_{\Delta\lambda \to 0} \frac{1}{\Delta\lambda} \left(\int_{\lambda_0}^{\lambda+\Delta\lambda/2} \left[\int_{\partial\Omega(\lambda')} \frac{\Phi \, \mathrm{d}\mathcal{S}}{|\nabla\lambda'|} \right] \mathrm{d}\lambda' - \int_{\lambda_0}^{\lambda-\Delta\lambda/2} \left[\int_{\partial\Omega(\lambda')} \frac{\Phi \, \mathrm{d}\mathcal{S}}{|\nabla\lambda'|} \right] \mathrm{d}\lambda' \right) \tag{51.18c}$$

$$= \lim_{\Delta\lambda \to 0} \frac{1}{\Delta\lambda} \left(\int_{\lambda-\Delta\lambda/2}^{\lambda+\Delta\lambda/2} \left[\int_{\partial\Omega(\lambda')} \frac{\Phi \, \mathrm{d}\mathcal{S}}{|\nabla\lambda'|} \right] \mathrm{d}\lambda' \right) \tag{51.18d}$$

$$= \lim_{\Delta\lambda \to 0} \frac{1}{\Delta\lambda} \int_{\Omega(\lambda \pm \Delta\lambda/2)} \Phi \, \mathrm{d}V. \tag{51.18e}$$

The final expression is a volume integral of Φ computed within a thin layer surrounding the λ-interface.

Setting $\Phi = \rho\,\dot{\lambda}$ in equation (51.18e) renders an expression for the water mass transformation across the λ-interface

$$G(\lambda) = \int_{\partial\Omega(\lambda)} \rho\, w^{\mathrm{dia}}\, \mathrm{d}\mathcal{S} = \int_{\partial\Omega(\lambda)} \frac{\rho\,\dot{\lambda}\,\mathrm{d}\mathcal{S}}{|\nabla\lambda|} = \frac{\partial}{\partial\lambda}\int_{\Omega(\lambda_0\le\lambda)} \rho\,\dot{\lambda}\,\mathrm{d}V = \lim_{\Delta\lambda\to 0}\frac{1}{\Delta\lambda}\int_{\Omega(\lambda\pm\Delta\lambda/2)} \rho\,\dot{\lambda}\,\mathrm{d}V. \tag{51.19}$$

The final expression lends itself to a straightforward binning algorithm that forms the basis for the practice of water mass transformation analysis (see Section 7.5 of *Groeskamp et al.* (2019)). In this manner, calculation of water mass transformation in a model or from observational measurements requires the various contributions to the material time change, $\dot{\lambda}$, a weighting of those time changes according to the mass of the fluid element, $\rho\,\mathrm{d}V$, and binning of $\rho\,\dot{\lambda}\,\mathrm{d}V$ according to λ-classes. Notably, many scalars materially evolve according to $\rho\,\dot{\lambda} = -\nabla\cdot\boldsymbol{J}$, so that the volume integral in equation (51.19) reduces, through Gauss' divergence theorem, to surface integrals of $\hat{\boldsymbol{n}}\cdot\boldsymbol{J}$ along the layer boundaries. However, in practice it is more convenient to accumulate $\rho\,\dot{\lambda}\,\mathrm{d}V$ within the thin layer according to the discretized volume integral, rather than to perform the area integral of the flux around the domain boundary. The reason to prefer the volume integral is that it is amenable to binning whereas the area integral is not.

51.3.4 Transport crossing open boundaries and ocean surface

According to Figure 51.3, we need two more transport terms to develop a layer mass budget. The mass transport leaving the layer through an interior open boundary is given by

$$\Delta\Psi(\lambda\pm\Delta\lambda/2) = \int_{\partial\Omega_{\mathrm{in}}(\lambda\pm\Delta\lambda/2)} \rho\,(\boldsymbol{v}-\boldsymbol{v}^{(b)})\cdot\hat{\boldsymbol{n}}\,\mathrm{d}\mathcal{S}, \tag{51.20}$$

where $\boldsymbol{v}^{(b)}$ is the velocity for a point on the boundary and $\hat{\boldsymbol{n}}$ is the outward normal along the boundary. As an explicit example, set the boundary to be at a particular latitude so that $\boldsymbol{v}^{(b)} = 0$ and $\hat{\boldsymbol{n}} = \hat{\boldsymbol{y}}$, in which case

$$\Delta\Psi(\lambda\pm\Delta\lambda/2) = \int_{\partial\Omega_{\mathrm{in}}(\lambda\pm\Delta\lambda/2)} \rho\, v\,\mathrm{d}x\,\mathrm{d}z, \tag{51.21}$$

where $\partial\Omega_{\partial\Omega_{\mathrm{in}}(\lambda\pm\Delta\lambda/2)}$ specifies the depth and longitude range over the constant latitude boundary.

In a similar manner we write the mass transport crossing the ocean free surface

$$\Delta W(\lambda\pm\Delta\lambda/2) = -\int_{\partial\Omega_{\mathrm{out}}(\lambda\pm\Delta\lambda/2)} \rho\,(\boldsymbol{v}-\boldsymbol{v}^{(b)})\cdot\hat{\boldsymbol{n}}\,\mathrm{d}\mathcal{S} = \int_{\partial\Omega_{\mathrm{out}}(\lambda\pm\Delta\lambda/2)} Q_{\mathrm{m}}\mathrm{d}A, \tag{51.22}$$

where we made use of the surface kinematic boundary condition (equation (14.70c)) to write

$$\rho\,(\boldsymbol{v}-\boldsymbol{v}^{(b)})\cdot\hat{\boldsymbol{n}}\,\mathrm{d}\mathcal{S} \equiv -Q_{\mathrm{m}}\,\mathrm{d}A, \tag{51.23}$$

where $Q_{\mathrm{m}}\,\mathrm{d}A$ is the mass transport of water crossing the free surface ($Q_{\mathrm{m}} > 0$ for water entering the ocean) and $\mathrm{d}A = \mathrm{d}x\,\mathrm{d}y$ is the horizontal area element.

51.3.5 The area between two λ-interfaces intersecting a surface

The transports $\Delta W(\lambda\pm\Delta\lambda/2)$ and $\Delta\Psi(\lambda\pm\Delta\lambda/2)$ are computed over the area of a surface bounded by where the interfaces $\lambda\pm\Delta\lambda/2$ intersect that surface. We can express these area integrals in a manner analogous to the volume integrals considered previously. For that purpose consider the

area between two λ values along a surface, such as the ocean surface $\partial\Omega_{\text{out}}(\lambda \pm \Delta\lambda/2)$ or the open interior boundary $\partial\Omega_{\text{in}}(\lambda \pm \Delta\lambda/2)$. Figure 51.6 illustrates the case for the ocean surface. From this figure we write the integral of an arbitrary function Φ over the region of a surface with $\lambda_0 \leq \lambda$

$$\Sigma_\Phi(\lambda_0 \leq \lambda) = \int_{\partial\Omega(\lambda_0 \leq \lambda)} \Phi \, \mathrm{d}\mathcal{S}. \tag{51.24}$$

By construction the area integral of an arbitrary function within the layer $[\lambda - \Delta\lambda/2, \lambda + \Delta\lambda/2]$ is given by the difference

$$\Delta\Sigma_\Phi(\lambda \pm \Delta\lambda/2) = \Sigma_\Phi(\lambda_0 \leq \lambda + \Delta\lambda/2) - \Sigma_\Phi(\lambda_0 \leq \lambda - \Delta\lambda/2) = \int_{\partial\Omega(\lambda \pm \Delta\lambda/2)} \Phi \, \mathrm{d}\mathcal{S}, \tag{51.25}$$

with the reference value, λ_0, dropping out. We make use of this result in Section 51.3.6 when developing the layer mass budget.

Figure 51.6: *A closed ocean basin and the area between two λ-interfaces along the ocean surface where the λ-interfaces outcrop. The reference value λ_0 is arbitrary, and it drops out from the area integral within the band, $[\lambda - \Delta\lambda/2, \lambda + \Delta\lambda/2]$, as shown by equation (51.25).*

51.3.6 The layer mass budget

Bringing the above pieces together leads to the mass budget for a layer

$$\frac{\mathrm{d}\Delta M}{\mathrm{d}t} + \Delta\Psi = \Delta W - \Delta G, \tag{51.26}$$

where

$$\Delta M(\lambda \pm \Delta\lambda/2) = \int_{\Omega(\lambda \pm \Delta\lambda/2)} \rho \, \mathrm{d}V \tag{51.27}$$

is the layer mass, and for brevity we dropped $\lambda \pm \Delta\lambda/2$ arguments on all terms. It is common to define the layer mass *formation* as the mass accumulation within the layer

$$\Delta F \equiv \frac{\mathrm{d}\Delta M}{\mathrm{d}t} + \Delta\Psi = \Delta W - \Delta G. \tag{51.28}$$

The first equality indicates that water mass formation occurs in the layer if there is a time change in the mass within the layer (sometimes referred to as the *storage term*), as well as net mass crossing the interior open boundary. The second equality indicates that water mass formation occurs if there is mass crossing the ocean surface and/or mass converging through transformation across interior layer interfaces.

We arrive at a differential equation for the mass budget (51.28) by dividing through by the layer increment, $\Delta\lambda$, and taking the limit as this increment tends to zero

$$\frac{\partial F}{\partial \lambda} = \frac{\partial}{\partial t}\left[\frac{\partial M}{\partial \lambda}\right] + \frac{\partial \Psi}{\partial \lambda} = \frac{\partial W}{\partial \lambda} - \frac{\partial G}{\partial \lambda}. \tag{51.29}$$

Interchanging the λ and time derivatives and integrating from a reference value λ_0 to λ leads to

$$\int_{\lambda_0}^{\lambda} \frac{\partial \Psi}{\partial \lambda}\, \mathrm{d}\lambda = \int_{\lambda_0}^{\lambda} \frac{\partial}{\partial \lambda}\left[-\frac{\partial M}{\partial t} + W - G\right] \mathrm{d}\lambda \Longrightarrow \Psi = -\frac{\partial M}{\partial t} + W - G. \tag{51.30}$$

We dropped the contribution from the reference value at the lower limit since we can, without loss of generality, assume that it sits outside of the ocean domain. The differential water mass equation (51.30) is a continuous version of the discrete equation (51.28).

Figure 51.7: *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 (e.g., buoyancy fluxes in the case of $\lambda = \gamma$). Movement of the $\lambda - \Delta\lambda/2$-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 $\partial\Omega_{in}(\lambda \pm \Delta\lambda/2)$ moves in the opposite direction 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 water mass transformation leaves the layer through circulation at the boundary $\partial\Omega_{in}(\lambda \pm \Delta\lambda/2)$.*

51.3.7 Circulation in λ-space

The layer mass budget (51.28) and its continuous expression (51.30) provide the basis for inferences about circulation and transformation. As an illustration, consider the continuous expression (51.30)

and integrate from the reference value up to λ

$$\int_{\lambda_0}^{\lambda} \Psi \, \mathrm{d}\lambda' = \int_{\lambda_0}^{\lambda} \left[-\frac{\partial M}{\partial t} + W - G \right] \mathrm{d}\lambda'. \tag{51.31}$$

The left hand side is an expression for the circulation in λ-space. We thus see 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

$$\int_{\lambda_0}^{\lambda} \Psi \, \mathrm{d}\lambda' = -\int_{\lambda_0}^{\lambda} G \, \mathrm{d}\lambda'. \tag{51.32}$$

We depict an example in Figure 51.7 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 $\partial\Omega_{\mathrm{in}}(\lambda \pm \Delta\lambda/2)$. Figure 51.8 shows another case, 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 the key power of layer water mass analysis.

Figure 51.8: *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 $\gamma_{1/2} \leq \gamma \leq \gamma_{3/2}$).*

51.3.8 Further study

Walin (1977) and *Walin* (1982) developed a layer formalism for studying water mass transformation and the associated circulation. *Marshall et al.* (1999) introduced the formalism surrounding

equation (51.17), thus providing a mathematical framework for the practice of water mass analysis. A somewhat pedantic point of note is that *Marshall et al.* (1999) refer to equation (51.17) as a *generalized Leibniz rule.* Instead, it is a direct result of the fundamental theorem of calculus.

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 papers.

51.4 Tracer mass analysis

Consider a tracer concentration, C, localized in a region within a layer of buoyancy (commonly approximated by potential density) as depicted in Figure 51.9. The upper panels to this figure illustrate the tracer patch in geographic/depth space along with isolines of buoyancy, γ, 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 49.3, 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.

Figure 51.9: *This figure depicts 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 49.3). 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 the motion of the buoyancy surfaces need not coincide with motion of the tracer patch.*

51.4.1 General form of the mass budget

As depicted in Figure 51.9 for buoyancy layers, and Figure 51.10 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 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 15.3.4 to a λ-layer

$$\frac{\mathrm{d}}{\mathrm{d}t}\Delta M_C(\lambda \pm \Delta\lambda/2) = \int_{\Omega(\lambda\pm\Delta\lambda/2)} \rho\,\dot{C}\,\mathrm{d}V - \oint_{\partial\Omega(\lambda\pm\Delta\lambda/2)} \rho\,C\,(\boldsymbol{v} - \boldsymbol{v}^{(b)})\cdot\hat{\boldsymbol{n}}\,\mathrm{d}\mathcal{S}, \tag{51.33}$$

where

$$\Delta M_C(\lambda \pm \Delta\lambda/2) = \int_{\Omega(\lambda\pm\Delta\lambda/2)} \rho\,C\mathrm{d}V \tag{51.34}$$

is the mass of tracer within the layer. The volume integral on the right hand side of equation (51.26) arises from material time changes to the tracer within the layer, whereas the surface integral arises from dia-surface transport across the layer boundary.

51.4.2 Tracer processes

We determine the material time changes for a "conservative" tracer according to the convergence of a flux

$$\rho\,\dot{C} = \rho\,\frac{\mathrm{D}C}{\mathrm{D}t} = -\nabla\cdot\boldsymbol{J}. \tag{51.35}$$

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, $-\nabla\cdot\boldsymbol{J}$, 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 $\boldsymbol{J}$

$$\boldsymbol{J}\cdot\hat{\boldsymbol{n}}\,\mathrm{d}\mathcal{S} = \text{boundary tracer transport}. \tag{51.36}$$

Note that when there is an advective/skew diffusive component to the subgrid scale flux (Chapters 48 and 49), then it adds to the resolved advective component to render a residual mean material time operator

$$\rho\,\frac{\mathrm{D}^{\dagger}C}{\mathrm{D}t} = -\nabla\cdot\boldsymbol{J}_{\text{non-advect}}, \tag{51.37}$$

where

$$\frac{\mathrm{D}^{\dagger}}{\mathrm{D}t} = \frac{\partial}{\partial t} + (\boldsymbol{v} + \boldsymbol{v}^{*})\cdot\nabla, \tag{51.38}$$

with $\boldsymbol{v}^{*}$ an eddy-induced velocity (see Section 49.1). For the purposes of water mass transformation analysis, we write

$$\dot{C} = \frac{\mathrm{D}^{\dagger}C}{\mathrm{D}t}, \tag{51.39}$$

thus incorporating the eddy-induced stirring into the kinematic expression for the material time derivative.

There are many interior and boundary proceses that contribute to $\dot{C}$ within a layer. As a general expression for these contributions to the layer budget we write

$$\Delta E_C(\lambda \pm \Delta\lambda/2) = \int_{\Omega(\lambda\pm\Delta\lambda/2)} \rho\, \dot{C}\, \mathrm{d}V, \tag{51.40}$$

which is sometimes usefully decomposed into interior and boundary processes

$$\Delta E_C^{\text{interior}}(\lambda \pm \Delta\lambda/2) + \Delta E_C^{\text{boundary}}(\lambda \pm \Delta\lambda/2) = \int_{\Omega(\lambda\pm\Delta\lambda/2)} \rho\, (\dot{C}^{\text{interior}} + \dot{C}^{\text{boundary}})\, \mathrm{d}V, \tag{51.41}$$

where diffusion is the canonical mixing process contributing to $\dot{C}^{\text{interior}}$.

Figure 51.10: *As for the schematic of a layer seawater mass budget depicted in Figure 51.3, 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 domain $\Omega(\lambda \pm \Delta\lambda/2)$. 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 (51.41)). This term has no associated transfer of seawater mass and thus is absent from the water mass budget in Figure 51.3. 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 (51.42)); transport across the surface domain boundary, ΔW_C, arising from precipitation, evaporation, runoff, and melt (equation (51.43)); and transport within the circulation crossing an interior domain boundary, $\Delta\Psi_C$ (equation (51.44)).*

51.4.3 Transport across an interior layer interface

The surface integral in the budget (51.33) involves transport across the layer interfaces, with this transport requiring motion of the interface relative to a fluid particle. The same formalism introduced in Section 51.3 can be used to compute this transport. That is, we can generalize the transformation equation (51.19) to write

$$G_C(\lambda) = \int_{\partial\Omega(\lambda)} \rho\, C\, (\boldsymbol{v} - \boldsymbol{v}^{(b)}) \cdot \hat{\boldsymbol{n}}\, \mathrm{d}\mathcal{S} = \frac{\partial}{\partial\lambda} \int_{\Omega(\lambda_0 \le \lambda)} \rho\, \dot{\lambda}\, C\, \mathrm{d}V. \tag{51.42}$$

As a sanity 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 51.5 when studying budgets over regions bounded by a tracer isosurface.

51.4.4 Transport across interior and surface boundaries

We follow the discussion for water mass analysis in Section 51.3.4 to 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 (15.53)

$$\Delta W_C = \int_{\partial\Omega_{\rm out}(\lambda\pm\Delta\lambda/2)} \rho\, C\, (\boldsymbol{v} - \boldsymbol{v}^{(b)}) \cdot \hat{\boldsymbol{n}}\, \mathrm{d}\mathcal{S} = \int_{\partial\Omega_{\rm out}(\lambda\pm\Delta\lambda/2)} Q_{\rm m}\, C_{\rm m}\, \mathrm{d}A, \tag{51.43}$$

where $C_{\rm m}$ is the tracer concentration within the mass transported across the boundary.[1] As a sanity check, note that in the special case of a constant tracer concentration in the mass transported across the boundary, then $\Delta W_C = C_{\rm m}\, \Delta W$, where ΔW is the water mass transported across the ocean free surface as given by equation (51.22).

For the interior open boundary the contribution is written in the generic manner

$$\Delta \Psi_C = \int_{\partial\Omega_{\rm in}(\lambda\pm\Delta\lambda/2)} C\, \rho\, (\boldsymbol{v} - \boldsymbol{v}^{(b)}) \cdot \hat{\boldsymbol{n}}\, \mathrm{d}\mathcal{S}. \tag{51.44}$$

Again, in the special case where the tracer concentration is a constant along the interior boundary, $C_{\rm b}$, then $\Delta\Psi_C = C_{\rm b}\, \Delta\Psi$, where $\Delta\Psi$ is the seawater mass transport given by equation (51.20).

51.4.5 The layer tracer budget

Bringing all terms together leads to the layer tracer mass budget

$$\frac{\mathrm{d}\Delta M_C}{\mathrm{d}t} + \Delta\Psi_C = \Delta E_C + \Delta W_C - [G_C(\lambda + \Delta\lambda/2) - G_C(\lambda - \Delta\lambda/2)], \tag{51.45}$$

which is directly analogous to the seawater layer mass budget (51.26), with the added term ΔE_C arising from material tracer changes. As for the seawater mass budget discussed in Section 51.3.6, the layer tracer budget (51.45) provides the framework for rather general inferences about tracer transport within λ-classes.

51.4.6 Further study

As in Section 51.3, 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.

51.5 Regions bounded by a tracer contour/surface

In Section 51.4 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. We will find some rather useful simplifications arise as a result of this assumption.

[1] Note that equation (26) in *Groeskamp et al.* (2019) incorrectly writes the integrand in equation (51.43) as $Q_{\rm m}\,(C_{\rm m} - C)$.

Figure 51.11: *Left panel: a closed three-dimensional region, $\widetilde{\mathcal{R}}$, with its boundary, $\partial\widetilde{\mathcal{R}}$, defined by a surface of constant tracer concentration, $C = \widetilde{C}$. Right panel: the analog closed two-dimensional region with its boundary defined by a contour of constant tracer concentration, $C = \widetilde{C}$.*

As in Section 51.4, our starting point is the Leibniz-Reynolds budgets from Section 15.3.4 for seawater mass and tracer mass computed over an arbitrary region, $\mathcal{R}$

$$\frac{\mathrm{d}}{\mathrm{d}t}\left[\int_{\mathcal{R}} \rho\, C\, \mathrm{d}V\right] = -\int_{\partial\mathcal{R}} \left[\rho\, C\, (\boldsymbol{v} - \boldsymbol{v}^{(b)}) + \boldsymbol{J}\right] \cdot \hat{\boldsymbol{n}}\, \mathrm{d}\mathcal{S} \tag{51.46a}$$

$$\frac{\mathrm{d}}{\mathrm{d}t}\left[\int_{\mathcal{R}} \rho\, \mathrm{d}V\right] = -\int_{\partial\mathcal{R}} \left[\rho\, (\boldsymbol{v} - \boldsymbol{v}^{(b)})\right] \cdot \hat{\boldsymbol{n}}\, \mathrm{d}\mathcal{S}. \tag{51.46b}$$

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

$$M = \int_{\mathcal{R}} \rho\, \mathrm{d}V \qquad \text{region seawater mass} \tag{51.47a}$$

$$M_C = \int_{\mathcal{R}} C\, \rho\, \mathrm{d}V \qquad \text{region tracer mass} \tag{51.47b}$$

$$\langle C \rangle = \frac{1}{M}\int_{\mathcal{R}} C\, \rho\, \mathrm{d}V = \frac{M_C}{M} \qquad \text{region averaged tracer concentration.} \tag{51.47c}$$

51.5.1 Closed region bounded by a tracer surface/contour

Consider a closed region, $\widetilde{R}$, bounded by a surface of constant tracer concentration, $C = \widetilde{C}$, such as depicted in Figure 51.11. The tracer budget (51.46a) for this region is given by

$$\frac{\mathrm{d}(M\, \langle C \rangle)}{\mathrm{d}t} = -\widetilde{C} \int_{\partial\widetilde{\mathcal{R}}} \rho\, (\boldsymbol{v} - \boldsymbol{v}^{(b)}) \cdot \hat{\boldsymbol{n}}\, \mathrm{d}\mathcal{S} - \int_{\partial\widetilde{\mathcal{R}}} \boldsymbol{J} \cdot \hat{\boldsymbol{n}}\, \mathrm{d}\mathcal{S} \tag{51.48}$$

where we pulled the tracer concentration outside of the surface integral since, by construction, it is constant on the boundary, $\partial\widetilde{\mathcal{R}}$. Use of the mass budget (51.46b) then leads to the rather tidy result

$$\frac{\mathrm{d}[M\, (\langle C \rangle - \widetilde{C})]}{\mathrm{d}t} = -\int_{\partial\widetilde{\mathcal{R}}} \boldsymbol{J} \cdot \hat{\boldsymbol{n}}\, \mathrm{d}\mathcal{S}. \tag{51.49}$$

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,

Figure 51.12: *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.*

$\widetilde{C}$. 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 (51.49) reaches a steady state whereby

$$\frac{\mathrm{d}}{\mathrm{d}t}\left[M\left(\langle C\rangle - \widetilde{C}\right)\right] = 0 \Longleftrightarrow \int_{\partial\widetilde{\mathcal{R}}} \boldsymbol{J}\cdot\hat{\boldsymbol{n}}\,\mathrm{d}\mathcal{S} = 0. \tag{51.50}$$

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 51.12, 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 51.12. In this figure, the tracer contour defining the region boundary is a thick shell defined by a uniform concentration $C = \widetilde{C}$. The diffusive flux vanishes at each point within the boundary shell since the tracer concentration is uniform. Hence, the steady budget (51.50) leads to

$$\left(\langle C\rangle - \widetilde{C}\right)\frac{\mathrm{d}M}{\mathrm{d}t} + M\,\frac{\mathrm{d}\langle C\rangle}{\mathrm{d}t} = 0. \tag{51.51}$$

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 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

$$C = \tilde{C} \qquad \text{steady state tracer throughout a closed tracer region.} \tag{51.52}$$

Diffusion thus expells tracer gradients from steady state regions bounded by closed tracer contours, thus leaving a homogenous interior.

51.5.2 Region with $C \geq \tilde{C}$

As a second application 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 51.13. For this domain, there is no inner boundary like that shown in Figure 51.10. We introduce the following terms in developing both the seawater mass budget and the tracer substance budget

$$\tilde{M} = \int_{C \geq \tilde{C}} \rho \, \mathrm{d}V \qquad \text{seawater mass in the region } C \geq \tilde{C} \tag{51.53a}$$

$$\tilde{M}_{\tilde{C}} = \int_{C \geq \tilde{C}} C \, \rho \, \mathrm{d}V = \tilde{M} \, \langle C \rangle \qquad \text{tracer mass in the region } C \geq \tilde{C} \tag{51.53b}$$

$$\tilde{G} = -\int_{C=\tilde{C}} \rho \, (\boldsymbol{v} - \boldsymbol{v}^{(b)}) \cdot \hat{\boldsymbol{n}} \, \mathrm{d}\mathcal{S} = -\int_{C=\tilde{C}} \rho \, w^{\mathrm{dia}} \, \mathrm{d}\mathcal{S} \qquad \text{watermass transformation across } \tilde{C} \tag{51.53c}$$

$$\tilde{G}_{\tilde{C}} = -\int_{C=\tilde{C}} C \, \rho \, (\boldsymbol{v} - \boldsymbol{v}^{(b)}) \cdot \hat{\boldsymbol{n}} \, \mathrm{d}\mathcal{S} = -\int_{C=\tilde{C}} C \, \rho \, w^{\mathrm{dia}} \, \mathrm{d}\mathcal{S} \qquad \text{tracer advected across } \tilde{C} \tag{51.53d}$$

$$\tilde{W} = -\int_{\partial\Omega_{\mathrm{surf}}(\tilde{C})} \rho \, (\boldsymbol{v} - \boldsymbol{v}^{(b)}) \cdot \hat{\boldsymbol{n}} \, \mathrm{d}\mathcal{S} = \int_{\partial\Omega_{\mathrm{surf}}(\tilde{C})} Q_{\mathrm{m}} \, \mathrm{d}A \qquad \text{surface mass transport} \tag{51.53e}$$

$$\tilde{W}_{\tilde{C}} = -\int_{\partial\Omega_{\mathrm{surf}}(\tilde{C})} C \, \rho \, (\boldsymbol{v} - \boldsymbol{v}^{(b)}) \cdot \hat{\boldsymbol{n}} \, \mathrm{d}\mathcal{S} = \int_{\partial\Omega_{\mathrm{surf}}(\tilde{C})} C \, Q_{\mathrm{m}} \, \mathrm{d}A \qquad \text{surface mass } C\text{-transport} \tag{51.53f}$$

$$(E_{\tilde{C}})_{\mathrm{interior}} = -\int_{C=\tilde{C}} \boldsymbol{J} \cdot \hat{\boldsymbol{n}} \, \mathrm{d}\mathcal{S} \qquad \text{subgrid } C\text{-transport across } \tilde{C} \tag{51.53g}$$

$$(E_{\tilde{C}})_{\mathrm{surf}} = -\int_{\partial\Omega_{\mathrm{surf}}(\tilde{C})} \boldsymbol{J} \cdot \hat{\boldsymbol{n}} \, \mathrm{d}\mathcal{S} \qquad \text{subgrid } C\text{-transport across surface} \tag{51.53h}$$

$$(E_{\tilde{C}})_{\mathrm{bot}} = -\int_{\partial\Omega_{\mathrm{bot}}(\tilde{C})} \boldsymbol{J} \cdot \hat{\boldsymbol{n}} \, \mathrm{d}\mathcal{S} \qquad \text{subgrid } C\text{-transport across bottom} \tag{51.53i}$$

Recall that $\hat{\boldsymbol{n}}$ is the *outward* normal on a boundary so that positive values for the above transports increase the tracer mass within the region. For equations (51.53c) and (51.53d), we introduced the dia-surface transport velocity according to equation (51.5) for flow across the $C = \tilde{C}$ layer interface. Likewise, for equations (51.53e) and (51.53f) we made use of the surface kinematic boundary condition (14.70c)

$$\rho \, (\boldsymbol{v} - \boldsymbol{v}^{(s)}) \cdot \hat{\boldsymbol{n}} \, \mathrm{d}\mathcal{S} = -Q_{\mathrm{m}} \, \mathrm{d}A, \tag{51.54}$$

where Q_{m} is the mass transport across the free surface, with $Q_{\mathrm{m}} > 0$ adding mass to the ocean, and $\mathrm{d}A$ is the horizontal projection of the surface area element. By inspection of Figure 51.13, the seawater mass and tracer mass budgets for this region are given by

$$\frac{\mathrm{d}\tilde{M}}{\mathrm{d}t} = \tilde{G} + \tilde{W} \tag{51.55a}$$

$$\frac{\mathrm{d}[\tilde{M} \, \langle C \rangle]}{\mathrm{d}t} = \tilde{C} \, \tilde{G} + \tilde{W}_{\tilde{C}} + (E_{\tilde{C}})_{\mathrm{surf}} + (E_{\tilde{C}})_{\mathrm{bot}} + (E_{\tilde{C}})_{\mathrm{interior}}, \tag{51.55b}$$

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 $\tilde{G}_{\tilde{C}} = \tilde{C} \, \tilde{G}$.

Just as we did in Section 51.5.1, the tracer budget (51.55b) can be simplified by making use of the seawater mass budget (51.55a) to eliminate the water mass transformation contribution $\tilde{G}$, thus rendering

$$\frac{\mathrm{d}\tilde{M}_{I\tilde{C}}}{\mathrm{d}t} = [\tilde{W}_{\tilde{C}} - \tilde{W} \, \tilde{C}] + (E_{\tilde{C}})_{\mathrm{surf}} + (E_{\tilde{C}})_{\mathrm{bot}} + (E_{\tilde{C}})_{\mathrm{interior}}. \tag{51.56}$$

In this equation we introduced the *internal tracer mass* according to

$$\tilde{M}_{I\tilde{C}} \equiv \tilde{M} \, (\langle C \rangle - \tilde{C}) = \int_{C \geq \tilde{C}} (C - \tilde{C}) \, \rho \, \mathrm{d}V. \tag{51.57}$$

Figure 51.13: *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 and surface boundary, $(E_{\tilde{C}})_{\text{interior}}$, advection at the surface, $\tilde{W}_{\tilde{C}}$, and the tracer weighted water mass transformation across the interior layer boundary, $\tilde{C}\,\tilde{G}$.*

For completeness we express the internal tracer mass budget (51.56) in its integral form

$$\frac{\mathrm{d}}{\mathrm{d}t}\int_{C\geq\tilde{C}}(C-\tilde{C})\,\rho\,\mathrm{d}V = \int_{\partial\Omega_{\text{out}}}\Big[Q_{\text{m}}\,(C-\tilde{C})\,\mathrm{d}A - \boldsymbol{J}\cdot\hat{\boldsymbol{n}}\,\mathrm{d}\mathcal{S}\Big] - \int_{\partial\Omega_{\text{bot}}}\boldsymbol{J}\cdot\hat{\boldsymbol{n}}\,\mathrm{d}\mathcal{S} - \int_{C=\tilde{C}}\boldsymbol{J}\cdot\hat{\boldsymbol{n}}\,\mathrm{d}\mathcal{S}. \tag{51.58}$$

Elimination of the watermass transformation, $\tilde{G}$, from the internal tracer mass budget equations (51.56) and (51.58) 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 (51.56) for internal tracer mass, rather than the budget (51.55b) for the total tracer mass. We further discuss the internal tracer mass, and its complement the external tracer mass, in Section 51.6.

51.5.3 Further reading

The two-dimensional version of the result (51.52) 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 47.9. It is satisfying to see their result follow from the present formalism based on Leibniz-Reynolds.

Material in Section 51.5.2 was motivated by the study of ocean heat budgets from *Holmes et al.* (2019). We extend that discussion in Section 51.6 by working with tracer mass distributions.

51.6 Tracer mass distributions

The seawater mass and tracer mass in a region with $C \geq \tilde{C}$ were written in equations (51.53a) and (51.53b) as integrals over geographic/depth space for regions where $C \geq \tilde{C}$. Writing $\rho\,\mathrm{d}V = \mathrm{d}M$

for the mass element renders

$$\tilde{M} = \int_{C \geq \tilde{C}} \rho \, \mathrm{d}V = \int_{C \geq \tilde{C}} \mathrm{d}M \tag{51.59a}$$

$$\tilde{M}_{\tilde{C}} = \int_{C \geq \tilde{C}} C \, \rho \, \mathrm{d}V = \int_{C \geq \tilde{C}} C \, \mathrm{d}M. \tag{51.59b}$$

In this section we develop a complementary perspective for these expressions, and their corresponding mass budgets, by working with tracer mass distributions and the corresponding tracer mass density function. In this description we dispense with information concerning the position of a fluid element in geographic/depth space. Instead, we assume knowledge of the mass of seawater and mass of tracer within a tracer concentration range, $[C - \mathrm{d}C/2, C + \mathrm{d}C/2]$, or bin, without concern for where in the ocean the seawater lives. Extension to Conservative Temperature, Θ, allows us to consider the enthalpy content as the direct analog to tracer mass. The formalism developed here has a one-to-one relation to that used to study probability for continuous random variables.

51.6.1 Tracer mass density function

Let us write the mass of tracer within the infinitesimal tracer concentration range $[C - \mathrm{d}C/2, C + \mathrm{d}C/2]$ according to

$$\mathrm{d}M = M[C + \mathrm{d}C/2] - M[C - \mathrm{d}C/2] = \frac{\mathrm{d}M}{\mathrm{d}C} \, \mathrm{d}C \equiv m(C) \, \mathrm{d}C, \tag{51.60}$$

where we dropped higher order terms since we assume the increments are infinitesimal. This expression introduced the tracer mass density function, $m(C)$, which is a measure of the mass of seawater with tracer concentration C, per increment of tracer concentration. Figure 51.14 provides an example tracer mass density function.

51.6.2 Seawater mass and tracer mass

Use of the tracer mass density function allows us to write the seawater mass and tracer mass with tracer concentration $C \geq \tilde{C}$ in the form

$$\tilde{M} = \int_{C \geq \tilde{C}} \mathrm{d}M = \int_{\tilde{C}}^{C_\infty} m(C) \, \mathrm{d}C \tag{51.61a}$$

$$\tilde{M}_{\tilde{C}} = \int_{C \geq \tilde{C}} C \, \mathrm{d}M = \int_{\tilde{C}}^{C_\infty} m(C) \, C \, \mathrm{d}C. \tag{51.61b}$$

The upper limit C_∞ is a shorthand for an arbitrary value of tracer concentration that is larger than the maximum in the ocean. Notably, the tracer mass density function is zero when evaluated at C_∞ and larger

$$m(C) = 0 \qquad \text{if } C \geq C_\infty. \tag{51.62}$$

In a manner analogous to cumulative probability distribution functions encountered in probability theory, $\tilde{M}$ and $\tilde{M}_{\tilde{C}}$ represent cumulative mass distribution functions for seawater mass and tracer mass within the region with $C \geq \tilde{C}$. Leibniz's rule provides the relation

$$\frac{\mathrm{d}\tilde{M}}{\mathrm{d}\tilde{C}} = -m(\tilde{C}), \tag{51.63}$$

Figure 51.14: *An example tracer mass density function, $m(C) = \mathrm{d}M/\mathrm{d}C$, which measures the mass of tracer with concentration C, per increment in tracer concentration. Integration of the density function from the left up to a concentration $\tilde{C}$ leads to the mass of seawater and mass of tracer with tracer concentration less than or equal to $\tilde{C}$ (see equations (51.61a) and (51.61b)). The complement integration from $\tilde{C}$ above leads to the complement mass of seawater.*

which then renders the alternative expressions for the cumulative mass distribution functions

$$\tilde{M} = \int_{\tilde{C}}^{C_\infty} \left[-\frac{\mathrm{d}\tilde{M}}{\mathrm{d}C} \right] \mathrm{d}C = -\int_{C \geq \tilde{C}} \mathrm{d}\tilde{M} \tag{51.64a}$$

$$\tilde{M}_{\tilde{C}} = \int_{\tilde{C}}^{C_\infty} \left[-\frac{\mathrm{d}\tilde{M}}{\mathrm{d}C} \right] C\,\mathrm{d}C = -\int_{C \geq \tilde{C}} C\,\mathrm{d}\tilde{M}. \tag{51.64b}$$

Although the final expressions are seemingly quite trivial, they offer useful relations for working with tracer distributions in tracer space. Note that the minus sign arises merely since we define $\tilde{M}$ and $\tilde{M}_{\tilde{C}}$ as the mass for the region with tracer concentration *greater than or equal to* $\tilde{C}$, whereas there is no minus sign if working in the complement region.

As stated earlier, all information about geographical position and depth are absent when working with the tracer mass distributions and the corresponding tracer mass density function. Instead, we only have information about the tracer mass as a function of tracer concentration, with this information estimated in practice via a binning approach. Consequently, the tracer integrals in equation (51.64b) are one-dimensional in tracer space. For many purposes, it is natural to work with the tracer distributions, particularly when the region over which tracer concentration satisfies $C \geq \tilde{C}$ is complex, and possibly disconnected, when viewed in geographic/depth space.

51.6.3 Internal and external tracer masses

An integration by parts on the expression (51.64b) for tracer mass leads to the two-part decomposition

$$\tilde{M}_{\tilde{C}} = -\int_{C\geq\tilde{C}} C\,\mathrm{d}\tilde{M} = -\int_{\tilde{C}}^{C_\infty}\Big[\mathrm{d}[C\,\tilde{M}] - \tilde{M}\,\mathrm{d}C\Big] = \underbrace{\tilde{C}\,\tilde{M}}_{\text{external tracer mass}} + \underbrace{\int_{\tilde{C}}^{C_\infty}\tilde{M}\,\mathrm{d}C}_{\text{internal tracer mass}}. \tag{51.65}$$

Note that we dropped the $C = C_\infty$ term since there is zero mass for tracer concentrations at and above this value (see equation (51.62)).

Equation (51.65) decomposes the tracer mass into two terms with distinct properties. The first term is the *external tracer mass*

$$\tilde{M}_{E\tilde{C}} \equiv \tilde{C}\,\tilde{M}, \tag{51.66}$$

which is the product of the layer mass, $\tilde{M}$, and the interface tracer concentration, $\tilde{C}$. The only tracer dependence contained in the external tracer mass comes through the tracer concentration, $\tilde{C}$, of the interface defining the layer boundary. We choose the moniker "external" since the external tracer mass increases in direct proportion to the mass crossing the ocean layer boundaries, including the external boundaries, regardless the amount of tracer contained in that mass. The second term in equation (51.65) is the *internal tracer mass* defined by equation (51.57). Making use of tracer distributions it takes the form

$$\tilde{M}_{I\tilde{C}} \equiv \tilde{M}_{\tilde{C}} - \tilde{M}_{E\tilde{C}} = \tilde{M}\,[\langle C\rangle - \tilde{C}] = \int_{\tilde{C}}^{C_\infty}\tilde{M}\,\mathrm{d}C = \int_{C\geq\tilde{C}}(\tilde{C} - C)\,\mathrm{d}M. \tag{51.67}$$

To reach the second equality we set $\tilde{M}_{\tilde{C}} = \tilde{M}\,\langle C\rangle$, whereas the final equality arises from converting back to geographic/depth coordinates where $\mathrm{d}M = \rho\,\mathrm{d}V$ as per equations (51.59a) and (51.59b). In Section 51.5.2 we derived the budget equation (51.56) for the internal tracer mass, with this budget having the notable property that it is not dependent on the water mass transformation, $\tilde{G}$, crossing the $\tilde{C}$ layer interface.

51.6.4 Further reading

Material in this section was motivated by the study of the ocean internal heat budget in *Holmes et al.* (2019).

51.7 $\Theta - S$ water mass analysis

Bibliography

Abernathey, 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. [889, 893, 905]

Acheson, D., *Elementary Fluid Dynamics*, Oxford Applied Mathematics and Computing Science Series, Oxford, Oxford, 1990. [26, 615]

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. [685, 686, 687]

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. [885]

Andrews, D., and M. McIntyre, An exact theory of nonlinear waves on a Lagrangian-mean flow, *Journal of Fluid Mechanics*, *89*, 609–646, 1978a. [847, 856]

Andrews, D. G., and M. E. McIntyre, On wave action and its relatives, *Journal of Fluid Mechanics*, *89*, 647–664, 1978b. [847, 856]

Anstey, J., and L. Zanna, A deformation-based parametrization of ocean mesoscale eddy Reynolds stresses, *Ocean Modelling*, *112*, 99–111, 2017. [328]

Apel, J. R., *Principles of Ocean Physics*, *International Geophysics Series*, vol. 38, Academic Press, London, 1987. [150, 763]

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. [764]

Aris, R., *Vectors, Tensors and the Basic Equations of Fluid Mechanics*, Dover Publishing, New York, 1962. [5, 83, 191, 213, 215, 223, 229, 251, 256, 310]

Batchelor, G. K., *An Introduction to Fluid Dynamics*, Cambridge University Press, Cambridge, England, 1967. [189, 239, 400, 781]

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. [891]

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. [891]

Bennett, A., The geometry of neutral paths, *Journal of Physical Oceanography*, *49*, 3037–3044, doi:10.1175/JPO-D-19-0047.1, 2019. [367]

Bishop, S., R. Small, and F. Bryan, The global sink of available potential energy by mesoscale air-sea interaction, *Journal of Advances in Modeling the Earth System*, *in review*, 2020. [450]

Bleck, R., Finite difference equations in generalized vertical coordinates. Part I: Total energy conservation, *Contributions to Atmospheric Physics*, *51*, 360–372, 1978. [115]

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. [669, 672]

Bretherton, F. P., Critical layer instability in baroclinic flows, *Quarterly Journal of the Royal Meteorological Society*, *92*, 325–334, 1966. [445]

Brown, E., *Waves, tides, and shallow-water processes*, 227 pp., The Open University, Milton Keys, UK, 1999. [763]

Bryan, K., A numerical method for the study of the circulation of the world ocean, *Journal of Computational Physics*, *4*, 347–376, 1969. [700, 739]

Bühler, O., *Waves and mean flows*, 2nd ed., Cambridge University Press, Cambridge, UK, doi:10.1017/CBO9781107478701, 2014. [791, 847, 856]

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. [451]

Callen, H. B., *Thermodynamics and an Introduction to Thermostatics*, John Wiley and Sons, New York, 493 + xvi pp, 1985. [121, 369]

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. [741]

Chaikin, P. M., and T. C. Lubensky, *Principles of Condensed Matter Physics*, Cambridge University Press, Cambridge, United Kingdom, 1995. [65]

Chandrasekhar, S., *Hydrodynamic and Hydromagnetic Stability*, Dover Publications, New York, 654 pp, 1961. [437]

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. [645]

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. [905]

Courant, R., and D. Hilbert, *Methods of Mathematical Physics Volume I*, Wiley-Interscience, New York, 1953. [819]

Courant, R., and D. Hilbert, *Methods of Mathematical Physics Volume II: Partial Differential Equations*, Wiley-Interscience, 1962. [819]

Cox, M. D., Isopycnal diffusion in a z-coordinate ocean model, *Ocean Modelling*, *74*, 1–5, 1987. [891, 892]

Cushman-Roisin, B., and J.-M. Beckers, *Introduction to Geophysical Fluid Dynamics*, Academic-Press, Amsterdam, 828, 2011. [x, 504]

Cushmin-Roisin, B., Subduction, in *Dynamics of the oceanic surface mixed-layer*, pp. 181–196, Hawaii Institude of Geophysical Special Publications, 1987. [297]

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, 2003. [669]

Davis, R. E., Diapycnal mixing in the ocean: equations for large-scale budgets, *Journal of Physical Oceanography*, *24*, 777–800, 1994. [378, 400]

DeGroot, S. R., and P. Mazur, *Non-Equilibrium Thermodynamics*, Dover Publications, New York, 510 pp, 1984. [251, 256, 395, 400, 408]

DeSzoeke, R. A., and A. F. Bennett, Microstructure fluxes across density surfaces, *Journal of Physical Oceanography*, *23*, 2254–2264, 1993. [871]

Drazin, P., and R. Johnson, *Solitons: an introduction*, Cambridge University Press, Cambridge, UK, 226 pp, 1989. [790, 791]

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. [741]

Duchateau, P., and D. Zachmann, *Partial differential equations*, Schaum's Outline Series in Mathematics, McGraw-Hill, New York, 1986. [41, 60]

Dukowicz, J. K., and R. D. Smith, Stochastic theory of compressible turbulent fluid transport, *Physics of Fluids*, *9*, 3523–3529, 1997. [892]

Dunne, J. P., J. G. John, R. W. Hallberg, S. M. Griffies, E. N. Shevliakova, R. J. Stouffer, J. P. Krasting, L. A. Sentman, P. C. D. Milly, S. L. Malyshev, A. J. Adcroft, W. Cooke, K. A. Dunne, M. J. Harrison, H. Levy, B. L. Samuels, M. Spelman, M. Winton, A. T. Wittenberg, P. J. Phillips, and N. Zadeh, GFDL's ESM2 global coupled climate-carbon Earth System Models Part I: Physical formulation and baseline simulation characteristics, *Journal of Climate*, *25*, 6646––6665, 2012. [908]

Eckart, C., An analysis of the stirring and mixing processes in incompressible fluids, *Journal of Marine Research*, *7*, 265–275, 1948. [811, 905]

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. [813, 905]

Ertel, H., Ein neuer hydrodynamicscher wirbelsatz, *Meteorol. Z.*, *59*, 271–281, 1942. [647, 650]

Falkovich, G., *Fluid Mechanics: A short course for physicists*, Cambridge University Press, 167pp, 2011. [781]

Farrell, W., and J. Clark, On postglacial sea level, *Geophysical Journal of the Royal Astronomical Society*, *46*, 646–667, 1976. [763]

Feistel, R., Equilibrium thermodynamics of seawater revisited, *Progress in Oceanography*, *31*, 101–179, 1993. [381]

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. [893, 905]

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. [893]

Ferreira, D., and J. Marshall, Formulation and implementation of a residual-mean ocean circulation model, *Ocean Modelling*, *13*, 86–107, 2006. [887]

Fetter, A. L., and J. D. Walecka, *Theoretical Mechanics of Particles and Continua*, McGraw-Hill Book Company, New York, 570 pp, 1980. [780, 781, 791]

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. [37, 859]

Garrett, C., Turbulent dispersion in the ocean, *Progress in Oceanography*, *70*, 113–125, 2006. [858, 862]

Gent, P. R., and J. C. McWilliams, Isopycnal mixing in ocean circulation models, *Journal of Physical Oceanography*, *20*, 150–155, 1990. [477, 891, 892]

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, 1995. [458, 828, 871, 879, 883, 884, 885, 886, 887, 888, 889, 894]

Gill, A., *Atmosphere-Ocean Dynamics*, *International Geophysics Series*, vol. 30, Academic Press, London, 662 + xv pp, 1982. [x, 156, 158, 185, 329, 361, 412, 473, 484, 487, 764, 901, 905, 908]

Gnanadesikan, A., A global model of silicon cycling: Sensitivity to eddy parameterization and dissolution, *Global Biogeochemical Cycles*, *13*, 199–220, 1999. [891]

Goldstein, H., *Classical Mechanics*, Addison-Wesley Publishing, Reading, MA, 1980. [140]

Graham, F., and T. McDougall, Quantifying the nonconservative production of Conservative Temperature, potential temperature, and entropy, *Journal of Physical Oceanography*, *43*, 838–862, 2013. [878]

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. [271, 915]

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. [887]

Gregg, M. C., Entropy generation in the ocean by small-scale mixing, *Journal of Physical Oceanography*, *14*, 688–711, 1984. [400]

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. [310, 764, 915]

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 the Earth System*, p. in revision, 2020. [193, 299, 893]

Griffies, S. M., The Gent-McWilliams skew-flux, *Journal of Physical Oceanography*, *28*, 831–841, 1998. [826, 884, 892]

Griffies, S. M., *Fundamentals of Ocean Climate Models*, Princeton University Press, Princeton, USA, 518+xxxiv pages, 2004. [28, 83, 113, 189, 277, 311, 328, 684, 700, 819, 826, 847, 871, 881]

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. [437]

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. [248, 271, 847, 880, 883, 897, 915]

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, 1998. [819, 891, 892]

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. [686, 889]

Griffies, S. M., R. Pacanowski, M. Schmidt, and V. Balaji, Tracer conservation with an explicit free surface method for z-coordinate ocean models, *Monthly Weather Review*, *129*, 1081–1098, 2001. [908]

Griffies, S. M., A. Biastoch, C. W. Böning, F. Bryan, G. Danabasoglu, E. Chassignet, M. H. England, R. Gerdes, H. Haak, R. W. Hallberg, W. Hazeleger, J. Jungclaus, W. G. Large, G. Madec, A. Pirani, B. L. Samuels, M. Scheinert, A. S. Gupta, C. A. Severijns, H. L. Simmons, A. M. Treguier, M. Winton, S. Yeager, and J. Yin, Coordinated Ocean-ice Reference Experiments (COREs), *Ocean Modelling*, *26*, 1–46, doi:10.1016/j.ocemod.2008.08.007, 2009. [908]

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. Ilicak, 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. Voldoire, 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. [426, 520, 915]

Groeskamp, S., R. P. Abernathey, and A. Klocker, Water mass transformation by cabbeling and thermobaricity, *Geophysical Research Letters*, doi:10.1002/2016GL070860, 2016. [897]

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. [290, 897, 911, 921, 922, 926, 930, 933]

Hallberg, R., A thermobaric instability in Lagrangian vertical coordinate ocean models, *Ocean Modelling*, *8*, 227–300, 2005. [801]

Hallberg, R., and P. Rhines, Buoyancy-driven circulation in a ocean basin with isopycnals intersectin the sloping boundary, *Journal of Physical Oceanography*, *26*, 913–940, 1996. [680]

Haney, R. L., On the pressure gradient force over steep topography in sigma-coordinate ocean models, *Journal of Physical Oceanography*, *21*, 610–619, 1991. [686]

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. [666, 669]

Haynes, P. H., and M. E. McIntyre, On the conservation and impermeability theorems for potential vorticity, *Journal of Atmospheric Sciences*, *47*, 2021–2031, 1990. [669]

Hildebrand, F., *Advanced Calculus for Applications*, Prentice-Hall Publishers, Englewood Cliffs, New Jersey, 1976. [41, 53]

Holloway, G., Eddy transport of thickness and momentum in layer and level models, *Journal of Physical Oceanography*, *27*, 1153–1157, 1997. [889]

Holloway, G., and P. Rhines, Angular momenta of modeled ocean gyres, *Journal of Geophysical Research*, *27*, 843–846, 1991. [701]

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. [937, 940]

Holton, J. R., *An Introduction to Dynamic Meteorology*, Academic Press, San Diego, USA, 507 pp, 1992. [315, 316, 345, 352, 426, 479, 484, 487, 488, 489, 490, 491, 592, 594, 628]

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. [757, 758, 759]

Hoskins, B. J., and I. N. James, *Fluid Dynamics of the Midlatitude Atmosphere*, Wiley Blackwell, Chichester, UK, 2014. [228, 229]

Huang, K., *Statistical Mechanics*, John Wiley and Sons, New York, 493 pp, 1987. [181, 190, 369, 816]

Huang, R. X., Real freshwater flux as a natural boundary condition for the salinity balance and thermohaline circulation forced by evaporation and precipitation, *Journal of Physical Oceanography*, *23*, 2428–2446, 1993. [908]

Huang, R. X., and R. W. Schmitt, Goldsbrough-Stommel circulation of the World Oceans, *Journal of Physical Oceanography*, *23*, 1277–1284, 1993. [736]

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. [352]

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. [357, 359, 381, 400, 408, 878, 903, 907]

Jackson, J. D., *Classical Electrodynamics*, John Wiley and Sons, New York, USA, 848 pp, 1975. [827]

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. [604]

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, 2015. [700]

Killworth, P. D., On the parameterization of eddy transfer Part I: Theory, *Journal of Marine Research*, *55*, 1171–1197, 1997. [889]

Klocker, A., and R. Abernathey, Global patterns of mesoscale eddy properties and diffusivities, *Journal of Physical Oceanography*, *44*, 1030–1046, 2014. [905]

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. [897]

Kundu, P., I. Cohen, and D. Dowling, *Fluid Mechanics*, Academic Press, 921 + xxiv pp, 2012. [189, 212, 215, 326, 328, 781, 816]

Kushner, P. J., and I. M. Held, Potential vorticity thickness fluxes and wave-mean flow interaction, *Journal of Atmospheric Sciences*, *56*, 948–958, 1999. [871]

Landau, L. D., and E. M. Lifshitz, *Mechanics*, Pergamon Press, Oxford, UK, 170 pp, 1976. [163]

Landau, L. D., and E. M. Lifshitz, *Fluid Mechanics*, Pergamon Press, Oxford, UK, 539 pp, 1987. [395, 396, 400, 405, 408, 781]

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. [876, 905, 911]

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. [871]

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. [891]

Lilly, J., Kinematics of a fluid ellipse in a linear flow, *Fluids*, *3*(16), doi:10.3390/fluids3010016, 2018. [229]

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. [685, 686, 687]

Lorenz, E. N., Available potential energy and the maintenance of the general circulation, *Tellus*, *7*, 157–167, 1955. [444]

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. [403, 879, 883, 905]

Maddision, 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. [37]

Marion, J. B., and S. T. Thornton, *Classical Dynamics of Particles and Systems*, Harcourt Brace Jovanovich, San Diego, USA, 602 pp, 1988. [140, 586]

Markowski, P., and Y. Richardson, *Mesoscale Meteorology in Midlatitudes*, Wiley-Blackwell Publishers, Oxford, UK, 2010. [164]

Marshall, D., Vertical fluxes of potential vorticity and the structure of the thermocline, *Journal of Physical Oceanography*, *30*, 3102–3112, 2000. [659, 672]

Marshall, J., and R. A. Plumb, *Atmosphere, Ocean, and Climate Dynamics*, 1st ed., Academic Press, Amsterdam, 319pp, 2008. [315, 472, 495, 504, 517, 544, 548]

Marshall, J., D. Jamous, and J. Nilsson, Reconciling thermodynamic and dynamic methods of computation of water-mass transformation rates, *Deep-Sea Research I*, *46*, 545–572, 1999. [929, 930]

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. [659, 669, 671, 672, 680]

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. [842]

McDougall, T. J., Neutral surfaces, *Journal of Physical Oceanography*, *17*, 1950–1967, 1987a. [367, 905]

McDougall, T. J., Thermobaricity, cabbeling, and water-mass conversion, *Journal of Geophysical Research*, *92*, 5448–5464, 1987b. [367, 881, 882, 895, 896, 905]

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. [357, 400, 410, 411, 903]

McDougall, T. J., and R. Feistel, What causes the adiabatic lapse rate, *Deep-Sea Research*, *50*, 1523–1535, 2003. [386]

McDougall, T. J., and D. R. Jackett, On the helical nature of neutral trajectories in the ocean, *Progress in Oceanography*, *20*, 153–183, 1988. [367]

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. [367]

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. [835, 847, 864, 871, 876, 884]

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. [897]

McDougall, T. J., S. Groeskamp, and S. M. Griffies, On geometric aspects of interior ocean mixing, *Journal of Physical Oceanography*, *44*, 2164–2175, 2014. [367, 879, 881, 905]

McIntosh, P. C., and T. J. McDougall, Isopycnal averaging and the residual mean circulation., *Journal of Physical Oceanography*, *26*, 1655–1660, 1996. [871, 884]

McWilliams, J., *Fundamentals of Geophysical Fluid Dynamics*, Cambridge University Press, Cambridge, Cambridge, UK, 2006. [693, 700]

McWilliams, J. C., The emergence of isolated coherent vorticies in turbulent flow, *Journal of Fluid Mechanics*, *146*, 21–43, 1984. [693]

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. [686]

Mertz, G., and D. Wright, Interpretations of the JEBAR term, *Journal of Physical Oceanography*, *22*, 301–305, doi:10.1175/1520-0485(1992)022⟨0301:IOTJT⟩2.0.CO;2, 1992. [741]

Middleton, J. F., and J. W. Loder, Skew fluxes in polarized wave fields, *Journal of Physical Oceanography*, *19*, 68–76, 1989. [826, 858, 859, 860, 862, 894]

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. [763, 764]

Moffatt, H., Transport effects associated with turbulence with particular attention to the influence of helicity, *Reports on Progress in Physics*, *46*, 621–664, 1983. [826, 862]

Morse, P. M., and H. Feshbach, *Methods of Theoretical Physics Part I and II*, McGraw-Hill Book Company, New York, 1953. [37, 41, 60, 158]

Müller, P., Ertel's potential vorticity theorem in physical oceanography, *Reviews of Geophysics*, *33*, 67–97, 1995. [651]

Müller, P., *The Equations of Oceanic Motions*, 1st ed., Cambridge University Press, Cambridge, 302pp, 2006. [617]

Müller, P., and C. Garrett, From stirring to mixing in a stratified ocean, *Oceanography*, *15*, 12–19, 2002. [905]

Munk, W., and E. Palmén, Note on the dynamics of the Antarctic Circumpolar Current, *Tellus*, *3*, 53–55, 1951. [538, 541]

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. [905]

Munk, W. H., Abyssal recipes, *Deep-Sea Research*, *13*, 707–730, 1966. [905]

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. [905]

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. [680]

Nurser, A. G., and S. M. Griffies, Relating diffusive surface salinity fluxes to boundary freshwater and salt fluxes, *Journal of Physical Oceanography*, *in revision*, 2019. [903]

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. [871]

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. [871]

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, 1999. [444]

Olbers, D. J., J. Willebrand, and C. Eden, *Ocean Dynamics*, 1st ed., Springer, Berlin, Germany, 704 pages, 2012. [x, 189, 191, 194, 223, 367, 378, 408, 411, 878]

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. Maroztke, 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. [914]

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. [604]

Pedlosky, J., *Geophysical Fluid Dynamics*, 2nd ed., Springer-Verlag, Berlin Heidelberg New York, 710 + xv pp, 1987. [x, 709]

Pedlosky, J., L. Pratt, M. Spall, and K. Helfrich, Circulation around islands and ridges, *Journal of Marine Research*, *55*, 1199–1251, 1997. [599]

Plumb, R. A., Eddy fluxes of conserved quantities by small-amplitude waves, *Journal of Atmospheric Sciences*, *36*, 1699–1704, 1979. [862]

Polton, J., and D. Marshall, Overturning cells in the Southern Ocean and subtropical gyres, *Ocean Science*, *3*, 17–30, 2007. [659, 672]

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. [905]

Pugh, D. T., *Tides, surges, and mean sea-level*, 472 pp., John Wiley and Sons, 1987. [763]

Redi, M. H., Oceanic isopycnal mixing by coordinate rotation, *Journal of Physical Oceanography*, *12*, 1154–1158, 1982. [892]

Reichl, L. E., *A Modern Course in Statistical Physics*, John Wiley and Sons, New York, 822 pp, 1987. [369]

Reif, F., *Fundamentals of Statistical and Thermal Physics*, McGraw-Hill, New York, 1965. [181, 190, 369, 816]

Rhines, P. B., Basic dynamics of the large-scale geostrophic circulation, in *WHOI 1982 Summer Study Program*, Woods Hole Oceanographic Institute, 1982. [868]

Rhines, P. B., and W. R. Young, Homogenization of potential vorticity in planetary gyres, *Journal of Fluid Mechanics*, *122*, 347–367, 1982. [937]

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. [352]

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. [352]

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. [352]

Salmon, R., *Lectures on Geophysical Fluid Dynamics*, Oxford University Press, Oxford, England, 378 + xiii pp., 1998. [181, 189, 191, 326, 514, 587]

Samelson, R., *The Theory of Large-Scale Ocean Circulation*, Cambridge University Press, Cambridge, UK, 193 pp., 2011. [472, 735, 741]

Schär, C., A generalization of Bernoulli's Theorem, *Journal of the Atmospheric Sciences*, *50*, 1437–1443, doi:10.1029/2003JC001823, 1993. [659, 671, 672]

Schey, H., *Div, grad, curl and all that: an informal text on vector calculus*, W.W. Norton and Company, Inc., 176 pp, 2004. [19]

Schmitt, R. W., Double diffusion in oceanography, *Annual Review of Fluid Mechanics*, *26*, 255–285, 1994. [882, 905]

Schneider, E., D. DeWitt, A. Rosati, B. Kirtman, J. Link, and J. Tribbia, Retrospective enso forecasts: sensitivity to atmospheric model and ocean resolution, *Monthly Weather Review*, *131*, 3038–3060, 2003. [445]

Schutz, B. F., *A First Course in General Relativity*, Cambridge University Press, Cambridge, UK, 392 pp, 1985. [76]

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. [686]

Smagorinsky, J., General circulation experiments with the primitive equations: I. The basic experiment, *Monthly Weather Review*, *91*, 99–164, 1963. [418]

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. [328]

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. [889, 893]

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. [473]

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. [889]

Smith, R. D., The primitive equations in the stochastic theory of adiabatic stratified turbulence, *Journal of Physical Oceanography*, *29*, 1865–1880, 1999. [876]

Spiegel, M., *Theory and Problems of Vector Analysis*, Schaum's Outline Series, McGraw-Hill International Book Company, New York, 1974. [19]

Stakgold, I., *Boundary value problems of mathematical physics, volume I*, SIAM, Philadelphia, 340 pp, 2000a. [41, 57, 60]

Stakgold, I., *Boundary value problems of mathematical physics, volume II*, SIAM, Philadelphia, 408 pp, 2000b. [41, 57, 60]

Stanley, G. J., Neutral surface topology, *Ocean Modelling*, doi:10.1016/j.ocemod.2019.01.008, 2019. [367]

Starr, V. P., A quasi-Lagrangian system of hydrodynamical equations, *Journal of Meteorology*, *2*, 227–237, 1945. [115]

Stewart, R., *An Introduction to Physical Oceanography*, 345 pp., Texas A& M, College Station, Texas, USA, 2008. [763]

Straub, D. N., On thermobaric production of potential vorticity in the ocean, *Tellus*, *51A*, 314–325, 1999. [654]

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. [801]

Talley, L. D., G. L. Pickard, W. J. Emery, and J. H. Swift, *Descriptive Physical Oceanography*, 6th ed., Elsevier, 555pp, 2011. [519]

Taylor, G., Diffusion by continuous movements, *Proceedings of the London Mathematical Society*, *20*, 196–212, 1921. [813, 905]

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. [669, 680]

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. [680]

Thorne, K., and R. Blandford, *Modern Classical Physics*, Princeton University Press, Princeton, USA, 1511 + xl pp, 2017. [2, 83, 405, 408, 620]

Tomczak, M., and J. S. Godfrey, *Regional Oceanography: An Introduction*, Pergamon Press, Oxford, England, 422 + vii pp, 1994. [426, 519, 520, 599]

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. [889, 893, 905]

Truesdell, C., *The kinematics of vorticity*, Dover Publications, Mineola, New York, 1954. [259]

Vallis, G. K., *Atmospheric and Oceanic Fluid Dynamics: Fundamentals and Large-scale Circulation*, 1st ed., Cambridge University Press, Cambridge, 745 + xxv pp, 2006. [522, 548, 749, 751, 754]

Vallis, G. K., *Atmospheric and Oceanic Fluid Dynamics: Fundamentals and Large-scale Circulation*, 2nd ed., Cambridge University Press, Cambridge, 946 + xxv pp, 2017. [x, xix, xxii, 108, 158, 160, 189, 310, 329, 352, 359, 386, 408, 413, 426, 437, 443, 451, 456, 472, 477, 484, 504, 514, 519, 564, 567, 568, 593, 604, 608, 715, 724, 756, 816, 826, 839, 878, 879]

Vallis, G. K., *Essentials of Atmospheric and Oceanic Dynamics*, 1st ed., Cambridge University Press, Cambridge, 2019. [x, xix]

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. [491]

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 the Earth System*, doi:10.1029/2018ms001336, 2018. [911]

Veronis, G., Large scale ocean circulation, *Advances in Applied Mechanics*, *13*, 2–92, 1973. [158]

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. [929]

Walin, G., On the relation between sea-surface heat flow and thermal circulation in the ocean, *Tellus*, *34*, 187–195, 1982. [929]

Wallace, J., and P. Hobbs, *Atmospheric Science: An Introductory Survey*, Academic Press, 2006. [548]

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. [905]

Weiss, J., The dynamics of enstrophy transfer in two-dimensional hydrodynamics, *Physica D*, *273–294*, 1991. [229]

Welander, P., Studies of the general development of motion in a two-dimensional ideal fluid, *Tellus*, *7*, 141–156, 1955. [278]

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. [905]

Williams, J., and S. Elder, *Fluid Physics for Oceanographers and Physicists: an introduction to incompressible flow*, Pergamon Press, Oxford, 1989. [212]

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. [871]

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. [441, 442, 443]

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. [876, 887]

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. [887]

Index

absolute momentum, 164
acceleration, 137
    centrifugal, 148, 152
    centripetal, 148
    Coriolis, 148, 152
    inertial, 145, 150
    planetary Cartesian, 148
    spherical, 151
    spherical metric, 152
action/reaction law, 317
advection equation, 42
advection operator, 202
advection-diffusion equation, 263
advective tracer flux, 255
ageostrophic flow, 494
anelastic approximation, 270, 437
angular momentum, 138, 162, 171, 322
    axial, 169
    conservation, 172
anti-cyclonic, 165
arc length, 29
arc-length, 67, 272
area
    surface, 71
area evolution
    incompressible, 277
aspect ratio, 420, 453
available potential energy, 443
    approximate, 449
    exact, 448
Avogadro's number, 186
axial angular momentum, 148, 313
    atmosphere, 315
    ring of air, 314
    zonal acceleration, 175

baroclinicity
    generalized vertical coordinates, 689
barycenter, 254
barycentric velocity, 251, 253, 260, 319
basic one-form, 144
basis one-forms, 87
basis vector, 144
basis vectors, 81, 86
    changes, 143
Beltrami flow, 39
Bernoulli potential, 804
Bernoulli's Principle, 406
Bernoulli's theorem, 406
beta-plane approximation, 418
bi-orthogonality relation, 87
body forces, 317
boldface notation, 81
boundary velocity, 260
Boussinesq approximation
    generalized, 454
    ocean equations, 435
    traditional, 437
Boussinesq dynamic enthalpy, 442
Boussinesq reference density, 270
budget analysis, 257
buoyancy coordinates, 79
buoyancy work, 440

cabbeling, 435
capillary-gravity waves, 329
Cartesian coordinates, 102

summary, 103
Cauchy problem
hyperbolic, 53
parabolic, 50
Cauchy's stress principle, 318
center of mass, 254
center of mass motion, 248
centrifugal acceleration
particle motion, 165
planetary orbital, 769
planetary rotation, 769
spherical, 153
characteristic curves, 44
advection equation, 45
wave equation, 53
Christoffel symbols, 96
related to metric, 98
circuit, 30, 66
circulation, 30, 36, 576
solid-body rotation, 580
compatibility
total mass + tracer mass, 255
conservation equation
Eulerian flux-form, 257
material form, 257
conservation laws, 137, 163, 167
local, 267
conservation of labels, 81
conservative force, 139
conservative vector field, 39
constitent mass flux, 254
constitutive relation, 326
contact force
pressure, 320
contact forces, 317
contact stresses
local equilibrium, 319, 330
continuity, 269
continuity equation, 232
continuum hypothesis, 181, 182
contravariant tensor index, 78
coordinate representation
Cartesian, 6
Coriolis acceleration, 148, 172
large-scale motion, 153
shallow atmosphere, 153
spherical, 153
Coriolis parameter, 153
covariance, 76
covariant curl, 99
covariant derivative
metric tensor, 98
one-form, 97
scalar, 96
vector, 96
covariant differentiation, 77
covariant divergence
vector, 98
covariant Laplacian
scalar, 99
covariant tensor index, 78
Crocco's theorem, 413
curl of vector, 24
curl-free vector, 25
curvature
curve, 69
Gaussian, 72
surface, 71
curvature circle, 69
cyclonic, 165

D'Alembert wave solution, 54
deformation tensor, 198
determinant
exponential, 99
developing flow, 202
deviatoric stress, 326
dia-surface transport, 264
diagnostic equation, 205
diapycnal transport velocity, 246
diapycnal velocity component, 802
diffusion equation, 50
diffusive fluxes, 251, 260
diffusive tracer flux, 255
diffusivity
eddy, 437
Dirac delta, 58
Dirichlet boundary condition, 48
distance, 7
distance to polar axis, 148
divergence of curl, 26
divergence operator
Cartesian, 23
divergence theorem, 32
scalar fields, 33
double integral identity, 62

drag coefficient, 439
duality vectors & one-forms, 88
Duhamel's integral
    heat equation, 52
    wave equation, 56
dynamic topography, 425
dynamical constraints, 162, 163, 167

Eötvös correction, 153
earth radius, 764
earth's angular velocity, 140
earth's mass, 155
earth's radius, 140, 155
earth's solid body speed, 140
eddy diffusivity, 437
eddy viscosity, 438
eddy-induced velocity, 437, 458
Einstein index convention, 6
Einstein summation convention, 81
Ekman
    balance, 494
Ekman balance
    Ekman layer thickness, 500
    Ekman number, 499
    horizontal transport, 501
    Laplacian friction, 497
    mass transport, 501
    natural coordinates, 495
    Rayleigh drag, 495
    spiral motion, 497
    vertical transport, 502
elliptic PDE
    classification, 46
    properties, 46
energy
    total, 403
enthalpy
    budget, 399
equation of motion, 309
    Cartesian, 156
    particle, 156
    rotating, 310
    spherical, 157, 310
equation of state
    linear, 435
equatorial radius, 158
equilibrium tide, 765
Euclidean metric, 7
Euclidean space, 76
Eulerian coordinates
    Cartesian, 78
    cylindrical, 78
    spherical, 78
Eulerian flux-form, 267
Eulerian reference frame, 180, 195
Eulerian region, 258
Eulerian time derivative, 78, 202
exact differential, 35, 139
exact differentials, 273
Exner function, 427
extensive property, 257
external forces, 317

f-plane approximation, 165, 418
First law of thermodynamics
    moving fluid, 398
    potential temperature, 400
flow lines, 210
fluid element, 194, 252, 254
fluid material region, 194
fluid parcel, 193
    constituents, 252
fluid particle, 193
fluid particle trajectories, 196
fluid particle trajectory, 210
fluid region, 194
flux, 40
force couplets, 324
force potential, 139
forces
    body, 308
form stress, 320
Fourier's heat law, 264
free particle, 167
free vortex, 577, 581
friction
    vertical shear, 327
friction tensor, 326
frictional dissipation, 404
Froude number, 204, 453
functional degrees of freedom, 270
fundamental theorem of calculus, 30

Galilean boost, 137
Galilean invariance, 137, 140, 205, 326
    kinetic energy, 168

Galilean transformation, 137, 206, 257
    material time operator, 206
gauge symmetry, 273, 276
gauge transformation, 37
Gauss's divergence theorem, 32, 100, 684
Gaussian curvature, 72
general orthogonal coordinates, 113
general vertical coordinates
    basis one-forms, 119
    basis vectors, 118
    circulation, 131
    common confusion, 118
    contravariant velocity, 124
    covariant velocity, 124
    diffusion operator, 129
    divergence, 128
    divergence theorem, 128
    Jacobian, 125
    layer integrated diffusion, 130
    Levi-Civita tensor, 126
    material time derivative, 128
    metric tensor, 125
    partial derivatives, 126
    position vector, 121
    related to Cartesian, 117
    specific thickness, 125
    triple product identity, 120
    vector, 123
    vector cross product, 126
    velocity, 123
    volume element, 126
    vorticity, 130
generalized vertical coordinates, 197
geopotential, 156
    thickness, 424
geopotential coordinates, 158
geostrophy
    isopycnal models, 804
geothermal heating, 264
gradient operator, 20
    Cartesian and GVC, 127
    general vertical coordinate, 127
gravitational acceleration, 155, 766
    effective, 156
gravitational potential, 764
    earth-moon, 769
    general case, 764
gravitational potential energy, 155, 400
gravity waves, 575
Green's function method, 57
Green's identities, 33

Hadley circulation, 315
Hamilton's operator, 20
harmonic function, 24, 26, 274, 576
    mean-value property, 47
harmonic functions, 47
heat equation, 50
Helmholz decomposition, 37, 274
Helmholz equation, 56
Hooke's Law, 327
hydrostatic approximation, 312, 416
    ocean model equations, 435
    scaling, 419
    vertical motion, 416, 423
hydrostatic balance, 312
hydrostatic fluid, 175
hydrostatic scaling
    with stratification, 451
hyperbolic PDE
    classification, 46
    domain of influence, 55
hyperbolic PDEs, 53

ideal gas law, 186
incompressible flow, 269
incropping buoyancy, 444
index convention, 6
index gymnastics, 14, 81
inertial acceleration, 145
inertial oscillations, 165
    ocean, 166
    period, 166
inertial reference frame, 138
inertial velocity, 145
    spherical, 151
inexact differential, 35
    integrating factor, 36
injection work, 405
integral curve, 193
integral surface, 44
intensive property, 257
internal forces, 317
irreversible thermodynamics, 395
irrotational flow, 275, 575
irrotational vector, 25

isopycnal coordinates, 79
isopycnal primitive equations, 800
isopycnals, 246

Jacobian
    ratio of volumes, 199
Jacobian of transformation, 86, 90
Joule heating, 404
Joule unit of energy, 187

kinematic b.c.
    buoyancy surface, 246
    incompressible flow, 270
    material interface, 242
    moving material, 240
    ocean free surface, 246
    permeable surface, 244
    static material, 239
kinematic boundary condition, 239
kinetic energy, 138, 168, 401
    axial angular momentum, 169
    Boussinesq, 439
    cartesian coordinates, 168
    hydrostatic Boussinesq, 455
    spherical coordinates, 169
kinetic energy dissipation, 402
kinetic stress, 324
kinetic theory, 186
Knudsen number, 183
Kronecker tensor, 7, 84, 144

laboratory reference frame, 138, 195
Lagrangian coordinates, 79
Lagrangian reference frame, 180, 195
Lagrangian time derivative, 78, 202
Laplace's equation, 47, 576
    properties, 47
Laplacian operator, 24
law of cosines, 769
Leibniz's Rule, 259, 633
Leibniz-Reynolds
    brute force, 265
Leibniz-Reynolds transport theorem, 259
length scale
    gradient, 185
    macroscopic, 182
    mean free path, 182, 187
    measurement, 185
    molecular, 182
Levi-Citiva tensor
    general coordinates, 92
    volume element, 93
Levi-Civita tensor, 9
linear momentum, 137
    conservation, 137
    finite volume, 325
    Lagrangian volume, 325
local thermodynamic equilibrium, 395

mass budget
    fluid column, 246, 248
mass budgets, 260
mass conservation, 180
    arbitrary Eulerian region, 235
    Eulerian flux-form, 233
    finite Eulerian region, 234
    Jacobian derivation, 232
    Lagrangian, 232
    Lagrangian derivation, 237
    multi-components, 253
mass equation
    general vertical coordinates, 682
mass flux through a surface, 244
material constant, 257
material coordinates, 79, 196, 197
    choices, 197
material fluid parcel, 193
material invariance, 204, 205
material invariant, 257
material surface, 204
material time derivative, 78, 202
    invariance, 207
    isopycnal models, 802
materially invariant volume, 269
matter conservation, 267
mechanical energy, 400, 403
    Boussinesq, 439, 456
    conservation, 170
    particle, 170
mechanical energy dissipation, 329
mechanical pressure, 326
mechanical stress, 324
mechanical work, 138, 168
metric tensor
    Cartesian coordinates, 84
    coordinate representation, 85
    coordinate transformation, 86

determinant, 90
general, 85
inverse, 88, 89
relating vectors & one-forms, 88
metricity condition, 98
moist static energy, 412
molecular composition
air, 186
water, 186
molecular diffusivity, 456
molecular viscosity, 329, 456
moment arm, 148
moment-arm, 313
momentum
angular, 138
approximate axial angular, 175
axial angular, 148, 167, 171, 313, 426
axial angular and Coriolis, 172
linear, 167
potential, 163, 167
momentum equation
flux form, 324, 683
general vertical coordinates, 682
horizontal, 683
spherical, 310
vector-invariant, 311, 683
vertical, 683
Monge gauge, 65
Montgomery potential, 801

natural coordinates
isobars, 494
Navier-Stokes equation, 310
nearly horizontal motion, 175
Neumann boundary condition, 48
Newton's equation of motion, 309
Newton's gravitational constant, 155
Newton's gravitational law, 155
Newton's second law, 137
Newton's third law, 317, 319, 330
pressure, 319
Newtonian fluid, 327
Newtonian gravity field, 769
Newtonian time, 76, 78
no normal flow b.c., 239
no-slip boundary condition, 331
Noether's Theorem, 163
non-conservative forces, 139
non-conservative process, 267
non-dimensionalization
Ekman balance, 498
hydrostatic approximation, 452
non-divergent velocity field, 269
non-inertial reference frame, 138
non-Newtonian constitutive relation, 328
non-orthogonal coordinates, 79
normal derivative, 21
normal direction, 21, 66
normal vector, 272

oblate spheroid, 158
ocean-atmosphere equations, 411
one-forms, 87
basis, 87
coordinate representation, 87
one-way wave equation, 53
orbital motion, 769
orthogonal curvilinear coordinates, 113
oscillator equation, 165
outcropping buoyancy, 444
outer product, 77, 324
outward normal
solid-earth bottom, 240
overturning streamfunction, 278

parabolic PDE
classification, 46
parabolic PDEs, 50
smoothing property, 51
partial derivatives
transformation, 95
particle mechanics, 137
passive tracers, 256
path, 66
orientable, 66
simple, 66
path integral
arc length, 29
scalar function, 28
vector function, 29
pathlines, 210
analog to car flow, 210
permutation symbol, 9, 91
product identity, 14
physics as geometry, 76
planar curves, 67

planetary Cartesian coordinates, 147
planetary centrifugal acceleration, 148
planetary centripetal acceleration, 148
point transformation, 197
points, 6
Poisson's equation, 49
    max-min principle, 49
polar radius, 158
position coordinates, 197
position vector
    Cartesian representation, 146
    spherical representation, 146
potential energy, 139
    Boussinesq, 440
potential flow, 575
potential momentum, 157, 167
potential vorticity
    Boussinesq, 690
    generalized vertical coordinates, 687
    isopycnal coordinates, 805
power, 168
pressure
    contact force, 686
    contact stress, 684
    gradient force, 684
    interfacial form stress, 686
    Pascal unit, 187
    Poisson equation, 455
    stagnation, 406
    standard atmosphere, 186
    stress, 318
    total, 406
pressure form stress, 320
pressure gradient
    baroclinic, 422
    bottom, 422
    hydrostatic, 422
primitive equations, 173, 416
    nomenclature, 418
prognostic equation, 60, 205
pseudo angular momentum, 164
Pythagoras' Theorem, 7

quasi-equilibrium thermodynamics, 395
quasi-Lagrangian coordinates, 79
quasi-static process, 181, 188, 395

radius of curvature, 69, 71
Rayleigh drag, 139, 496
reference density, 270
reference frames
    absolute, 140
    inertial, 140
    non-inertial, 140
    rotating, 140
region following the flow, 236
Reynolds number, 184, 328, 499
    atmosphere, 329
    Gulf Stream rings, 329
Reynolds stress, 324
Reynolds transport theorem, 238
    linear momentum, 325
    multi-component fluid, 261
Richardson number, 453
right hand rule, 9
Rossby number, 499

scalar product, 7
    general tensors, 89
    invariance, 17
self-attraction and loading, 773
shallow fluid approximation, 173, 175, 416
slope of a surface, 245
solid body rotation, 142
solid body velocity
    spherical, 148
solid-body rotation, 578, 581
spherical coordinates, 103
    basis one-forms, 105
    basis vectors, 105
    differential operators, 107
    Levi-Civita tensor, 106
    metric, 106
    position, 106
    summary, 108
    transformation from Cartesian, 104
    vector components, 107
    vector cross product, 107
    velocity, 106
    volume element, 106
static equilibrium sea level, 764, 765
steady flow, 202
steady state, 205, 257
steepest descent, 20
Stoke's theorem, 100
Stokes' theorem, 31

streaklines, 212
streamfunction
    meridional-depth overturning, 278
    streamlines, 272
    transport, 272, 275, 276
    two-dimensional flow, 271
    vector, 275
    vertical gauge, 276
streamlines, 211, 239, 272
streamlines versus pathlines, 211
streamtubes, 211
stress
    contact, 308
    deviatoric, 308
    friction, 308, 309
    pressure, 308
stress boundary condition
    Lagrangian interface, 332
    permeable interface, 332
    solid boundary, 330
stress on an interface, 329
stress tensor, 321
    symmetry, 322
stress-strain relation, 326
strong formulation, 192, 252, 263
subharmonic function, 49
surface
    orientable, 66
surface area, 71
    horizontal projection, 245, 266
surface tension, 318, 329
surface velocity, 241
symmetries, 163, 167
symmetry, 137

tangent direction, 21, 66
tangent plane approximation, 418
tangent plane Cartesian coordinates, 147
tangent vector, 272
Taylor series, 20
tensor analusis
    covariant differentiation, 77
tensor analysis
    algebra summarized, 143
    contraction of indices, 77
    permissible operations, 76
    quotient rule, 77
    tensor fields, 6
    tensors, 6
    tensors and matrices, 17
    uses for GFM, 2
test particles, 193
tetrahedron fluid region, 321
thermal expansion, 271
thermobaricity, 435
thermodynamic pressure, 326
    incompressible fluid, 326
thermodynamics
    moving fluid, 396
thermohaline circulation, 431
thickness
    specific, 285, 682
    specific thickness, 802
thickness equation
    isopycnal models, 802
tidal acceleration, 766
tides
    earth-moon system, 767
    realistic effects, 773
    semi-diurnal, 767
time
    coordinate, 78
    parameter, 78
time averages, 60
time derivative notation, 239
time derivatives
    summary, 263
time scale
    molecular collisions, 188
time tendency, 60, 202
torque, 138
total time derivative, 201
tracer boundary conditions, 263
    air-sea boundary, 265
    bottom, 264
    no-normal derivative, 264
tracer budgets, 260, 263
tracer concentration, 255
tracer equation, 252, 255
    derived, 254
    general vertical coordinates, 682
tracers
    passive, 256
Traditional approximation, 416
trajectories, 196
trajectory, 77, 137, 210

transformation
Cartesian, 16
Cartesian coordinates, 14
geometric, 16
inverse, 15
jacobian, 198
material and position, 197
orthogonal, 15
rotation, 15
tensor, 17
transformation matrix
coordinates, 86
determinant, 91, 94
discrete algorithm, 198
general vertical coordinates, 117
inverse, 95
Jacobian, 86, 90
space-time components, 94
translation motion, 578
transport, 40
transport theorem
scalar field, 260
turbulence, 328
turbulent energy cascade, 329

unit vector, 7
unit vectors
rotation, 22
universal gas constant, 186

vanishing layers, 805
vector product, 9
general coordinate, 93
geometric, 11
vector rotation, 142
vector streamfunction, 275
vectors, 6
velocity, 137
basis vectors, 145
coordinate, 145
particle motion, 145
spherical, 151
surface, 241
velocity equation
vector-invariant, 311, 683, 804
velocity potential, 274, 277
velocity scale
molecular rms speed, 188
velocity vector, 80
planetary Cartesian representation, 147
spherical representation, 147
viscosity, 326
air, 185
dynamic, 185
eddy, 438
kinematic, 185
water, 185
volume
n-space, 13
defined by vectors, 12
element for integration, 13
volume budget
material region, 236
volume element
general coordinates, 93
invariant, 90
volume equation, 233
volume evolution, 260
incompressible, 277
vortical flow, 274
vorticity, 36
absolute, 579
bottom boundary, 634
depth integrated, 633
general vertical coordinate, 683
generalized vertical coordinates, 687
isopycnal coordinates, 806
isopycnal models, 804
planetary, 579
relative, 579, 683
surface boundary, 634

wave equation
domain of influence, 55
weak formulation, 192, 253, 263
wing flow, 406
work, 168

zero buoyancy layer thickness, 444