Jon Pierre Fortney

A Visual Introduction to Differential Forms and Calculus on Manifolds

Birkhäuser

A Visual Introduction to Differential Forms and Calculus on Manifolds

Jon Pierre Fortney

A Visual Introduction to Differential Forms and Calculus on Manifolds

Birkhäuser

Jon Pierre Fortney
Department of Mathematics and Statistics
Zayed University
Dubai, United Arab Emirates

ISBN 978-3-319-96991-6 ISBN 978-3-319-96992-3 (eBook)
https://doi.org/10.1007/978-3-319-96992-3

Library of Congress Control Number: 2018952359

Mathematics Subject Classification (2010): 53-01, 57-01, 58-01, 53A45, 58C20, 58C35

This book is published under the imprint Birkhäuser, www.birkhauser-science.com by the registered company Springer Nature Switzerland AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

To my parents, Daniel and Marlene Fortney, for all of their love and support.

Preface

Differential forms, while not quite ubiquitous in mathematics, are certainly common. And the role differential forms play appears in a wide range of mathematical fields and applications. Differential forms, and their integration on manifolds, are part of the foundational material with which it is necessary to be proficient in order to tackle a wide range of advanced topics in both mathematics and physics. Some upper-level undergraduate books and numerous graduate books contain a chapter on differential forms, but generally the intent of these chapters is to provide the computational tools necessary for the rest of the book, not to aid students in actually obtaining a clear understanding of differential forms themselves. Furthermore, differential forms often do not show up in the typically required undergraduate mathematics or physics curriculums, making it both unlikely and difficult for students to gain a deep and intuitive feeling for them. One of the two aims of this book is to address and remedy exactly this gap in the typical undergraduate mathematics and physics curriculums.

Additionally, it is during the second year and third year that undergraduate mathematics majors are making the transition from the concrete computation-based subjects generally found in high school and lower-level undergraduate courses to the abstract topics generally found in upper-level undergraduate and graduate courses. This is a tricky and challenging time for many undergraduate students, and it is during this period that most undergraduate programs see the highest attrition rates. Furthermore, while many undergraduate mathematics programs require mathematical structures or introduction to proofs class, there are also many programs do not. And often a single course meant to help students' transition from the concrete computations of calculus to the abstract notions of theoretical mathematics is not enough; a majority of students need more support in making this transition. The second aim of this book has been to help students make this transition to a mathematically more abstract and mature way of thinking.

Thus, the intended audience for this book is quite broad. From the perspective of the topics covered, this book would be completely appropriate for a modern geometry course; in particular, a course that is meant to help students make the jump from Euclidian/Hyperbolic/Elliptic geometry to differential geometry, or it could be used in the first semester of a two-semester sequence in differential geometry. It would also be appropriate as an advanced calculus course that is meant to help students' transition to calculus and analysis on manifolds. Additionally, it would be appropriate for an undergraduate physics program, particularly one with a more theoretical bent, or in a physics honors program; it could be used in a geometry for physics course. Finally, from this perspective, it is also a perfect reference for graduate students entering any field where a thorough knowledge of differential forms is necessary and who find they lack the necessary background. Though graduate students are not the intended audience, they could read and assimilate the ideas quite quickly, thereby enabling them to gain a fairly deep insight into the basic nature of differential forms before tackling more advanced material.

But from the perspective of helping undergraduate students make the transition to abstract mathematics, this book is absolutely appropriate for any and all second- or third-year undergraduate mathematics majors. Its mathematical prerequisites are light; a course in vector calculus is completely sufficient and the few necessary topics in linear algebra are covered in the introductory chapter. However, this book has been carefully written to provide undergraduates the scaffolding necessary to aid them in the transition to abstract mathematics. In fact, this material dove-tails with vector calculus, with which students are already familiar, making it a perfect setting to help students transition to advanced topics and abstract ways of thinking. Thus this book would be ideal in a second- or third-year course whose intent is to aid students in transitioning to upper-level mathematics courses.

As such, I have employed a number of different pedagogical approaches that are meant to complement each other and provide a gradual yet robust introduction to both differential forms in particular and abstract mathematics in general. First, I have made a great deal of effort to gradually build up to the basic ideas and concepts, so that definitions, when made, do not appear out of nowhere; I have spent more time exploring the "how" and "why" of things than is typical for most post-calculus math books. Additionally, the two major proofs that are done in this book (the generalized Stokes' theorem and the Poincaré lemma) are done very slowly and carefully, providing more detail than is usual. Second, this book tries to explain and help the reader develop, as much as possible, their geometric intuition as it relates to differential forms. To aid in this endeavor there are over 250 figures in the book. These images play a crucial role in aiding the student to understand and visualize the concepts being discussed and are an integral part of the exposition. Third, Students benefit from seeing the same idea presented and explained multiple times and from different perspectives; the repetition aids in learning and internalizing the idea. A number of the more important topics are discussed in both the $\mathbb{R}^n$ setting as well as in the setting of more abstract manifolds. Also, many topics are discussed from a visual/geometric approach as well as from a computational approach. Finally, there are over 200 exercises interspersed with the text and about 200 additional end-of-chapter exercises. The end of chapter questions are primarily computational, meant to help students gain familiarity and proficiency with the notation and concepts. Questions interspersed in the text range from trivial to challenging and are meant to help students genuinely engage with the readings, absorb fundamental ideas, and look carefully and critically at various steps of the computations done in the text. Taken together, these questions will not only allow students to gain a deeper understanding of the material, but also gain confidence in their abilities and internalize the essential notation and ideas.

Putting all of these pedagogical strategies together may result in an exposition that, to an expert, would seem at times to be unnecessarily long, but this book is based on my own experiences and reflections in both learning and teaching and is entirely written with students fairly new to mathematics in mind. I want my readers to truly understand and internalize these ideas, to gain a deeper and more accurate perception of mathematics, and to see the beautiful interconnectedness of the subject; I want my readers to walk away feeling that they have genuinely mastered a body of knowledge and not simply learned a set of disconnected facts.

Covering the full book is probably too much to ask of most students in a one-semester course, but there are a number of different pathways through the book based on the overall emphasis of the class. Provided below are the ones I consider most appropriate:

1. For schools on the quarter system or for a seminar class: 1 (optional), 2, 3, 4, 6, 7, 9 (optional).
2. Emphasizing differential forms and geometry: 1 (optional), 2–9, 10 (optional), 11, Appendix B1–2 (optional).
3. Emphasizing physics: 1 (optional), 2–7, 9, 11, 12, Appendix A (optional), Appendix B3–5 (optional).
4. Emphasizing the transition to abstract mathematics: 1 (optional), 2–4, 6–11.
5. Advanced students or as a first course to an upper-level sequence in differential geometry: 1 (optional), 2–11, Appendix A (optional), Appendix B (optional).

A word of warning, Appendix A on tensors was included in order to provide the proof of the global formula for exterior differentiation, a proof I felt was essential to provide in this book and which relies on the lie derivative. However, from a pedagogical perspective Appendix A is probably too terse and lacks the necessary examples to be used as a general introduction to tensors, at least if one wishes to cover anything beyond the mere definitions and basic identities. Instructors should keep this in mind when deciding whether or not to incorporate this appendix into their classes.

Finally, I would like to express my sincere appreciation to Ahmed Matar and Ron Noval for all of their invaluable comments and suggestions. Additionally, I would like to thank my friend Rene Hinojosa for his ongoing encouragement and support.

Dubai, United Arab Emirates
May 2018

Jon Pierre Fortney

Contents

Chapter 1
Background Material

We will begin by introducing, or reviewing, three basic topics that are necessary for the remainder of this book; vector spaces, determinants, and derivatives of multivariable functions. In all three cases the topics are presented in a manor that is consistent with the needs of the rest of this book.

In section one we introduce vector spaces, moving immediately to the specific vector space that will concern us, namely $\mathbb{R}^n$. The dual space of $\mathbb{R}^n$ is then discussed at length since dual spaces play a central role with regards to differential forms. Section two introduces the determinant as a function that finds the n-dimensional volume of an n-dimensional parallelepiped. We use our intuitive ideas of what volume is to derive the formula for the determinant of a matrix. This is the key relationship necessary to understand the wedgeproduct of differential forms. Finally, in section three we introduce very briefly derivatives to multivariable functions. The exterior derivative introduced in this book builds on and generalizes derivatives of multivariable functions.

Since there is no assumption that students reading this book have completed a linear algebra course the sections on vector spaces and determinants are covered in greater detail. However, since we assume most students have recently completed a multivariable calculus class fewer details are given in the section on the derivatives of multivariable functions.

1.1 Review of Vector Spaces

We will start out with a review of **vector spaces**, which are usually introduced in a linear algebra course. Abstractly a vector space is a set V equipped with binary operations of addition $+$ and scalar multiplication $\cdot$, an additive identity 0, and an "operation of negation," which satisfies the following eight axioms. Here we have $u, v, w \in V$ and $c, d \in \mathbb{R}$. The elements of a vector space are called **vectors**. The elements c and d are often called **scalars**.

1. $v + w = w + v$
2. $(u + v) + w = u + (v + w)$
3. $v + 0 = v = 0 + v$
4. $v + (-v) = 0$
5. $1 \cdot v = v$
6. $c \cdot (d \cdot v) = (c \cdot d) \cdot v$
7. $c \cdot (v + w) = c \cdot v + c \cdot w$
8. $(c + d) \cdot v = c \cdot v + d \cdot v$

Since we are actually only interested in one concrete example of a vector space we immediately move to the vector spaces we will be concerned with, the vector space $\mathbb{R}^n$. Addition and scalar multiplication are done pointwise. Using $\mathbb{R}^2$ as an example that means that

$$c \cdot \begin{bmatrix} u_1 \\ u_2 \end{bmatrix} + \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = \begin{bmatrix} c \cdot u_1 + v_1 \\ c \cdot u_2 + v_2 \end{bmatrix}.$$

J. P. Fortney, *A Visual Introduction to Differential Forms and Calculus on Manifolds*,
https://doi.org/10.1007/978-3-319-96992-3_1

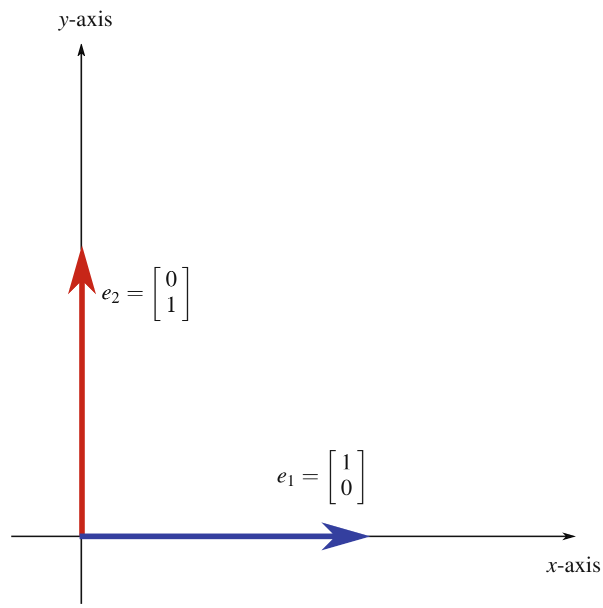


Fig. 1.1 Cartesian plane $\mathbb{R}^2$ with vectors e_1 and e_2 shown

Question 1.1 Consider $\mathbb{R}^2$ with $u = \begin{bmatrix} u_1 \\ u_2 \end{bmatrix}$, $v = \begin{bmatrix} v_1 \\ v_2 \end{bmatrix}$, and $w = \begin{bmatrix} w_1 \\ w_2 \end{bmatrix}$ and $c, d \in \mathbb{R}$. Show that elements of $\mathbb{R}^2$ satisfy the above eight axioms.

If you have taken a course in linear algebra you may recall that a basis for a vector space V is a subset of elements $\beta = \{v_1, \dots, v_n\}$ of V which is both **linearly independent** and **span** the vector space V. Linear independence means that if $a_1v_1 + a_2v_2 + \dots + a_nv_n = 0$ for some scalars $a_1, a_2, \dots, a_n$ then necessarily $a_1 = a_2 = \cdots = a_n = 0$. Spanning means that for any $v \in V$ it is possible to find scalars $a_1, a_2, \dots, a_n$ such that $v = a_1v_1 + a_2v_2 + \dots + a_nv_n$. This is not a course in linear algebra, our main goal here is simply to remind you of some facts about linear algebra and to give you a better intuitive feel of the concepts so we will not go into any more detail.

The Euclidian vector space $\mathbb{R}^2$ is one you are very familiar with, it is usually pictured as a cartesian plane as in Fig. 1.1. The two Euclidian unit vectors $e_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$ and $e_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$ are shown. A typical vector in $\mathbb{R}^2$ would look like $\begin{bmatrix} x \\ y \end{bmatrix}$, for some $x, y \in \mathbb{R}$. It is straightforward to see that e_1 and e_2 are linearly independent. Suppose we have an a_1 and an a_2 such that

$$a_1 \begin{bmatrix} 1 \\ 0 \end{bmatrix} + a_2 \begin{bmatrix} 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}.$$

The only way this can happen is if $a_1 = 0$ and $a_2 = 0$, which is what we needed to show in order to conclude e_1 and e_2 were linearly independent. To see that e_1 and e_2 span $\mathbb{R}^2$ then we must show that for any vector $\begin{bmatrix} x \\ y \end{bmatrix}$ in $\mathbb{R}^2$ we can find a_1 and a_2 such that

$$\begin{bmatrix} x \\ y \end{bmatrix} = a_1 \begin{bmatrix} 1 \\ 0 \end{bmatrix} + a_2 \begin{bmatrix} 0 \\ 1 \end{bmatrix}.$$

But this is clearly accomplished if we let $a_1 = x$ and $a_2 = y$. Thus e_1 and e_2 span $\mathbb{R}^2$ and are called **standard basis** vectors or sometimes **Euclidian basis** vectors. Clearly the standard basis vectors are easy to work with. Of course, the set of Euclidian unit vectors $\{e_1, e_2\}$ is not the only possible basis for $\mathbb{R}^2$.

Question 1.2 Show that the set of vectors $\left\{ \begin{bmatrix} 2 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \end{bmatrix} \right\}$ is also a basis for $\mathbb{R}^2$. Then find another basis for $\mathbb{R}^2$.

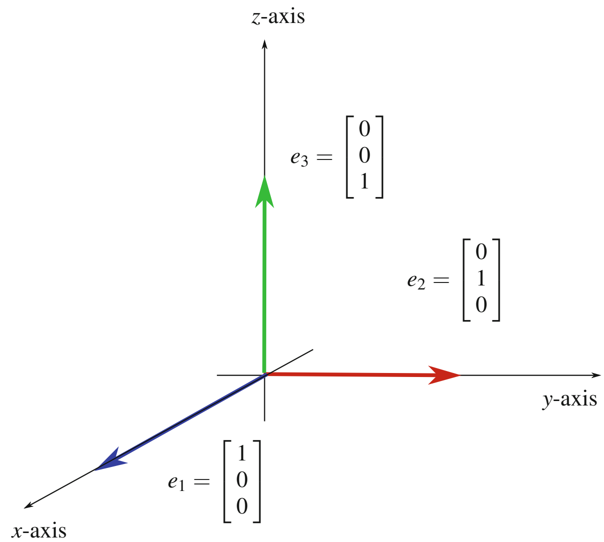


Fig. 1.2 Cartesian plane $\mathbb{R}^3$ with vectors e_1, e_2 and e_3 shown

Question 1.3 Now consider the vector space $\mathbb{R}^3$. This vector space is usually pictured as in Fig. 1.2. Consider the three vectors $e_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$, $e_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$, and $e_3 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$

(a) What does a typical vector in $\mathbb{R}^3$ look like?
(b) Show that e_1, e_2, e_3 are linearly independent.
(c) Show that e_1, e_2, e_3 span $\mathbb{R}^3$

The vector space $\mathbb{R}^n$ is an n-tuple of numbers

$$\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}.$$

It is impossible to actually imagine but we generally resort to using a "cartoon" picture of $\mathbb{R}^n$ that is basically the same as our picture of $\mathbb{R}^3$ The Euclidian unit vectors

$$e_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \ e_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \ e_3 = \begin{bmatrix} 0 \\ 0 \\ 1 \\ \vdots \\ 0 \end{bmatrix}, \ \ldots, e_n = \begin{bmatrix} 0 \\ 0 \\ 0 \\ \vdots \\ 1 \end{bmatrix},$$

are a basis of the vector space $\mathbb{R}^n$ called the standard basis.

One of the things that we will be doing throughout this book is trying to help you develop good and appropriate pictures, cartoons, and mental images to help you gain an in-depth and intuitive understanding of differential forms and their uses. Occasionally we will attempt to draw things that cannot accurately be portrayed in two or three dimensions. You must always use care when thinking in terms of these cartoons, a lot of information is actually lost in them. However, we believe that having the appropriate approximate pictures will go a long way in helping you understand the situation at hand. Finally, we emphasise that **we will always treat elements of vector spaces as column vectors and never as row vectors**. The reason for this will become apparent later on.

Now we will consider a certain type of transformation between vector spaces called a **linear transformation**, also sometimes called a **linear map** or a **linear operator**. Suppose T is a mapping between $\mathbb{R}^n$ and $\mathbb{R}^m$, that is, $T : \mathbb{R}^n \longrightarrow \mathbb{R}^m$. Then T is called a linear transformation if it has the following two properties

1. $T(v + w) = T(v) + T(w)$
2. $T(c \cdot v) = c \cdot T(v)$

where $v, w \in \mathbb{R}^n$ and $c \in \mathbb{R}$.

The first property, $T(v + w) = T(v) + T(w)$, means that performing the transformation on the sum of two vectors in the domain vector space has the same result as performing the transformation on the two vectors separately and then adding the results in the codomain vector space. The second property, $T(c \cdot v) = c \cdot T(v)$, means that performing the transformation on a scalar multiple of a vector has the same result as performing the transformation on a vector and then multiplying the result by the scalar. If T is a linear transformation from $\mathbb{R}^n$ to $\mathbb{R}$ then it is often simply called a **linear function** or a **linear functional**.

We begin by examining a few concrete examples. Consider the mapping $T : \mathbb{R}^2 \longrightarrow \mathbb{R}^3$ defined by

$$T\left(\begin{bmatrix} x \\ y \end{bmatrix}\right) = \begin{bmatrix} x - y \\ x + y \\ 2x \end{bmatrix}.$$

First we consider what this mapping does to a specific element of $\mathbb{R}^2$,

$$T\left(\begin{bmatrix} 4 \\ 10 \end{bmatrix}\right) = \begin{bmatrix} -6 \\ 14 \\ 8 \end{bmatrix}.$$

Thus the mapping T sends the vector $\begin{bmatrix} 4 \\ 10 \end{bmatrix} \in \mathbb{R}^2$ to the vector $\begin{bmatrix} -6 \\ 14 \\ 8 \end{bmatrix} \in \mathbb{R}^3$. Our goal now is to see if the mapping T is in fact a linear transformation. To do this we need to show that both properties hold. We start with property one, does

$$T\left(\begin{bmatrix} a_1 \\ a_2 \end{bmatrix} + \begin{bmatrix} b_1 \\ b_2 \end{bmatrix}\right) = T\left(\begin{bmatrix} a_1 \\ a_2 \end{bmatrix}\right) + T\left(\begin{bmatrix} b_1 \\ b_2 \end{bmatrix}\right)?$$

We first find an expression for the left hand side,

$$T\left(\begin{bmatrix} a_1 \\ a_2 \end{bmatrix} + \begin{bmatrix} b_1 \\ b_2 \end{bmatrix}\right) = T\left(\begin{bmatrix} a_1 + b_1 \\ a_2 + b_2 \end{bmatrix}\right) = \begin{bmatrix} a_1 + b_1 - a_2 - b_2 \\ a_1 + b_1 + a_2 + b_2 \\ 2a_1 + 2b_1 \end{bmatrix}.$$

We next find an expression for the right hand side

$$T\left(\begin{bmatrix} a_1 \\ a_2 \end{bmatrix}\right) + T\left(\begin{bmatrix} b_1 \\ b_2 \end{bmatrix}\right) = \begin{bmatrix} a_1 - a_2 \\ a_1 + a_2 \\ 2a_1 \end{bmatrix} + \begin{bmatrix} b_1 - b_2 \\ b_1 + b_2 \\ 2b_2 \end{bmatrix} = \begin{bmatrix} a_1 + b_1 - a_2 - b_2 \\ a_1 + b_1 + a_2 + b_2 \\ 2a_1 + 2b_1 \end{bmatrix}.$$

Comparing these two expressions we see that indeed property one holds. Next we look at property two, does

$$T\left(c \cdot \begin{bmatrix} a_1 \\ a_2 \end{bmatrix}\right) = c \cdot T\left(\begin{bmatrix} a_1 \\ a_2 \end{bmatrix}\right)?$$

Finding an expression for the left hand side we have

$$T\left(c \cdot \begin{bmatrix} a_1 \\ a_2 \end{bmatrix}\right) = T\left(\begin{bmatrix} ca_1 \\ ca_2 \end{bmatrix}\right) = \begin{bmatrix} ca_1 - ca_2 \\ ca_1 + ca_2 \\ 2ca_2 \end{bmatrix}.$$

We next find an expression for the right hand side

$$c \cdot T\left(\begin{bmatrix} a_1 \\ a_2 \end{bmatrix}\right) = c \cdot \begin{bmatrix} a_1 - a_2 \\ a_1 + a_2 \\ 2a_2 \end{bmatrix} = \begin{bmatrix} ca_1 - ca_2 \\ ca_1 + ca_2 \\ 2ca_2 \end{bmatrix}.$$

Comparing these two expressions we see that property two holds. Thus, since both property one and two hold the mapping T is indeed a linear transformation.

Question 1.4 Consider $T : \mathbb{R}^2 \longrightarrow \mathbb{R}^2$ defined by

$$T\left(\begin{bmatrix} x \\ y \end{bmatrix}\right) = \begin{bmatrix} x + y \\ y + 5 \end{bmatrix}.$$

Is T a linear transformation?

Determining if a mapping $T : \mathbb{R}^n \to \mathbb{R}$ is a linear functional is likewise straightforward. Consider the mapping $T\left(\begin{bmatrix} x \\ y \end{bmatrix}\right) = 2x + y$. To determine if T is a linear functional we again have to show that properties one and two apply. Since both

$$T\left(\begin{bmatrix} a_1 \\ a_2 \end{bmatrix} + \begin{bmatrix} b_1 \\ b_2 \end{bmatrix}\right) = 2a_1 + 2b_1 + a_2 + b_2$$

and

$$T\left(\begin{bmatrix} a_1 \\ a_2 \end{bmatrix}\right) + T\left(\begin{bmatrix} b_1 \\ b_2 \end{bmatrix}\right) = 2a_1 + 2b_1 + a_2 + b_2$$

then the first property holds. And since

$$T\left(c \cdot \begin{bmatrix} a_1 \\ a_2 \end{bmatrix}\right) = 2ca_1 + ca_2$$

and

$$c \cdot T\left(\begin{bmatrix} a_1 \\ a_2 \end{bmatrix}\right) = 2ca_1 + ca_2$$

then property two holds and T is a linear functional.

Question 1.5 Consider the $T : \mathbb{R}^2 \longrightarrow \mathbb{R}$ defined below. Are they linear functionals?

(a) $T\left(\begin{bmatrix} x \\ y \end{bmatrix}\right) = x^2 + y$

(b) $T\left(\begin{bmatrix} x \\ y \end{bmatrix}\right) = xy$

(c) $T\left(\begin{bmatrix} x \\ y \end{bmatrix}\right) = 2x + 3y + 7$

Before we turn our attention to the relationship between linear transformation and matrices we very quickly review both matrices and matrix multiplication. A **matrix** is, simply put, a rectangular array of numbers, generally written enclosed in square brackets. For example,

$$\begin{bmatrix} 7 & 4 & -3 \\ 2 & -9 & 5 \end{bmatrix}, \quad \begin{bmatrix} 2 & -6 \\ 9 & 4 \\ 11 & 1 \\ 6 & 8 \end{bmatrix}, \quad \text{and} \quad \begin{bmatrix} 2 & -4 & 6 & 7 \\ -1 & 8 & 0 & 10 \\ 5 & 2 & -3 & 7 \\ -2 & -4 & 1 & 3 \end{bmatrix}$$

are all examples of matrices. The size of the matrix is indicated with the number of rows and columns, usually written as (rows)×(columns). For example, the sizes of the above matrices are 2×3, 4×2, and 4×4. The entries in the matrix are sometimes called matrix elements and are denoted with variables that have two subscripts, the first for the row number and the second for the column number. Here a 3×3 matrix is written using variables with subscripts,

$$\begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}.$$

It is important to remember that the row subscript comes first and the column subscript comes second, a_{rc}.

Matrix multiplication can seem somewhat complicated if you have never seen it before. First of all, the number of columns the first matrix has must be equal to the number of rows the second matrix has. Second, to get the ijth entry in the product matrix you multiply the ith row of the first matrix with the jth column of the second. Before giving the general formula we illustrate this with an example,

$$\begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \begin{bmatrix} 5 & 6 \\ 7 & 8 \end{bmatrix}.$$

We get the a_{11} term of the product matrix my multiplying the red row with the blue column like this

$$a_{11} = 1 \cdot 5 + 2 \cdot 7 = 19.$$

Similarly

$$\begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \begin{bmatrix} 5 & 6 \\ 7 & 8 \end{bmatrix} \Rightarrow a_{12} = 1 \cdot 6 + 2 \cdot 8 = 22,$$

$$\begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \begin{bmatrix} 5 & 6 \\ 7 & 8 \end{bmatrix} \Rightarrow a_{21} = 3 \cdot 5 + 4 \cdot 7 = 43,$$

$$\begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \begin{bmatrix} 5 & 6 \\ 7 & 8 \end{bmatrix} \Rightarrow a_{22} = 3 \cdot 6 + 4 \cdot 8 = 50.$$

Putting this all together we have

$$\begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \begin{bmatrix} 5 & 6 \\ 7 & 8 \end{bmatrix} = \begin{bmatrix} 19 & 22 \\ 43 & 50 \end{bmatrix}.$$

We now give the general formula for matrix multiplication. If A is an $n \times m$ matrix and B is an $m \times p$ matrix,

$$\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1m} \\ a_{21} & a_{22} & \cdots & a_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nm} \end{bmatrix} \begin{bmatrix} b_{11} & b_{12} & \cdots & b_{1p} \\ b_{21} & b_{22} & \cdots & b_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ b_{m1} & b_{m2} & \cdots & b_{mp} \end{bmatrix} = \begin{bmatrix} c_{11} & c_{12} & \cdots & c_{1p} \\ c_{21} & c_{22} & \cdots & c_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ c_{n1} & c_{n2} & \cdots & c_{np} \end{bmatrix},$$

where

$$c_{ij} = \sum_{k=1}^{m} a_{ik} b_{kj}.$$

Notice that the product matrix has dimension $n \times p$.

We now turn our attention to the relationship between linear transformations and matrices. This is not a linear algebra class so we will not get fancy with either our vector spaces or our bases. We will stick to vector spaces $\mathbb{R}^n$ and the standard basis made up of the Euclidian unit vectors $e_1, e_2, \ldots, e_n$. Technically, in order to write linear transformations $T : \mathbb{R}^n \to \mathbb{R}^m$ as a matrix we have to have **ordered bases** for both $\mathbb{R}^n$ and $\mathbb{R}^m$; that is, our basis elements have to have a set order. But we generally think of our Euclidian unit vectors as having the intuitively obvious order, e_1 first, e_2 second, e_3 third, and so on. We proceed with a couple examples before giving a general definition.

Consider the example we had before, $T : \mathbb{R}^2 \longrightarrow \mathbb{R}^3$ defined by

$$T\left(\begin{bmatrix} x \\ y \end{bmatrix}\right) = \begin{bmatrix} x - y \\ x + y \\ 2x \end{bmatrix}.$$

Using the linear transformation properties and knowledge of how matrix multiplication works we have the following

$$\begin{aligned}
T\left(\begin{bmatrix} x \\ y \end{bmatrix}\right) &= T\left(\begin{bmatrix} x \\ 0 \end{bmatrix} + \begin{bmatrix} 0 \\ y \end{bmatrix}\right) \\
&= T\left(x\begin{bmatrix} 1 \\ 0 \end{bmatrix} + y\begin{bmatrix} 0 \\ 1 \end{bmatrix}\right) \\
&= T\left(x\begin{bmatrix} 1 \\ 0 \end{bmatrix}\right) + T\left(y\begin{bmatrix} 0 \\ 1 \end{bmatrix}\right) && \text{Property One} \\
&= xT\left(\begin{bmatrix} 1 \\ 0 \end{bmatrix}\right) + yT\left(\begin{bmatrix} 0 \\ 1 \end{bmatrix}\right) && \text{Property Two} \\
&= x\begin{bmatrix} 1 \\ 1 \\ 2 \end{bmatrix} + y\begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix} \\
&= \begin{bmatrix} 1 & -1 \\ 1 & 1 \\ 2 & 0 \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix} && \text{Def of Matrix Mult.}
\end{aligned}$$

By equating the left and right sides we can see that we can write the linear transformation T as a matrix, namely,

$$T \equiv \begin{bmatrix} 1 & -1 \\ 1 & 1 \\ 2 & 0 \end{bmatrix}.$$

This computation probably seems obvious to you and you may wonder why we went through all of these steps. The main reason was to illustrate the way that the answer depends on how T acts on the basis vectors of $\mathbb{R}^2$. Let's consider this in more detail. First we let T act on the first basis vector of $\mathbb{R}^2$ to get

$$T\left(\begin{bmatrix} 1 \\ 0 \end{bmatrix}\right) = \begin{bmatrix} 1 \\ 1 \\ 2 \end{bmatrix} = 1\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} + 1\begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} + 2\begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}.$$

Notice that the coefficients for the basis vectors of $\mathbb{R}^3$ are the entries of the first column of the matrix representation of T. Similarly, when we let T act on the second basis vector of $\mathbb{R}^2$ to get

$$T\left(\begin{bmatrix} 0 \\ 1 \end{bmatrix}\right) = \begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix} = -1\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} + 1\begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} + 0\begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}.$$

Here the coefficients for the basis vectors of $\mathbb{R}^3$ are the entries of the second column of the matrix representation of T. Now we give the formal definition of the matrix representation of a linear transformation. Suppose that $T : \mathbb{R}^n \to \mathbb{R}^m$. Let $e_1, e_2, \ldots, e_n$ be the standard basis of $\mathbb{R}^n$ and let $\tilde{e}_1, \tilde{e}_2, \ldots, \tilde{e}_m$ be the standard basis of $\mathbb{R}^m$. Then for $1 \leq j \leq n$ there are unique numbers a_{ij} such that

$$T(e_j) = \sum_{i=1}^{m} a_{ij}\tilde{e}_i.$$

Then the matrix representation of T is given by the $m \times n$ matrix with entries a_{ij}.

Question 1.6 Find the matrix representation of the linear transformation given by

$$T\left(\begin{bmatrix} x \\ y \end{bmatrix}\right) = \begin{bmatrix} 2x + 3y \\ 0 \\ 2x - 5y \end{bmatrix}.$$

Now we turn our attention to linear functionals, which are a special case of linear transformations when the codomain is simply $\mathbb{R}$. Consider the linear functional from before, $T\left(\begin{bmatrix} x \\ y \end{bmatrix}\right) = 2x + y$. We find the matrix representation of this linear functional,

$$T\left(\begin{bmatrix} x \\ y \end{bmatrix}\right) = 2x + y = [2, 1]\begin{bmatrix} x \\ y \end{bmatrix}.$$

We can see that the matrix representations of linear functionals are $1 \times n$ matrices, which are just row vectors. This is an important point, linear functionals are represented by row vectors. When we write $1 \times n$ matrices, or row vectors, we generally put commas between the entries just to make everything clear.

The last major topic in this section is the definition of the dual space of a vector space. Suppose we are given a vector space V. In this book the only vector spaces we will be considering are $\mathbb{R}^n$ for some integer n. The set of all linear functionals on V is called the **dual space** of V, which is denoted as V^*.

It turns out that V^*, the dual space of V, is itself a vector space. To show this we begin by explaining what addition of linear functionals means. We can add two linear functionals $S : V \to \mathbb{R}$ and $T : V \to \mathbb{R}$ to give a new linear functional $S + T : V \to \mathbb{R}$, with the functional $S + T$ being evaluated as you would expect it to be, $(S + T)(v) = S(v) + T(v)$, where $v \in V$. In order to show that the dual space V^* of V, or the set of linear functionals on V, is itself a vector space we have to show that the eight vector space axioms are satisfied.

Consider the first axiom which we had written as $v + w = w + v$, where v, w were elements in the vector space. But now we are dealing with V^*, with elements S and T. We want to show that $S + T = T + S$. How do we do that? We have to use our definition of how addition of linear functionals works. Given any element $v \in V$ we have $(S + T)(v) = S(v) + T(v)$. Now recalling that both $S(v)$ and $T(v)$ are simply real numbers, and that addition of real numbers is commutative, we have that $S(v) + T(v) = T(v) + S(v)$. And finally, using addition of linear functionals we have $T(v) + S(v) = (T + S)(v)$. Putting all of this together we have $(S + T)(v) = (T + S)(v)$. Since this is true for any $v \in V$ we can write $S + T = T + S$, thereby showing that the first axiom is satisfied.

Question 1.7 Suppose that $R, S, T : \mathbb{R}^n \to \mathbb{R}$ are linear functionals on V, that is $R, S, T \in V^*$ and c, d are real numbers. Show that the V^* is a vector space by showing that the other seven vector space axioms are also satisfied.

We will continue by considering the specific vector space $\mathbb{R}^2$. Any element of $\begin{bmatrix} x \\ y \end{bmatrix} \in \mathbb{R}^2$ can be written as $\begin{bmatrix} x \\ y \end{bmatrix} = x\begin{bmatrix} 1 \\ 0 \end{bmatrix} + y\begin{bmatrix} 0 \\ 1 \end{bmatrix}$. Suppose $T \in (\mathbb{R}^2)^*$, then we have

$$\begin{aligned} T\left(\begin{bmatrix} x \\ y \end{bmatrix}\right) &= T\left(x\begin{bmatrix} 1 \\ 0 \end{bmatrix} + y\begin{bmatrix} 0 \\ 1 \end{bmatrix}\right) \\ &= T\left(x\begin{bmatrix} 1 \\ 0 \end{bmatrix}\right) + T\left(y\begin{bmatrix} 0 \\ 1 \end{bmatrix}\right) \\ &= xT\left(\begin{bmatrix} 1 \\ 0 \end{bmatrix}\right) + yT\left(\begin{bmatrix} 0 \\ 1 \end{bmatrix}\right) \\ &= xc_1 + yc_2 \end{aligned}$$

for some real numbers $c_1 = T\left(\begin{bmatrix} 1 \\ 0 \end{bmatrix}\right)$ and $c_2 = T\left(\begin{bmatrix} 0 \\ 1 \end{bmatrix}\right)$. So how the linear functional $T \in (\mathbb{R}^2)^*$ acts on any vector $\begin{bmatrix} x \\ y \end{bmatrix} \in \mathbb{R}^2$ is determined entirely by how it acts on the basis vectors $\begin{bmatrix} 1 \\ 0 \end{bmatrix}$ and $\begin{bmatrix} 0 \\ 1 \end{bmatrix}$ of $\mathbb{R}^2$. Let us define two linear functionals T_1 and T_2 by how they act on these basis vectors as follows

$$T_1\left(\begin{bmatrix} 1 \\ 0 \end{bmatrix}\right) = 1 \text{ and } T_1\left(\begin{bmatrix} 0 \\ 1 \end{bmatrix}\right) = 0$$

and

$$T_2\left(\begin{bmatrix} 1 \\ 0 \end{bmatrix}\right) = 0 \text{ and } T_2\left(\begin{bmatrix} 0 \\ 1 \end{bmatrix}\right) = 1.$$

Question 1.8 Let T_1 and T_2 be the linear functionals on $\mathbb{R}^2$ defined above.

(a) Show $T_1\left(\begin{bmatrix} x \\ y \end{bmatrix}\right) = x$ and $T_2\left(\begin{bmatrix} x \\ y \end{bmatrix}\right) = y$
(b) Is it possible to write the above linear functional T as a linear combination of T_1 and T_2?
(c) Show that $\{T_1, T_2\}$ is a basis for $(\mathbb{R}^2)^*$, the set of linear functionals on $\mathbb{R}^2$.

Since the dual space is a vector space then of course the dual space has a basis. Using our standard notation for the Euclidian unit vectors,

$$e_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \text{ and } e_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

we could write $T_1(e_1) = 1$, $T_2(e_1) = 0$, $T_1(e_2) = 0$, $T_2(e_2) = 1$. Using a slightly different notation we could summarize this as $T_i(e_j) = \delta_{ij},\ i, j = 1, 2$, where δ_{ij} is called the **Kronecker delta function**, which is defined as

$$\delta_{ij} = \begin{cases} 1 & i = j \\ 0 & i \neq j. \end{cases}$$

Since $T_i(e_j) = \delta_{ij}$ then T_1 is said to be **dual** to the vector e_1 and T_2 is said to be dual to the vector e_2 and $\{T_1, T_2\}$ is said to be the **dual basis** of $\{e_1, e_2\}$.

Now for a little notation, often we write T_1 as e^1 and T_2 as e^2. Note that instead of subscripts we are using superscripts. So we have $\{e_1, e_2\}$ is the standard basis for $\mathbb{R}^2$ and $\{e^1, e^2\}$ is its dual basis, which is a basis for the dual space $(\mathbb{R}^2)^*$. Then we often write $e^i(e_j) = \delta^i_j$, which is exactly the Kronecker delta function defined above, but with one of the subscripts written as a superscript to match with the way the superscripts and subscripts are written on the basis and dual basis elements.

Also, notice that $e^i(e_j)$ is basically function notation, the linear functional e^i is eating as its argument the vector e_j. This is sometimes also written as $\langle e^i, e_j \rangle$ instead.

Also recall that earlier we said elements of vector spaces like $\mathbb{R}^2$ would always be written as column vectors, a convention that is often not used in calculus classes where vectors are often written as row vectors for convenience. The reason that we adhere to that convention here is because elements of dual spaces like $(\mathbb{R}^2)^*$ will always be written as row vectors. Since $e^1 = 1 \cdot e^1 + 0 \cdot e^2$ we will write $e^1 = [1, 0]$. Similarly, since $e^2 = 0 \cdot e^1 + 1 \cdot e^2$ we have $e^2 = [0, 1]$. This allows us to use matrix multiplication:

$$e^1(e_1) = \langle e^1, e_1 \rangle = [1, 0] \begin{bmatrix} 1 \\ 0 \end{bmatrix} = 1$$

$$e^1(e_2) = \langle e^1, e_2 \rangle = [1, 0] \begin{bmatrix} 0 \\ 1 \end{bmatrix} = 0$$

$$e^2(e_1) = \langle e^2, e_1 \rangle = [0, 1] \begin{bmatrix} 1 \\ 0 \end{bmatrix} = 0$$

$$e^2(e_2) = \langle e^2, e_2 \rangle = [0, 1] \begin{bmatrix} 0 \\ 1 \end{bmatrix} = 1$$

Notice that this matches what we would have obtained using the Kronecker delta notation. Just as any element v of $\mathbb{R}^2$ can be written as a linear combination of e_1 and e_2 and as a column vector,

$$v = xe_1 + ye_2 = \begin{bmatrix} x \\ y \end{bmatrix},$$

any element α of $(\mathbb{R}^2)^*$ can be written as a linear combination of e^1 and e^2 and as a row vector

$$\alpha = ae^1 + be^2 = [a, b].$$

Thus we can use matrix multiplication to find

$$\alpha(v) = \langle \alpha, v \rangle = [a, b] \begin{bmatrix} x \\ y \end{bmatrix} = ax + by.$$

All of this of course generalizes to any vector space V and you will often (though not always) see the basis elements of V written with subscripts and the basis elements of V^* written with superscripts. For example, if $\{v_1, v_2, \ldots, v_n\}$ is the basis of some vector space V then we would write the dual basis as $\{v^1, v^2, \ldots, v^n\}$, which is a basis of V^*. Thus any $v \in V$ can be written as a linear combination of $\{v_1, v_2, \ldots, v_n\}$ and hence as a column vector and any $\alpha \in V^*$ can be written as a linear combination of $\{v^1, v^2, \ldots, v^n\}$ and hence as a row vector. This allows $\alpha(v)$ to be computed using matrix multiplication as a row vector multiplied by a column vector. This also explains why we always write elements of the vector space as a column vector, because elements of the dual space are written as row vectors and it is very important to always distinguish between vector space and dual space elements. In vector calculus classes the dual space is not introduced so for convenience vectors are sometimes written as row vectors. We will never do that in this book.

Question 1.9 Consider the vector space $\mathbb{R}^3$ with the standard Euclidian basis

$$e_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \ e_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \ e_3 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$$

and let $v = \begin{bmatrix} x \\ y \\ z \end{bmatrix} = xe_1 + ye_2 + ze_3 \in \mathbb{R}^3$. Let $(\mathbb{R}^3)^*$ be the dual space of $\mathbb{R}^3$ with the dual basis

$$e^1 = [1, 0, 0],\ e^2 = [0, 1, 0],\ e^3 = [0, 0, 1].$$

Now consider the linear functional

$$f\left(\begin{bmatrix} x \\ y \\ z \end{bmatrix}\right) = 3x + 2y - 6z.$$

(a) Verify that f is indeed a linear functional.
(b) Let $v = \begin{bmatrix} -3 \\ 4 \\ 7 \end{bmatrix}$. Find $f(v)$.
(c) Find $f(e_i)$ for $i = 1, 2, 3$.
(d) Write f as a linear combination of e^1, e^2, e^3.
(e) Write $f \in (\mathbb{R}^3)^*$ as a row vector.
(f) Write $f(v)$ using matrix multiplication. (That is, as a row vector "f" times a column vector v.)

Question 1.10 Consider the vector space $\mathbb{R}^4$ with the standard Euclidian basis e_1, e_2, e_3, e_4 and the dual space $(\mathbb{R}^4)^*$ with the standard dual basis e^1, e^2, e^3, e^4. Let $v = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} \in \mathbb{R}^4$ and define the linear functional f by $f(v) = 5x_1 - 2x_2 + 7x_3 - 6x_4$.

(a) Verify that $f \in (\mathbb{R}^4)^*$. (That is, that f is a linear functional.)
(b) Find $f(e_i)$ for $i = 1, 2, 3, 4$ and use this to write f as a row vector.
(c) Use the row vector representing f to find

$$f\left(\begin{bmatrix} -7 \\ 3 \\ 2 \\ -1 \end{bmatrix}\right) \text{ and } f\left(\begin{bmatrix} 1/2 \\ -4 \\ -2/3 \\ 5 \end{bmatrix}\right).$$

Let us take a closer look at the dual basis elements. Using $(\mathbb{R}^4)^*$ as our example we have the dual basis elements e^1, e^2, e^3, e^4. Consider the following vector in $\mathbb{R}^4$,

$$v = \begin{bmatrix} 2 \\ -5 \\ 7 \\ -9 \end{bmatrix}.$$

Let us see how the dual basis elements act on v. We have

$$e^1(v) = [1, 0, 0, 0] \begin{bmatrix} 2 \\ -5 \\ 7 \\ -9 \end{bmatrix} = 2.$$

Similarly, we have $e^2(v) = -5$, $e^3(v) = 7$, and $e^4(v) = -9$. We can see that in essence the dual basis elements "pick out" the corresponding vector coordinates; e^1 picks out the first coordinate, e^2 picks out the second coordinate, e^3 picks out the third coordinate, and e^4 picks out the fourth coordinate.

Now we will explore one of the cartoons that is often used to try to picture linear functionals. We use the word cartoon deliberately. What we are developing are imperfect mental pictures or images so that we can try to visualize what is happening. It is important to remember that these pictures are imprecise, yet they allow us to develop our intuition regarding mathematical concepts. In our experience, the cartoon representation that we will start to develop here is most often seen in physics. We will start by using $\mathbb{R}^2$ and $(\mathbb{R}^2)^*$ as our example spaces.

Let $e_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$ and $e_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$ be a basis for the vector space $\mathbb{R}^2$. This basis *induces* the **dual basis** $e^1 = [1, 0]$ and $e^2 = [0, 1]$ of $(\mathbb{R}^2)^*$. Consider the following vectors in $\mathbb{R}^2$:

$$\begin{bmatrix} 1 \\ 3 \end{bmatrix}, \begin{bmatrix} 1 \\ 2 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ -1 \end{bmatrix}, \begin{bmatrix} 1 \\ -2 \end{bmatrix}, \begin{bmatrix} 1 \\ -3 \end{bmatrix}.$$

It is clear that e^1 applied to each of these vectors gives 1. Now consider the vectors:

$$\begin{bmatrix} 2 \\ 3 \end{bmatrix}, \begin{bmatrix} 2 \\ 2 \end{bmatrix}, \begin{bmatrix} 2 \\ 1 \end{bmatrix}, \begin{bmatrix} 2 \\ 0 \end{bmatrix}, \begin{bmatrix} 2 \\ -1 \end{bmatrix}, \begin{bmatrix} 2 \\ -2 \end{bmatrix}, \begin{bmatrix} 2 \\ -3 \end{bmatrix}.$$

We have e^1 sending all of these vectors to 2. In the same manner e^1 sends any vector of the form $\begin{bmatrix} 0 \\ y \end{bmatrix}$, where $y \in \mathbb{R}$, to 0. It sends any vector of the form $\begin{bmatrix} -1 \\ y \end{bmatrix}$ to -1, et cetera. Sketching these vectors as in Fig. 1.3 we can see that the endpoints of the vectors that e^1 sends to 1 lie on the vertical line $x = 1$, the endpoints of the vectors that e^1 sends to 2 lie on the vertical line $x = 2$, the endpoints of the vectors that e^1 sends to 0 lie on the vertical line $x = 0$, and the endpoints of the vectors that e^1 sends to -1 lie on the vertical line $x = -1$, et cetera. This gives a way to construct a cartoon image of the linear functional e^1 by drawing the integer *level sets* on the vector space $\mathbb{R}^2$ as in Fig. 1.4. Consider the vectors shown in this figure,

$$\mathbf{A} = \begin{bmatrix} 3.5 \\ 1.5 \end{bmatrix}, \quad \mathbf{B} = \begin{bmatrix} -3.5 \\ 1.5 \end{bmatrix}, \quad \mathbf{C} = \begin{bmatrix} 1.9 \\ -1 \end{bmatrix}, \quad \mathbf{D} = \begin{bmatrix} -1.5 \\ -1.5 \end{bmatrix},$$

Clearly $e^1(\mathbf{A}) = 3.5$, $e^1(\mathbf{B}) = -3.5$, $e^1(\mathbf{C}) = 1.9$, and $e^1(\mathbf{D}) = -1.5$.

Using this cartoon picture of e^1 only gives us approximate answers, but provides a very nice way to visualize and understand what e^1 does. Sometimes in physics you will hear it said that "the value of the vector," as given by the linear functional, is the number of level sets of the linear functional that the vector pierces. In essence this picture of the linear functional e^1 works by counting the number of vertical level sets that the vectors "pierces" or goes through, while still keeping track of the sign. That means we count the negative level sets as negative. If the endpoint of a vector lies on a level set than that level set is considered "pierced" and is included in the count. In Fig. 1.4, for example, the vector **A** pierces 3 level sets and the vector **C** pierces one level set, while the vector **B** pierces -3 level sets and the vector **D** pierces -1. Thus,

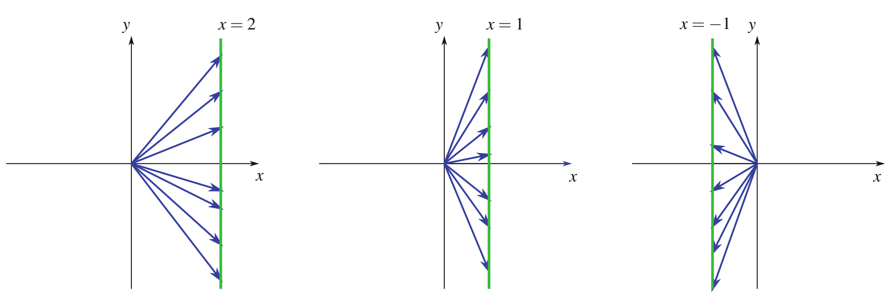


Fig. 1.3 Sketches of vectors with endpoints with x values of 2, 1, and -1 respectively

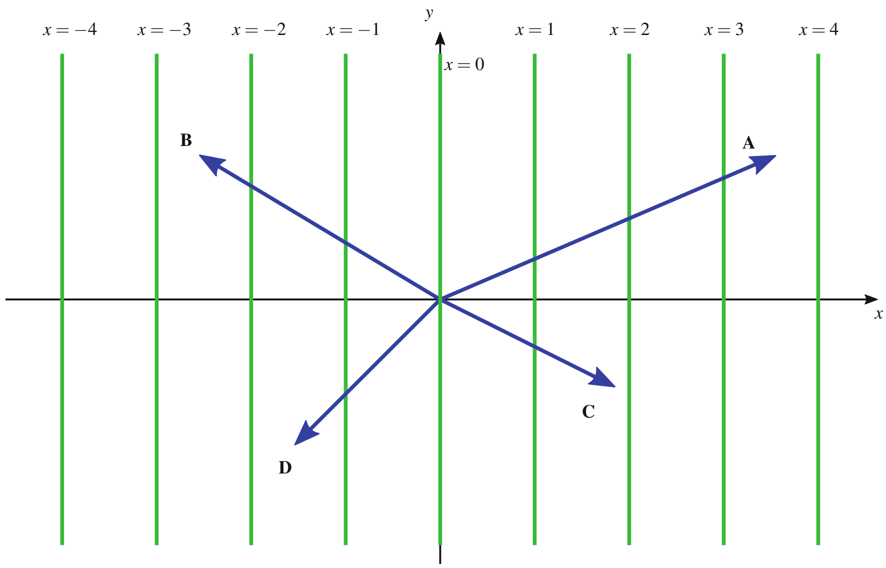


Fig. 1.4 The "level sets" on the vector space $\mathbb{R}^2$ that depict the linear functional e^1 along with vectors **A**, **B**, **C**, and **D**

according to our cartoon picture we have $e^1(\mathbf{A}) = 3$, $e^1(\mathbf{B}) = -3$, $e^1(\mathbf{C}) = 1$, and $e^1(\mathbf{D}) = -1$, which are close to the actual values. A similar analysis results in e^2 being pictured as in Fig. 1.5.

Question 1.11 Consider the linear functional $e^2 \in (\mathbb{R}^2)^*$.

(a) Give five vectors that e^2 send to 1; to 2; to 0; to -1;
(b) Sketch these vectors and use their endpoints to draw the level sets.

Think about the two cartoons of e^1 and e^2. The xy-coordinate axes represent the vector space $\mathbb{R}^2$. Yet e^1 and e^2 are actually in the dual space $(\mathbb{R}^2)^*$. In some sense what we are doing is strange, we are attempting to draw a picture of e^1 and e^2, which are elements of $(\mathbb{R}^2)^*$, in $\mathbb{R}^2$. Consider how different our picture of e^1 drawn in $\mathbb{R}^2$ is from our picture of e^1 drawn in $(\mathbb{R}^2)^*$ in Fig. 1.6. Similarly we picture of e^2 drawn in $\mathbb{R}^2$ and e^2 drawn in $(\mathbb{R}^2)^*$ in Fig. 1.7.

Now we take a look at a slightly more complicated linear functional $f = 2e^1 + e^2 \in (\mathbb{R}^2)^*$. First let us ask ourselves what vectors v does f send to 1? That is, find v such that $f(v) = 1$;

$$1 = f\left(\begin{bmatrix} x \\ y \end{bmatrix}\right) = 2e^1\left(\begin{bmatrix} x \\ y \end{bmatrix}\right) + e^2\left(\begin{bmatrix} x \\ y \end{bmatrix}\right) = 2x + y$$
$$\Rightarrow 2x + y = 1 \Rightarrow y = -2x + 1.$$

In other words, $\Rightarrow f\left(\begin{bmatrix} x \\ -2x+1 \end{bmatrix}\right) = 1$ where x is any real number. We have also found out that the endpoints of the vectors v that the linear functional f sends to 1 determine the line $y = -2x + 1$. Thus our level set for 1 is given by the equation $y = -2x + 1$.

Question 1.12 Find the vectors that f sends to the values 2, 3, 0, -1, and -2. Use these vectors to determine the level sets of for these values.

Sketching the level sets of f that we found in the question we obtain the picture of f on $\mathbb{R}^2$ as in Fig. 1.8. Consider the two vectors shown,

$$\mathbf{A} = \begin{bmatrix} 1/2 \\ 1 \end{bmatrix} \quad \text{and} \quad \mathbf{B} = \begin{bmatrix} 1 \\ -3 \end{bmatrix}.$$

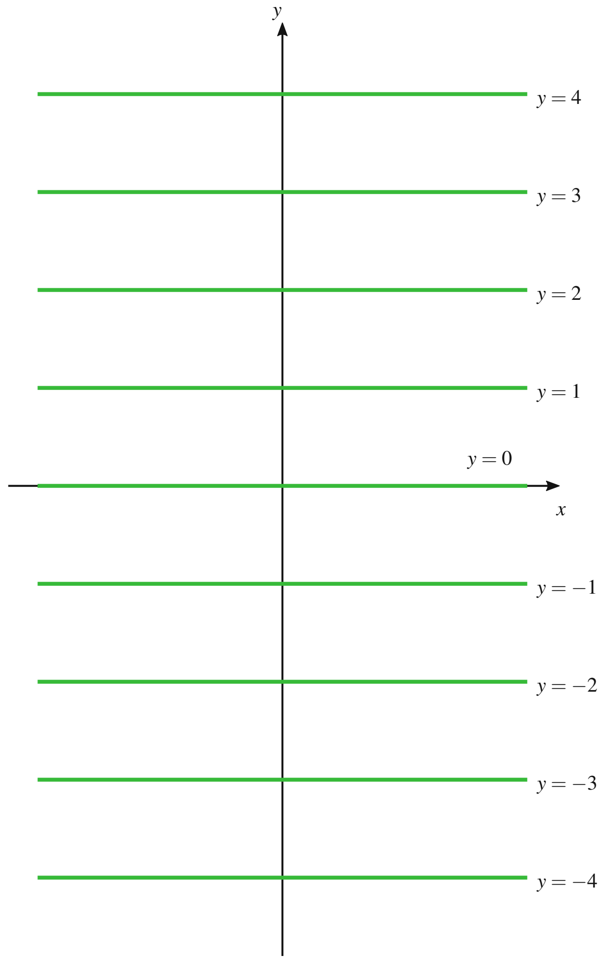


Fig. 1.5 The "level sets" on the vector space $\mathbb{R}^2$ that depict the linear functional e^2

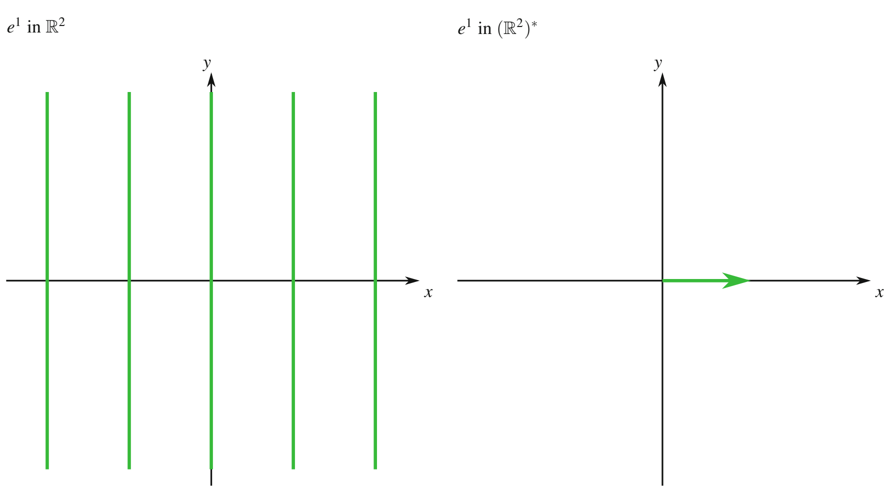


Fig. 1.6 The level sets that depict the linear functional e^1 drawn on the vector space $\mathbb{R}^2$ (left) and the linear functional e^1 drawn as a vector in the vector space $(\mathbb{R}^2)^*$ (right)

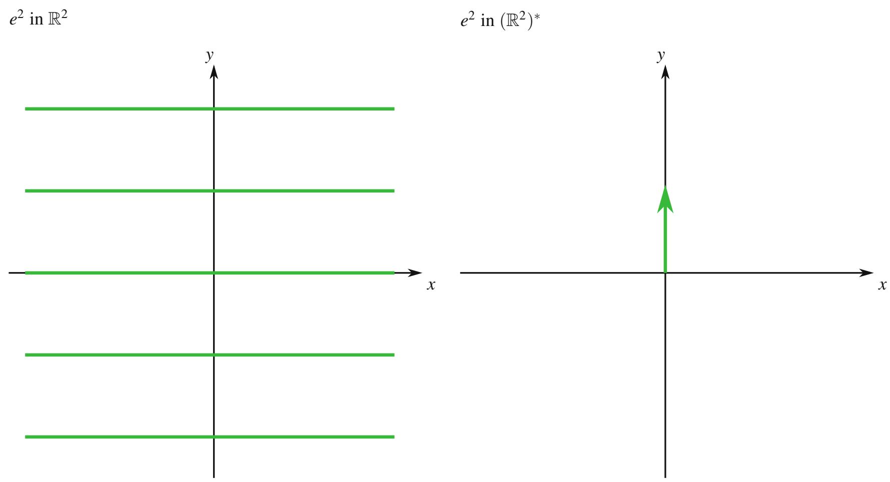


Fig. 1.7 The level sets that depict the linear functional e^2 drawn on the vector space $\mathbb{R}^2$ (left) and the linear functional e^2 drawn as a vector in the vector space $(\mathbb{R}^2)^*$ (right)

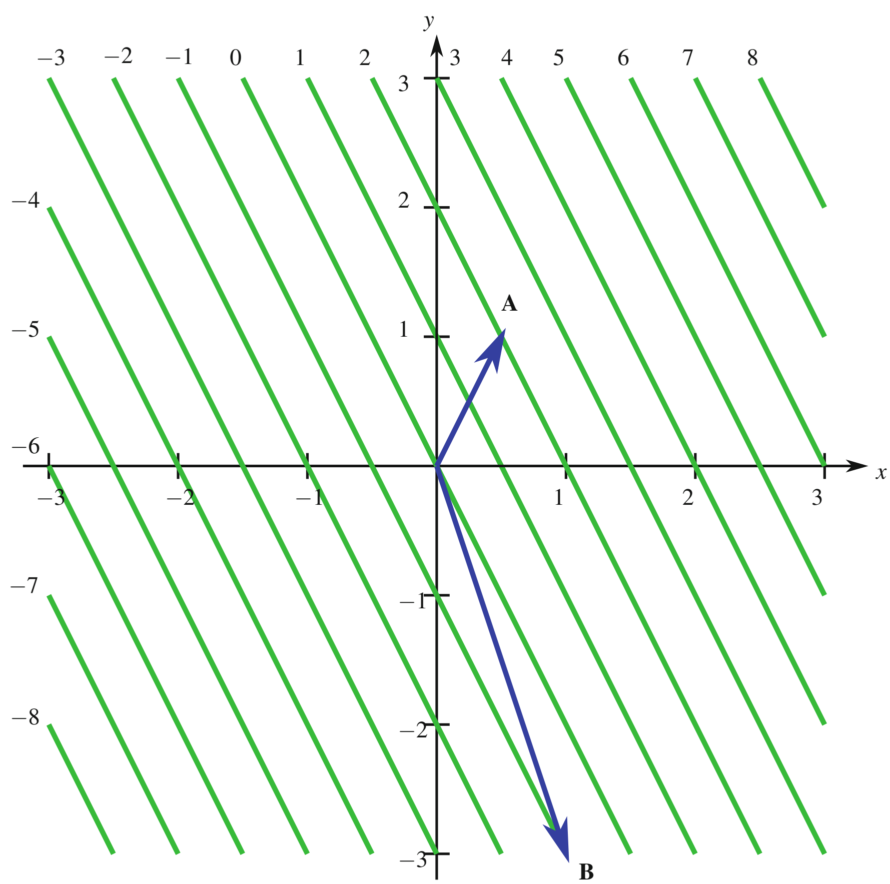


Fig. 1.8 The level sets that depict the linear functional $f = 2e^1 + e^2$ drawn on the vector space $\mathbb{R}^2$, along with the vectors **A** and **B**

It is clear that **A** pierces two level sets and so according to our picture $f(\mathbf{A}) = 2$ and **B** pierces -1 level set so according to our picture $f(\mathbf{B}) = -1$. Double checking our picture, computationally we have

$$f(\mathbf{A}) = 2e^1(\mathbf{A}) + e^1(\mathbf{A}) = 2\left(\frac{1}{2}\right) + 1 = 2$$

$$f(\mathbf{B}) = 2e^1(\mathbf{B}) + e^2(\mathbf{B}) = 2(1) - 3 = -1,$$

which is what we would expect. Our cartoon pictures of linear functionals can readily be extended to $\mathbb{R}^3$. We will not do that here but we will consider this case in a future chapter in a slightly different setting.

1.2 Volume and Determinants

Determinants of matrices have a wide range of properties and uses. However, for us, the most useful thing about the determinant will be how it relates to volume: the determinant of a matrix gives the **signed volume** of the parallelepiped that is generated by the vectors given by the matrix columns. The only difference with our intuitive notion of volume is that here the volume has either a positive or negative sign attached to it. So, while we are used to thinking of volumes as being positive (or zero) they can in fact be negative as well. And as if that wasn't strange enough, the truly bizarre thing is that our intuitive ideas of how a volume should behave force that sign upon us.

Determinants are introduced in a variety of different ways and, quite frankly, many of those ways are not at all clear. Since really understanding how determinants are related to volume is important to us we will actually use our intuitive understanding of volumes and three properties that we expect volume to have to derive the determinant. What is so amazing is that these three properties of how volume should behave uniquely determines the determinant.

So, how do we expect volume to behave? First, we expect a unit cube (or unit n-cube for that matter) to have a volume of one. This comes form the fact that 1^3 (or 1^n) equals 1. This, in effect, normalizes our volumes. Second, we expect a degenerate parallelepiped to have volume zero. Basically, in n dimensions any $n-1$ dimensional object has zero n-dimensional volume. For example, a two-dimensional unit square in $\mathbb{R}^3$ has zero three-dimensional volume, even if it has a two-dimensional volume of 1 in $\mathbb{R}^2$. Third, we expect volumes to be linear. Linearity means several things, but one of these is that if we hold $n - 1$ sides of an n-dimensional parallelepiped fixed and scale the remaining side by some factor then the volume of the parallelepiped changes by that factor. So, if we doubled the length of one of the sides of a 3-cube then the volume will double as well. Now, with the general ideas in hand let us move onto the actual mathematics.

Suppose we have a parallelepiped $\mathscr{P} \in \mathbb{R}^n$ whose edges are given by $v_1, v_2, \ldots, v_n \in \mathbb{R}^n$. We will say that the parallelepiped $\mathscr{P}$ is the **span** of the vectors $v_1, v_2, \ldots, v_n$ and write $\mathscr{P} = \text{span}\{v_1, v_2, \ldots, v_n\}$. Notice that this is a different use of the word span than in the last section where we said the basis of a vector space spans the vector space. We want to find a function $D : \mathbb{R}^{n\times n} \to \mathbb{R}$ which takes $v_1, v_2, \ldots, v_n$ (or similarly, the matrix M, which has $v_1, v_2, \ldots, v_n$ as its columns) to a real number which is the volume of $\mathscr{P}$. Allowing our notation to be flexible we can write both $D(M)$ or $D(v_1, v_2, \ldots, v_n)$. Now, we present our three intuitive properties of volume in mathematical rigor.

Property 1 $D(I) = 1$, where $I = [e_1, e_2, \ldots, e_n]$, or the identity matrix.

This says that the volume of the unit n-cube is 1. Examples of this for two and three dimensions are shown in Fig. 1.9. Notice that the unit n-cube is given by $\text{span}\{e_1, e_2, \ldots, e_n\}$ and the identity matrix I is given by $I = [e_1, e_2, \ldots, e_n]$. Thus we often also denote unit n-cubes with I.

Property 2 $D(v_1, v_2, \ldots, v_n) = 0$ if $v_i = v_j$ for any $i \neq j$.

This condition says that if any two edges of the n-parallelepiped are the same, that is, if the parallelepiped is degenerate, then the volume is zero. Figure 1.10 shows this for a two-dimensional parallelepiped in three-dimensional space.

Property 3 D is linear. That is, for any $j = 1, \ldots, n$

$$\begin{aligned}D(v_1, \ldots, v_{j-1}, v + cw, v_{j+1}, \ldots, v_n) &= D(v_1, \ldots, v_{j-1}, v, v_{j+1}, \ldots, v_n)\\ &\quad + cD(v_1, \ldots, v_{j-1}, w, v_{j+1}, \ldots, v_n).\end{aligned}$$

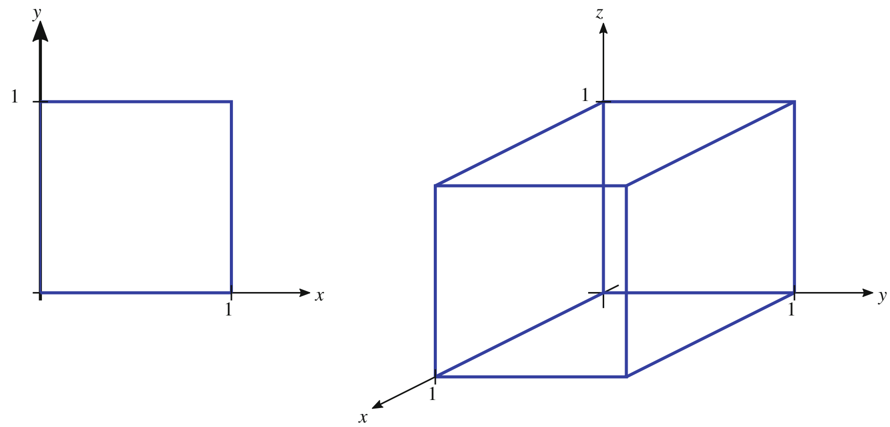


Fig. 1.9 The two-dimensional volume, also known as area, of a unit square is one (left) and the three-dimensional volume of a unit cube is also one (right)

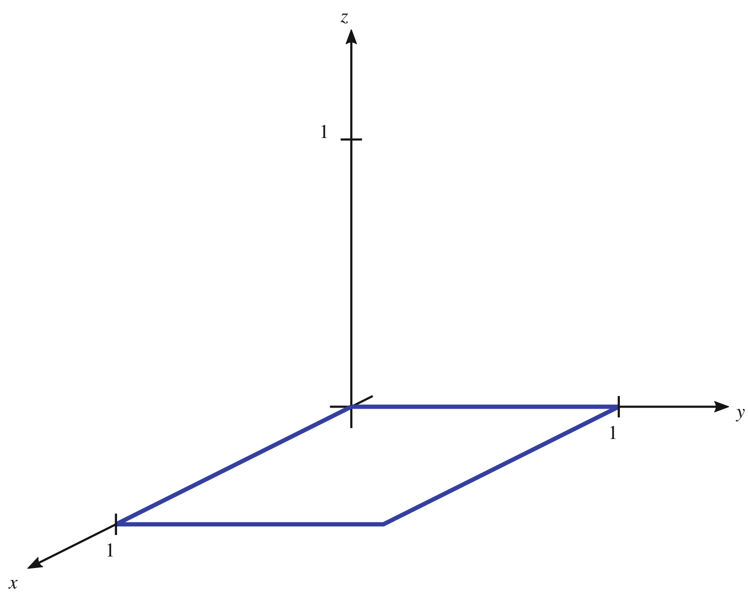


Fig. 1.10 The three-dimensional volume of a two-dimensional parallelepiped in three dimensional space is zero

In actually, this says slightly more than just "if we scale the length of one side the volume scales as well." This property includes shears as well, but from Euclidian geometry we know shears do not change volumes either. A scaling is shown in Fig. 1.11 for a two-dimensional parallelepiped.

We will use these three properties of volume to derive a specific formula for D in terms of the entries of $M = [v_1, v_2, \ldots, v_n]$. By deriving a specific formula for D we will have shown that D both exists and is unique. And of course, this formula for D will be the formula you know and love, or maybe detest, for the determinant of a matrix. There is unfortunately no getting around the fact that this formula is both horridly ugly and a real hassle to compute. But such is life (and math) at times.

We will first proceed by using the three properties above to determine several other properties that will be essential in deriving the formula for D. The first thing we will show is the surprising fact that volumes are signed.

Property A D is alternating, that is, if we switch any two vectors the sign changes.

$$D(v_1, \ldots, v_i, \ldots, v_j, \ldots, v_n) = -D(v_1, \ldots, v_j, \ldots, v_i, \ldots, v_n).$$

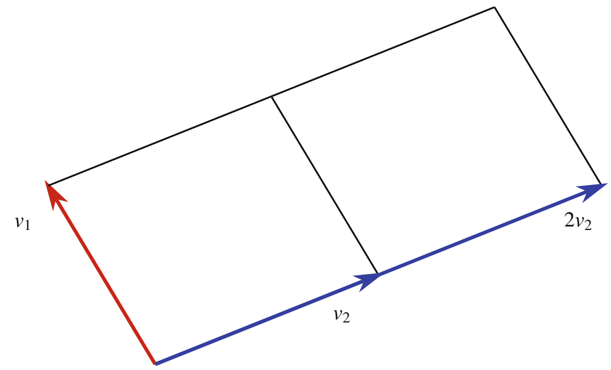


Fig. 1.11 Doubling the length of one side of this two-dimensional parallelepiped doubles the two-dimensional volume

Proof In an attempt to keep notation simple we could have simply written the above as $D(v_i, v_j) = -D(v_j, v_i)$ since v_i and v_j are the only vectors we are concerned with; the others remain unchanged. We use this simplified notation below.

$$\begin{aligned}
D(v_i, v_j) &= D(v_i, v_j) + D(v_i, v_i) && \text{Property 2} \\
&= D(v_i, v_j + v_i) && \text{Property 3} \\
&= D(v_i, v_j + v_i) - D(v_j + v_i, v_j + v_i) && \text{Property 2} \\
&= D(-v_j, v_j + v_i) && \text{Property 3} \\
&= -D(v_j, v_j + v_i) && \text{Property 3} \\
&= -D(v_j, v_j) - D(v_j, v_i) && \text{Property 3} \\
&= -D(v_j, v_i) && \text{Property 2}
\end{aligned}$$

□

Property B If the vectors $v_1, v_2, \ldots, v_n$ are linearly dependant, then

$$D(v_1, v_2, \ldots, v_n) = 0.$$

Proof By the definition of linearly independent vectors, at least one of the vectors can be written as a linear combination of the others. Without loss of generality suppose we can write $v_1 = c_2 v_2 + \ldots + c_n v_n$, where the c_i are scalars. By repeated uses of Property 3 we have

$$\begin{aligned}
D(v_1, v_2, \ldots, v_n) &= D(c_2 v_2 + \ldots + c_n v_n, v_2, \ldots v_n) \\
&= c_2 D(v_2, v_2, \ldots, v_n) + c_3 D(v_3, v_2, \ldots, v_n) + \ldots + c_n D(v_n, v_2, \ldots, v_n) \\
&= 0
\end{aligned}$$

where the last equality comes from repeated uses of Property 2. □

What did we mean by the phrase *without loss of generality* in the above proof? Notice that we assumed that we could write the first vector v_1 as a linear combination of the other vectors. Even though we did the computation in the proof for the specific example where $v_1 = c_2 v_2 + \ldots + c_n v_n$, had we changed our assumption and decided that instead $v_i = c_1 v_1 + \ldots + c_{i-1} v_{i-1} + c_{i+1} v_{i+1} + \ldots + c_n v_n$ for any v_i then the computation would have been essentially the same and given us the same result. Of course we are not going to redo what is essentially the same computation n times, that would be a waste of time and space and effort. The mathematician's way to get around this is to use the phrase *without loss of generality* and then do the computation for one specific case. It is understood that all the other cases would be exactly the same and give the same result if you chose to actually carry out the computation.

Property C Adding a multiple of one vector to another does not change the determinant.

Proof Without loss of generality suppose we are adding c times v_j to v_i.

$$\begin{aligned} D(v_1, \dots, v_i + cv_j, \dots, v_n) &= D(v_1, \dots, v_i, \dots, v_n) + cD(v_1, \dots, v_j, \dots, v_n) \\ &= D(v_1, \dots, v_i, \dots, v_n) \end{aligned}$$

Where the last equality follows from Property 2 since vector v_j is repeated. □

Question 1.13 Explain how the phrase *without loss of generality* is used in the proof of Property C.

With these properties in hand we are almost ready to derive the formula for determinant. The final ingredient we need to do this is permutations. Therefore we spend a few moments reviewing permutations. A **permutation** of a set $\{1, \dots, n\}$ is a bijective (one-to-one and onto) function that maps $\{1, \dots, n\}$ to itself. The set of permutations of $\{1, \dots, n\}$ is usually denoted by S_n. Thus each element of S_n is in fact a permutation function and there are $n!$ elements in S_n.

Question 1.14 Explain what one-to-one and onto mean in the context of permutation functions.

It is fairly straightforward to see that there are $n!$ elements in S_n. Let $\sigma \in S_n$. Then $\sigma : \{1, \dots, n\} \to \{1, \dots, n\}$. So there are n possible choices for $\sigma(1)$, any element of $\{1, \dots, n\}$. $\sigma(2)$ can be any element of $\{1, \dots, n\}$ except for $\sigma(1)$, since σ is one-to-one, so there are $n - 1$ possible choices for it. Similarly, $\sigma(3)$ can be any element of $\{1, \dots, n\}$ except for $\sigma(1)$ or $\sigma(2)$, so there are $n - 2$ possible choices. This continues until $\sigma(n)$ is determined by whatever element is left over. In summary, the total number of possible permutations of $\{1, \dots, n\}$ is $n \cdot (n - 1) \cdot (n - 2) \cdot \dots \cdot 1 = n!$.

We denote a particular permutation σ as

$$\begin{pmatrix} 1 & 2 & \cdots & n \\ \downarrow & \downarrow & \cdots & \downarrow \\ \sigma(1) & \sigma(2) & \cdots & \sigma(n) \end{pmatrix}.$$

While this notation, called *Cauchy's two-line notation*, is clear, it is also cumbersome. Taking only the second row of this array gives the more succinct *Cauchy's one-line notation*, $\big(\sigma(1), \sigma(2), \dots, \sigma(n)\big)$. Here σ is given as an ordered list. Thus, if $\sigma \in S_3$ we write $(2, 1, 3)$ to mean

$$\begin{pmatrix} 1 & 2 & 3 \\ \downarrow & \downarrow & \downarrow \\ 2 & 1 & 3 \end{pmatrix}.$$

Note, some books use the *cycle notation* where $(2, 1, 3)$ would mean

$$\begin{pmatrix} 1 & 2 & 3 \\ \downarrow & \downarrow & \downarrow \\ 3 & 1 & 2 \end{pmatrix}.$$

That is, the first element of the list is mapped to the second, the second mapped to the third, and the last mapped to the first, $2 \mapsto 1 \mapsto 3 \mapsto 2$. Be careful when looking in other books to make sure you know which notation they are using. In cycle notation the σ above would be written simply as $(1, 2)$ where the fixed element 3 that maps to itself is not written down. However, here we will stick to either Cauchy's one-line or two-line notations.

A permutation in which only two elements are exchanged is called a **transposition**. We will use a special notation for transpositions, $\tau_{i,j}$ means that i and j are exchanged and all other elements are fixed. We illustrate this with an example that we hope will make the idea clear; the transposition $\tau_{3,7} \in S_8$ is given by

$$\begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 \\ \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow \\ 1 & 2 & 7 & 4 & 5 & 6 & 3 & 8 \end{pmatrix}.$$

Next we notice that the composition of two permutations is also a permutation. For the purpose of this example we use S_5, which is large enough to not be trivial but small enough to be easily handled. Suppose that $\sigma \in S_5$ is given by

$$\begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ \downarrow & \downarrow & \downarrow & \downarrow & \downarrow \\ 4 & 3 & 1 & 5 & 2 \end{pmatrix}.$$

and the transposition $\tau_{1,3} \in S_5$ is given by

$$\begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ \downarrow & \downarrow & \downarrow & \downarrow & \downarrow \\ 3 & 2 & 1 & 4 & 5 \end{pmatrix}.$$

We write the composition $\tau_{1,3} \circ \sigma$ as

$$\begin{array}{cccccc} & 1 & 2 & 3 & 4 & 5 \\ \sigma & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow \\ & 4 & 3 & 1 & 5 & 2 \\ \tau_{1,3} & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow \\ & 4 & 1 & 3 & 5 & 2. \end{array}$$

Notice that since the middle row is not in ascending order this is not exactly Cauchy's two line notation, but what is happening should be clear. The transposition $\tau_{1,3}$ is marked in red to help make the diagram easier to read. Next, it should not be hard to see that one can find a series of transpositions that will transform any given permutation to the identity. For example, the permutation $\tau_{3,4} \circ \tau_{2,4} \circ \tau_{2,3} \circ \tau_{2,5} \circ \tau_{1,4} \circ \tau_{1,3} \circ \sigma$ is the identity:

$$\begin{array}{cccccc} & 1 & 2 & 3 & 4 & 5 \\ \sigma & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow \\ & 4 & 3 & 1 & 5 & 2 \\ \tau_{1,3} & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow \\ & 4 & 1 & 3 & 5 & 2 \\ \tau_{1,4} & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow \\ & 1 & 4 & 3 & 5 & 2 \\ \tau_{2,5} & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow \\ & 1 & 4 & 3 & 2 & 5 \\ \tau_{2,3} & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow \\ & 1 & 4 & 2 & 3 & 5 \\ \tau_{2,4} & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow \\ & 1 & 2 & 4 & 3 & 5 \\ \tau_{3,4} & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow \\ & 1 & 2 & 3 & 4 & 5 \end{array}$$

but notice that $\tau_{4,5} \circ \tau_{3,4} \circ \tau_{2,3} \circ \tau_{1,4} \circ \sigma$,

$$\begin{array}{cccccc} & 1 & 2 & 3 & 4 & 5 \\ \sigma & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow \\ & 4 & 3 & 1 & 5 & 2 \\ \tau_{1,4} & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow \\ & 1 & 3 & 4 & 5 & 2 \\ \tau_{2,3} & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow \\ & 1 & 2 & 4 & 5 & 3 \\ \tau_{3,4} & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow \\ & 1 & 2 & 3 & 5 & 4 \\ \tau_{4,5} & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow \\ & 1 & 2 & 3 & 4 & 5 \end{array}$$

also gives the identity. As does $\tau_{2,3} \circ \tau_{1,3} \circ \tau_{4,5} \circ \tau_{2,5} \circ \tau_{2,4} \circ \tau_{4,5} \circ \tau_{1,4} \circ \tau_{2,4} \circ \sigma$,

$$\begin{array}{cccccc}
 & 1 & 2 & 3 & 4 & 5 \\
\sigma & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow \\
 & 4 & 3 & 1 & 5 & 2 \\
\tau_{3,4} & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow \\
 & 3 & 4 & 1 & 5 & 2 \\
\tau_{1,4} & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow \\
 & 3 & 1 & 4 & 5 & 2 \\
\tau_{4,5} & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow \\
 & 3 & 1 & 5 & 4 & 2 \\
\tau_{2,4} & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow \\
 & 3 & 1 & 5 & 2 & 4 \\
\tau_{2,5} & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow \\
 & 3 & 1 & 2 & 5 & 4 \\
\tau_{4,5} & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow \\
 & 3 & 1 & 2 & 4 & 5 \\
\tau_{1,3} & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow \\
 & 1 & 3 & 2 & 4 & 5 \\
\tau_{2,3} & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow \\
 & 1 & 2 & 3 & 4 & 5.
\end{array}$$

So it is clear that the *number* of transpositions necessary to transform a permutation into the identity is not unique. In the first example six transpositions were required to transform σ into the identity, in the second example four transpositions were required, and in the third example eight transpositions were required. But notice that six, four, and eight are all even numbers. It turns out that the **parity** (evenness or oddness) is unique. This means that for any permutation $\sigma \in S_n$ the number of transpositions required to transform the permutation into the identity is either always even or always odd. This will not be proved here.

The **sign** of a permutation $\sigma \in S_n$ is a function $sgn : S_n \to \{+1, -1\}$. $sgn(\sigma) = +1$ if σ requires an even number of permutations and $sgn(\sigma) = -1$ if σ requires an odd number of permutations. You should be aware that sometimes the notation $sgn(\sigma) \equiv (-1)^{\sigma}$ is also used. The idea behind this notation is that when σ requires an even number of transpositions then $(-1)^{\text{even}} = 1$ and when σ requires an odd number of transpositions then $(-1)^{\text{odd}} = -1$.

And now with the permutation notation in hand we can take the next step in deriving the formula for the determinant. Suppose we have the unit vectors

$$e_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \quad e_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \quad \cdots \quad , \quad e_n = \begin{bmatrix} 0 \\ 0 \\ 0 \\ \vdots \\ 1 \end{bmatrix} \in \mathbb{R}^n$$

then $[e_1 \ \ e_2 \ \ \cdots \ \ e_n]$ is the identity matrix I and by Property 1 we have that $D\big([e_1 \ \ e_2 \ \ \cdots \ \ e_n]\big) = D(I) = 1$. Next, if $\sigma \in S_n$ then it defines a permutation of the unit vectors, so $e_{\sigma(i)}$ is the vector that has a 1 in the $\sigma(i)$-th row and zeros everywhere else. We can then define the matrix $E_\sigma = [e_{\sigma(1)} \ \ e_{\sigma(2)} \ \ \cdots \ \ e_{\sigma(n)}]$. Now we show the following property.

Property D $\;D\big(E_\sigma\big) = sgn(\sigma)$.

Proof Notice the similarity between $E_\sigma = [e_{\sigma(1)} \ \ e_{\sigma(2)} \ \ \cdots \ \ e_{\sigma(n)}]$ and the Cauchy one-line notation for σ, $\big(\sigma(1), \sigma(2), \ldots, \sigma(n)\big)$. A sequence of transformations that takes $\big(\sigma(1), \sigma(2), \ldots, \sigma(n)\big)$ to the identity permutation $(1, 2, \ldots, n)$ will also take $E_\sigma = [e_{\sigma(1)} \ \ e_{\sigma(2)} \ \ \cdots \ \ e_{\sigma(n)}]$ to the identity matrix I. By property A we have that each column exchange results in a multiplicative factor of (-1). Since $sgn(\sigma)$ keeps track of the parity of transpositions needed to transform σ into the identity permutation it is also the parity of the column changes needed to transform E_σ into the identity matrix I, so we have $D(E_\sigma) = sgn(\sigma) D(I) = sgn(\sigma)$. □

Now we have all the pieces necessary to find a formula that will give the volume of the parallelepiped spanned by n vectors. When these vectors are written as the columns of a matrix this volume formula is exactly the determinant of the matrix and is therefore denoted by D. First of all we let

$$v_1 = \begin{bmatrix} a_{11} \\ a_{21} \\ a_{31} \\ \vdots \\ a_{n1} \end{bmatrix}, \quad v_2 = \begin{bmatrix} a_{12} \\ a_{22} \\ a_{32} \\ \vdots \\ a_{n2} \end{bmatrix}, \quad \cdots \quad , \quad v_n = \begin{bmatrix} a_{1n} \\ a_{2n} \\ a_{3n} \\ \vdots \\ a_{nn} \end{bmatrix} \in \mathbb{R}^n.$$

We want to find the volume of the parallelepiped spanned by $v_1, v_2, \ldots, v_n$; that is, the determinant of the matrix that has these vectors as its columns.

$$\begin{aligned}
D\left(\begin{bmatrix} a_{11} & a_{12} & & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & & a_{nn} \end{bmatrix}\right) &= \sum_{i_1=1}^{n} a_{i_1 1}\, D\left(\begin{bmatrix} | & a_{12} & & a_{1n} \\ e_{i_1} & a_{22} & \cdots & a_{2n} \\ | & \vdots & & \vdots \\ | & a_{n2} & & a_{nn} \end{bmatrix}\right) \\
&= \sum_{i_1, i_2=1}^{n} a_{i_1 1}\, a_{i_2 2}\, D\left(\begin{bmatrix} | & | & a_{13} & & a_{1n} \\ e_{i_1} & e_{i_2} & a_{23} & \cdots & a_{2n} \\ | & | & \vdots & & \vdots \\ | & | & a_{n3} & & a_{nn} \end{bmatrix}\right) \\
&= \quad \vdots \\
&= \sum_{i_1, i_2, \cdots, i_n=1}^{n} a_{i_1 1}\, a_{i_2 2} \, \cdots \, a_{i_n n} \quad \underbrace{D\left(\begin{bmatrix} | & | & & | \\ e_{i_1} & e_{i_2} & \cdots & e_{i_n} \\ | & | & & | \end{bmatrix}\right)}_{\substack{=0 \text{ if } e_{i_j}=e_{i_k} \text{ for any } i_j, i_k \\ \text{so non-zero terms are when } e_{i_1}, \ldots, e_{i_n} \\ \text{are some permutation } \sigma \text{ of } \{1, \ldots, n\}}} \\
&= \sum_{\sigma \in S_n} a_{\sigma(1)1}\, a_{\sigma(2)2} \, \cdots \, a_{\sigma(n)n}\, D\left(\begin{bmatrix} | & | & & | \\ e_{\sigma(1)} & e_{\sigma(2)} & \cdots & e_{\sigma(n)} \\ | & | & & | \end{bmatrix}\right) \\
&= \sum_{\sigma \in S_n} a_{\sigma(1)1}\, a_{\sigma(2)2} \, \cdots \, a_{\sigma(n)n}\, sgn(\sigma) \\
&= \sum_{\sigma \in S_n} sgn(\sigma) \prod_{i=1}^{n} a_{\sigma(i)i}.
\end{aligned}$$

Notice that the first through nth equalities were Property D. Thus we arrive at our standard formula for the determinant,

Formula for the determinant of a matrix	$\begin{vmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{vmatrix} = \sum_{\sigma \in S_n} sgn(\sigma) \prod_{i=1}^{n} a_{\sigma(i)i}.$

Notice that here we are using a slight variation of matrix notation, straight lines | to either side of the array of numbers instead of brackets [or] to indicate the determinate of the matrix.

We would like to reiterate that we built this formula for the determinant up entirely based on our three intuitive ideas of what the volume of an n-parallelepiped generated by n vectors should be (along with a little permutation notation and knowing that permutation parity is unique.) By actually deriving this formula we have shown that there is a unique formula that satisfies properties 1–3.

Question 1.15 Our formula for determinant should certainly satisfy the three original properties that we used to derive the formula. Using the derived formula for determinant show that

(a) $D(I) = 1$,
(b) $D(A) = 0$ if any two columns are the same,
(c) if $n - 1$ columns are fixed then D is a linear function of the remaining column.

The **transpose** of a matrix A, denoted by A^T, is obtained by "reflecting across the diagonal." That is, the first column of A becomes the first row of A^T, the second column of A becomes the second row of A^T, etc. For example, if

$$A = \begin{bmatrix} 1 & 4 & 7 \\ 2 & 5 & 8 \\ 3 & 6 & 9 \end{bmatrix} \quad \text{then} \quad A^T = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix}.$$

Question 1.16 Using the formula for D derived above, show that $D(A) = D(A^T)$.

Question 1.17 Properties 1–3 and A-D were stated for columns. Since the columns of A become the rows of A^T and $D(A) = D(A^T)$, all of these properties can be reworded for rows instead of columns. Reword properties 1–3 and A-D for rows instead of columns.

1.3 Derivatives of Multivariable Functions

In this section we will introduce the idea of the derivative of a multivariable function. You may have seen some of these ideas in a multivariable calculus course, though a full introduction to these ideas usually happens in an introductory analysis course. Here we will simply provide the essentials necessary for this book.

Recall that if you have a function $f : \mathbb{R} \to \mathbb{R}$ the derivative of f at the point $x_0 \in \mathbb{R}$, denoted by $f'(x)$, is given by

$$f'(x_0) = \lim_{h \to 0} \frac{f(x_0 + h) - f(x_0)}{h}$$

if the limit exists. But notice

$$\begin{aligned} & f'(x_0) = \lim_{h \to 0} \frac{f(x_0 + h) - f(x_0)}{h} \\ \Rightarrow \quad & \lim_{h \to 0} \frac{f(x_0 + h) - f(x_0) - f'(x_0)h}{h} = 0 \\ \Rightarrow \quad & \lim_{x \to x_0} \frac{f(x) - f(x_0) - f'(x_0)(x - x_0)}{x - x_0} = 0 \\ \Rightarrow \quad & \lim_{x \to x_0} \frac{|f(x) - f(x_0) - f'(x_0)(x - x_0)|}{|x - x_0|} = 0. \end{aligned}$$

Since $f'(x_0)$ represents the slope of the line tangent to the graph of f at the point $(x_0, f(x_0))$, differentiability of f at the point x_0 means that there exists a number m such that

$$\lim_{x \to x_0} \frac{|f(x) - f(x_0) - m(x - x_0)|}{|x - x_0|} = 0.$$

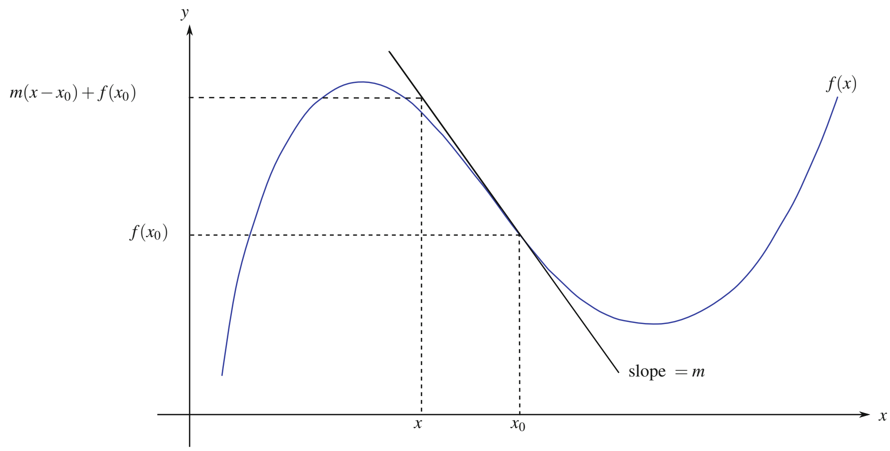


Fig. 1.12 The linear approximation to the function f at the point x_0

Now we want to consider the $m(x - x_0)$ part of the above expression. Clearly $x - x_0$ is just some number, let's call it s. Then consider the function $T : \mathbb{R} \to \mathbb{R}$, where $T(s) = ms$. Since

$$\begin{aligned} T(s+t) &= m(s+t) = ms + mt = T(s) + T(t) \quad \text{and} \\ T(cs) &= m(cs) = c(ms) = cT(s) \end{aligned}$$

then T is a linear function, or transformation. In fact, T is the linear function that most closely approximates the function f at the point $(x_0, f(x_0))$, see Fig. 1.12. Usually, for simplicity, we would simply say that T is the closest linear approximation of f at x_0. So for x values that are very close to x_0 we have

$$f(x) \approx \underbrace{m(x - x_0)}_{\substack{\text{linear approx.} \\ \text{of } f \text{ at } x_0}} + \underbrace{f(x_0)}_{\substack{\text{Shift lin. approx.} \\ \text{of } f \text{ up/down} \\ \text{appropriate amount}}} .$$

What we want to do now is generalize the concept of derivatives to functions of the form $f : \mathbb{R}^n \to \mathbb{R}^m$. We will assume the function f has the form

$$\begin{aligned} & f\big(x_1, x_2, \ldots x_n\big) \\ & = \Big(f_1(x_1, x_2, \ldots x_n), f_2(x_1, x_2, \ldots x_n), \ldots, f_m(x_1, x_2, \ldots x_n)\Big). \end{aligned}$$

What we want to find is the linear transformation, which we will denoted by Df, that most closely approximates this function $f : \mathbb{R}^n \to \mathbb{R}^m$ at some specific point $x_0 = (x_{1_0}, x_{2_0}, \ldots x_{n_0}) \in \mathbb{R}^n$. If f is differentiable at $x_0 \in \mathbb{R}^n$ then there exists a linear transformation $Df(x_0) : \mathbb{R}^n \to \mathbb{R}^m$ such that

$$\lim_{x \to x_0} \frac{\|f(x) - f(x_0) - Df(x_0)(x - x_0)\|}{\|x - x_0\|} = 0.$$

Here we have $Df(x_0)$ being a linear mapping, $(x - x_0)$ being a vector in $\mathbb{R}^n$ that originates at x_0 and terminates at x, and $Df(x_0)(x - x_0)$ being the linear mapping applied to the vector, which is an element in $\mathbb{R}^m$. If h is a vector in $\mathbb{R}^n$ then sometimes we write $Df(x_0) \cdot h$ instead of $Df(x_0)(h)$.

A quick word about what the $\|\cdot\|$ is and why it is needed. The $\|\cdot\|$ represents something called the Euclidian norm and is essentially the multi-dimensional version of the absolute value $|\cdot|$. Given a vector

$$v = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} \quad \text{then} \quad \|v\| = \sqrt{\sum_{i=1}^{n} x_i^2},$$

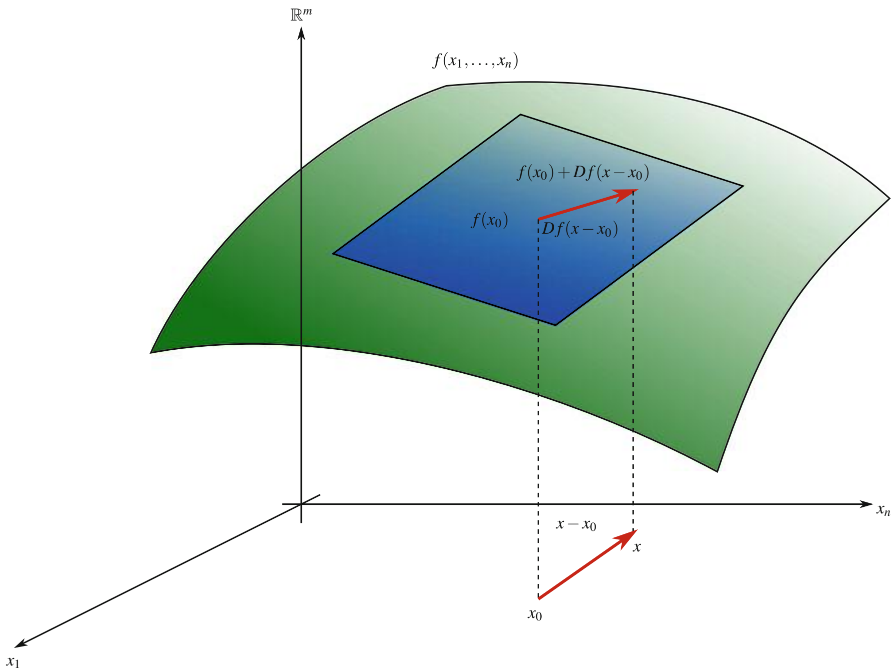


Fig. 1.13 The linear approximation to the function f at the point x_0, pictured in three-dimension

which is just the length of the vector. If you are thinking of the vector v as being the point $(x_0, \ldots, x_n)$ this gives the distance of the point from the origin. Without the norm both the numerator and the denominator in our definition would be a point or vectors in $\mathbb{R}^m$, depending on how you view them, and one can not divide a point or vector by another point or vector. The norm allows us to simply divide numbers.

It is easy to get a little confused. The notation can be a little ambiguous. But often a little ambiguity allows a great deal of flexibility, so the ambiguity is tolerated, or even embraced. Notice that Df, $Df(x)$, and $Df(x_0)$ are all used to denote the linear transformation which is an approximation of the function f. If we are being general and have not actually chosen a particular base point x_0 yet then we use Df or $Df(x)$. If we have chosen a base point x_0 then we use $Df(x_0)$. This is very analogous to saying that the derivative of $f(x) = x^2$ is f' or $f'(x)$, both of which are $2x$, but that the derivative of f at the point $x_0 = 3$ is $f'(3)$, which is 6. Also, drawing accurate pictures is no longer possible. To draw the graph of $f : \mathbb{R}^n \to \mathbb{R}^m$ would actually require drawing $\mathbb{R}^{n+m}$. The picture we draw is of course similar, but one must recognize that it is actually a cartoon as in Fig. 1.13.

As before, we have

$$f(x) \approx \underbrace{Df(x_0)(x-x_0)}_{\substack{\text{linear approx.}\\ \text{of } f \text{ at } x_0}} + \underbrace{f(x_0)}_{\substack{\text{Shift linear}\\ \text{approx. of } f\\ \text{appropriate}\\ \text{amount}}} .$$

We saw previously that we could write linear transformations as matrices, so our goal is to write $Df(x)$ as a matrix. Using the same notation as before, and assuming the basis of $\mathbb{R}^n$ is given by e_j and the basis of $\mathbb{R}^m$ is given by $\tilde{e}_i$ we want to find a_{ij} such that

$$Df(x) \cdot e_j = \sum_{i=1}^{m} a_{ij}\tilde{e}_j = \begin{bmatrix} a_{1j} \\ a_{2j} \\ \vdots \\ a_{mj} \end{bmatrix}$$

In other words, the ith component of the jth column of $Df(x)$ is just the ith component of $Df(x) \cdot e_j$.

In order find the matrix representation of $Df(x)$ recall from vector calculus that given a function $f : \mathbb{R}^n \to \mathbb{R}$ we defined the partial derivative of f with respect to x_j as

$$\frac{\partial f}{\partial x_j} = \lim_{h\to 0} \frac{f(x_1,\dots,x_j+h,\dots,x_n) - f(x_1,\dots,x_j,\dots,x_n)}{h}.$$

When we are given a function

$$\begin{aligned} f : \mathbb{R}^n &\longrightarrow \mathbb{R}^m \\ (x_1,\dots x_n) &\longmapsto \big(f_1(x_1,\dots x_n),\dots,f_m(x_1,\dots x_n)\big) \end{aligned}$$

we can of course take the partial derivatives of each f_i, $1 \le i \le m$, with respect to each x_j, $1 \le j \le n$,

$$\frac{\partial f_i}{\partial x_j} = \lim_{h\to 0} \frac{f_i(x_1,\dots,x_j+h,\dots,x_n) - f_i(x_1,\dots,x_j,\dots,x_n)}{h}.$$

There are a couple things to notice here. First, the x_j the coordinate corresponding to the basis vector e_j in $\mathbb{R}^n$ and the $f_i(x_1,\dots,x_n) \equiv f_i$ is the coordinate corresponding to the basis vector $\tilde{e}_i$ in $\mathbb{R}^m$. Also, we are holding $x_1,\dots,x_{j-1},x_{j+1},x_n$ fixed and only varying x_j. The function f_i with these variables fixed becomes a function from $\mathbb{R}$ to $\mathbb{R}$, and as such the closest linear approximation is simply a number representing a slope that we will denote a_{ij}. Thus we have

$$\frac{\partial f_i}{\partial x_j} = a_{ij}.$$

Consider our definition of $Df(x_0)$ as the linear transformation such that

$$\lim_{x\to x_0} \frac{\|f(x) - f(x_0) - Df(x_0)(x-x_0)\|}{\|x-x_0\|} = 0,$$

where $Df(x_0)(x-x_0)$ is a column vector in $\mathbb{R}^m$ whereas $x - x_0$ is a vector in $\mathbb{R}^n$.

To find a_{ij} of $Df(x_0)$ we need to find the ith element of $Df(x_0)\cdot e_j$. Letting

$$x = \begin{bmatrix} x_{1_0} \\ x_{2_0} \\ \vdots \\ x_{n_0} \end{bmatrix} + h \begin{bmatrix} 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \end{bmatrix} = x_0 + he_j$$

we can write the above as

$$\begin{aligned} &\lim_{x\to x_0} \frac{\|f(x) - f(x_0) - Df(x_0)(he_j)\|}{\|he_j\|} = 0 \\ &\Rightarrow \lim_{h\to 0} \frac{\|f(x_{1_0},\dots,x_{i_0}+h,\dots,x_{n_0}) - f(x_{1_0},\dots,x_{i_0},\dots,x_{n_0}) - hDf(x_0)\cdot e_j\|}{|h|} = 0. \end{aligned}$$

The numerator of this expression is a vector in $\mathbb{R}^m$ whose ith component is obtained by looking at the f_i in the definition of the function f. Thus the ith component of the jth column of $Df(x_0)$ is given by

$$\begin{aligned} &\lim_{h\to 0} \frac{|f_i(x_{1_0},\dots,x_{i_0}+h,\dots,x_{n_0}) - f_i(x_{1_0},\dots,x_{i_0},\dots,x_{n_0}) - ha_{ij}|}{|h|} = 0 \\ &\Rightarrow a_{ij} = \lim_{h\to 0} \frac{f_i(x_{1_0},\dots,x_{i_0}+h,\dots,x_{n_0}) - f_i(x_{1_0},\dots,x_{i_0},\dots,x_{n_0})}{h} \end{aligned}$$

but this is exactly $\frac{\partial f_i}{\partial x_j}$. Thus, the matrix representation of $Df(x)$ is given by a matrix called the **Jacobian matrix** of f,

Jacobian matrix of f	$Df(x) = \begin{bmatrix} \frac{\partial f_1}{\partial x_1} & \frac{\partial f_1}{\partial x_2} & \cdots & \frac{\partial f_1}{\partial x_n} \\ \frac{\partial f_2}{\partial x_1} & \frac{\partial f_2}{\partial x_2} & \cdots & \frac{\partial f_2}{\partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f_m}{\partial x_1} & \frac{\partial f_m}{\partial x_m} & \cdots & \frac{\partial f_m}{\partial x_n} \end{bmatrix} = \left[\frac{\partial f_i}{\partial x_j} \right]_{\substack{i = \text{row} \\ j = \text{column}}}.$

1.4 Summary, References, and Problems

1.4.1 Summary

The summary from each chapter includes the most important concepts and formulas that were covered in the chapter, along with any necessary concluding comments. The summary section is meant to serve as a quick reference as well as a chapter review.

A vector space is a set V equipped with addition $+$ and scalar multiplication $\cdot$, an additive identity 0, and an "operation of negation," which satisfies the following eight axioms. Given vector space elements $u, v, w \in V$ and scalars $c, d \in \mathbb{R}$ the eight axioms are

1. $v + w = w + v$
2. $(u + v) + w = u + (v + w)$
3. $v + 0 = v = 0 + v$
4. $v + (-v) = 0$
5. $1 \cdot v = v$
6. $c \cdot (d \cdot v) = (c \cdot d) \cdot v$
7. $c \cdot (v + w) = c \cdot v + c \cdot w$
8. $(c + d) \cdot v = c \cdot v + d \cdot v$

The most important example of a vector space is $\mathbb{R}^n$. Elements of $\mathbb{R}^n$ are written as column vectors. The basis of a vector space V is a set of elements $\{v_1, \dots, v_n\}$ that are linearly independent and span the vector space. A linear functional $T : \mathbb{R}^n \to \mathbb{R}$ is mapping that has the following two properties:

1. $T(v + w) = T(v) + T(w)$
2. $T(c \cdot v) = c \cdot T(v)$

where $v, w \in \mathbb{R}^n$ and $c \in \mathbb{R}$. The dual space of $\mathbb{R}^n$ is the set of all linear functionals on $\mathbb{R}^n$, that is, all mappings $T : \mathbb{R}^n \to \mathbb{R}$ that satisfy the two properties. The dual space of $\mathbb{R}^n$ is denoted by $(\mathbb{R}^n)^*$. Any basis of a vector space V induces a dual basis of V^*.

Using three properties for how we would expect the volume of a parallelepiped to behave we discover that volumes naturally come with a sign attached. Furthermore, using these properties we are able to derive the formula for the determinant of a matrix, which finds the signed volume of the parallelepiped spanned by the columns of the matrix.

Formula for the determinant of a matrix	$\begin{vmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{vmatrix} = \sum_{\sigma \in S_n} sgn(\sigma) \prod_{i=1}^{n} a_{\sigma(i)i}$

Finally, given a function $f : \mathbb{R}^n \to \mathbb{R}^m$ defined by $f = (f_1, \dots f_m)$ the derivative Df of this function is a linear transformation given by the Jacobian matrix

$$\boxed{\text{Jacobian matrix of } f \qquad Df(x) = \begin{bmatrix} \frac{\partial f_1}{\partial x_1} & \frac{\partial f_1}{\partial x_2} & \cdots & \frac{\partial f_1}{\partial x_n} \\ \frac{\partial f_2}{\partial x_1} & \frac{\partial f_2}{\partial x_2} & \cdots & \frac{\partial f_2}{\partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f_m}{\partial x_1} & \frac{\partial f_m}{\partial x_m} & \cdots & \frac{\partial f_m}{\partial x_n} \end{bmatrix} = \left[\frac{\partial f_i}{\partial x_j} \right]_{\substack{i=\text{ row} \\ j=\text{ column}}} .}$$

1.4.2 References and Further Reading

The material presented in this chapter is all fairly standard, though the exact approach taken in presenting the material is in line with what is required in the rest of the book. There are innumerable books on linear algebra, but the one by Friedberg, Insel, and Spense [22] is a very good reference. The relation between determinant and volume of parallelepipeds is also standard material, but this introduction largely followed a very nice anonymous and undated exposition on determinants [2]. Also see the expositions in Abraham, Marsden, and Ratiu [1]. For multivariable calculus the textbook by Stewart [43] is a standard one. For a somewhat more detailed mathematical introduction to multivariable calculus the book by Marsden and Hoffmen [31] is extremely good. In particular, Marsden tends to use some excellent notation which has been used throughout much of this book.

1.4.3 Problems

Question 1.18 Let $V = \{(x_1, x_2) | x_1, x_2 \in \mathbb{R}\}$. That is, V is the set of ordered pairs of real numbers. If $(a_1, a_2), (b_1, b_2) \in V$ and $c \in \mathbb{R}$ is a scalar define

$$(a_1, a_2) + (b_1, b_2) = (a_1 + b_1, a_2 b_2) \qquad \text{and} \qquad c(a_1, a_2) = (ca_1, a_2).$$

Is V a vector space under these operations?

Question 1.19 Let $V = \{(x_1, x_2) | x_1, x_2 \in \mathbb{R}\}$. If $(a_1, a_2), (b_1, b_2) \in V$ and $c \in \mathbb{R}$ is a scalar define

$$(a_1, a_2) + (b_1, b_2) = (a_1 + b_1, a_2 + b_2) \qquad \text{and} \qquad c(a_1, a_2) = (a_1, 0).$$

Is V a vector space under these operations?

Question 1.20 Let $V = \{(x_1, x_2) | x_1, x_2 \in \mathbb{R}\}$. If $(a_1, a_2), (b_1, b_2) \in V$ and $c \in \mathbb{R}$ is a scalar define

$$(a_1, a_2) + (b_1, b_2) = (a_1 + 3b_1, a_2 + 5b_2) \qquad \text{and} \qquad c(a_1, a_2) = (ca_1, ca_2).$$

Is V a vector space under these operations?

Question 1.21 Let $V = \{(x_1, x_2) | x_1, x_2 \in \mathbb{R}\}$. If $(a_1, a_2), (b_1, b_2) \in V$ and $c \in \mathbb{R}$ is a scalar define

$$(a_1, a_2) + (b_1, b_2) = (a_1 + b_1, a_2 + b_2) \qquad \text{and} \qquad c(a_1, a_2) = \begin{cases} (0, 0) & \text{if } c = 0 \\ \left(ca_1, \frac{a_2}{c}\right) & \text{if } c \neq 0. \end{cases}$$

Is V a vector space under these operations?

Question 1.22 Let V and W be vector spaces and let $Z = \{(v, w) | v \in V, w \in W\}$. If $(v_1, w_1), (v_2, w_2) \in Z$ and $c \in \mathbb{R}$ is a scalar define

$$(v_1, w_1) + (v_2, w_2) = (v_1 + v_2, w_1 + w_2) \qquad \text{and} \qquad c(v_1, w_2) = (cv_1, cw_1).$$

Show that Z is a vector space.

Question 1.23 Consider the vectors $v_1 = \begin{bmatrix} 1 \\ 3 \\ -4 \\ 2 \end{bmatrix}$, $v_2 = \begin{bmatrix} 1 \\ 1 \\ -2 \\ 0 \end{bmatrix}$, $v_3 = \begin{bmatrix} 1 \\ -3 \\ 2 \\ -4 \end{bmatrix}$, $v_4 = \begin{bmatrix} -2 \\ 0 \\ 2 \\ 0 \end{bmatrix}$. Show that these vectors are linearly dependent. That is, find scalars c_1, c_2, c_3, c_4, not all equal to zero, such that $c_1v_1 + c_2v_2 + c_3v_3 + c_4v_4 = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$.

Question 1.24 Consider the vectors $v_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \\ -1 \end{bmatrix}$, $v_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \\ -1 \end{bmatrix}$, $v_3 = \begin{bmatrix} 0 \\ 0 \\ 1 \\ -1 \end{bmatrix}$, $v_4 = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix}$. Show that these vectors are linearly independent. To do this show that the only linear combination that equals zero, $c_1v_1 + c_2v_2 + c_3v_3 + c_4v_4 = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$, is the one where all scalars c_1, c_2, c_3, c_4 are equal to zero.

Question 1.25 Give a set of three linearly dependant vectors in $\mathbb{R}^3$ such that none of the three vectors is a multiple of any one the others.

Question 1.26 Give a set of four linearly dependant vectors in $\mathbb{R}^4$ such that none of the four vectors is a multiple of any one of the others.

Question 1.27 Do the vectors $v_1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$, $v_2 = \begin{bmatrix} 1 \\ -2 \end{bmatrix}$, span $\mathbb{R}^2$? That is, for an arbitrary vector $v \in \mathbb{R}^2$ can one can find scalars c_1, c_2 such that $c_1v_1 + c_2v_2 = v$?

Question 1.28 Do the vectors $v_1 = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}$, $v_2 = \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}$, $v_3 = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}$, span $\mathbb{R}^3$? That is, for an arbitrary vector $v \in \mathbb{R}^3$ can one can find scalars c_1, c_2, c_3 such that $c_1v_1 + c_2v_2 + c_3v_3 = v$?

Question 1.29 Do the vectors $v_1 = \begin{bmatrix} 4 \\ 2 \\ 0 \end{bmatrix}$, $v_2 = \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}$, $v_3 = \begin{bmatrix} 2 \\ 2 \\ 1 \end{bmatrix}$, span $\mathbb{R}^3$? That is, for an arbitrary vector $v \in \mathbb{R}^3$ can one can find scalars c_1, c_2, c_3 such that $c_1v_1 + c_2v_2 + c_3v_3 = v$?

Question 1.30 If $T : V \to W$ is a linear transformation show that $T(\mathbf{0}_V) = \mathbf{0}_W$, where $\mathbf{0}_V$ is the zero vector in V and $\mathbf{0}_W$ is the zero vector in W.

Question 1.31 If $T : V \to W$ is a linear transformation show that $T(cx + y) = cT(x) + T(y)$.

Question 1.32 If $T : V \to W$ is a linear transformation show that $T\left(\sum_{i=1}^{n} a_i v_i\right) = \sum_{i=1}^{n} c_i T(v_i)$.

Question 1.33 Define $T : \mathbb{R}^2 \to \mathbb{R}^2$ by $T\left(\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}\right) = \begin{bmatrix} 2x_1 + x_2 \\ x_1 \end{bmatrix}$. Is T a linear transformation? If so find the matrix representation of T.

Question 1.34 Given any angle θ define $T_\theta : \mathbb{R}^2 \to \mathbb{R}^2$ by $T\left(\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}\right) = \begin{bmatrix} x_1 \cos\theta - x_2 \sin\theta \\ x_1 \sin\theta + x_2 \cos\theta \end{bmatrix}$. Show that T_θ is a linear transformation. The transformation T_θ is called rotation by θ. Find the matrix representation of T_θ.

Question 1.35 Define $T : \mathbb{R}^2 \to \mathbb{R}^2$ by $T\left(\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}\right) = \begin{bmatrix} x_1 \\ -x_2 \end{bmatrix}$. Show that T is a linear transformation. The transformation T is called reflection across the x-axis. Find the matrix representation of T.

Question 1.36 Let $\left\{\begin{bmatrix}2\\1\end{bmatrix},\begin{bmatrix}3\\1\end{bmatrix}\right\}$ be a basis for $\mathbb{R}^2$. Find the dual basis. That is, find the basis of $\left(\mathbb{R}^2\right)^*$ which is dual to this basis.

Question 1.37 Using the formula for the determinant, find the constant C such that

$$\begin{vmatrix} 2a_1 & 2a_2 & 2a_3 \\ 3b_1+4c_1 & 3b_2+4c_1 & 3b_3+4c_1 \\ 5c_1 & 5c_2 & 5c_3 \end{vmatrix} = C \begin{vmatrix} a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \\ c_1 & c_2 & c_3 \end{vmatrix}.$$

Question 1.38 Let $\left\{\begin{bmatrix}1\\0\\1\end{bmatrix},\begin{bmatrix}1\\2\\1\end{bmatrix},\begin{bmatrix}0\\0\\1\end{bmatrix}\right\}$ be a basis for $\mathbb{R}^3$. Find the dual basis.

Question 1.39 Using the formula for the determinant, find the constant C such that

$$\begin{vmatrix} 3a_1 & 3a_2 & 3a_3 \\ 3b_1 & 3b_2 & 3b_3 \\ 3c_1 & 3c_2 & 3c_3 \end{vmatrix} = C \begin{vmatrix} a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \\ c_1 & c_2 & c_3 \end{vmatrix}.$$

Question 1.40 Find the value of the determinant of the following matrices:

$$\begin{bmatrix} 2 & -6 \\ -2 & 5 \end{bmatrix}, \quad \begin{bmatrix} 3 & 0 & 5 \\ -2 & -1 & 4 \\ 7 & 6 & 0 \end{bmatrix}, \quad \begin{bmatrix} 0 & 1 & 3 \\ -1 & 0 & -3 \\ 2 & 3 & 0 \end{bmatrix}.$$

Question 1.41 Find the Jacobian matrix of the following functions:

a) $f(x, y) = xy$
b) $f(x, y) = \sin(x^3 + y^2)$
c) $f(x, y, z) = x^5 + y^4$
d) $f(x, y) = \Big(\sin\big(x\sin(y)\big), (x + y)^2\Big)$
e) $f(x, y) = \big(\cos(xy), \sin(xy), x^2y^2\big)$
f) $f(x, y, z) = \big(z^{xy}, x^2, \tan(xyz)\big)$

Chapter 2
An Introduction to Differential Forms

In this chapter we introduce one of the fundamental ideas of this book, the differential one-form. Later chapters will discuss differential k-forms for $k > 1$. We slowly and systematically build up to the concept of a one-forms by first considering a variety of necessary ingredients. We begin in section one by introducing the Cartesian coordinate functions, which play a surprisingly important role in understanding exactly what a one-form is, as well as playing a role in the notation used for one-forms.

In section two we discuss manifolds, tangent spaces, and vector fields in enough detail to give you a good intuitive idea of what they are, leaving a more mathematically rigorous treatment for Chap. 10. In section three we return to the concept of directional derivatives that you are familiar with from vector calculus. The way directional derivatives are usually defined in vector calculus needs to be tweaked just a little for our purposes. Once this is done a rather surprising identification between the standard Euclidian unit vectors and differential operators can be made. This equivalence plays a central role going forward.

Finally in section four differential one-forms are defined. Thoroughly understanding differential one-forms requires pulling together a variety of concepts; vector spaces and dual spaces, the equivalence between Euclidian vectors and differential operators, directional derivatives and the Cartesian coordinate functions. The convergence of all of these mathematical ideas in the definition of the one-form is the goal of the chapter.

2.1 Coordinate Functions

Now that we have reviewed the necessary background we are ready to start getting into the real meat of the course, differential forms. For the first part of this course, until we understand the basics, we will deal with differential forms on $\mathbb{R}^n$, with $\mathbb{R}^2$ and $\mathbb{R}^3$ being our primary examples. Only after we have developed some intuitive feeling for differential forms will we address differential forms on more general manifolds. You can very generally think of a manifold as a space which is locally Euclidian - that means that if you look closely enough at one small part of a manifold then it basically looks like $\mathbb{R}^n$ for some n.

For the time being we will just work with our familiar old Cartesian coordinate system. We will discuss other coordinate systems later on. We are used to seeing $\mathbb{R}^2$ pictured as a plane with an x and a y axis as in Fig. 2.1. Here we have plotted the points $(2, 1)$, $(-4, 3)$, $(-2, -2)$, and $(1, -4)$ on the xy-plane. Similarly, we are used to seeing $\mathbb{R}^3$ pictured with an x, y, and z axis as show in Fig. 2.2. Here we have plotted the points $(2, 3, 4)$, $(-3, -4, 2)$, and $(4, 1, -3)$. Also, notice that we follow the "right-hand rule" convention in this book when drawing $\mathbb{R}^3$.

Notice the difference in the notations between points in $\mathbb{R}^n$ and vectors in $\mathbb{R}^n$. For example, points in $\mathbb{R}^2$ or $\mathbb{R}^3$ are denoted by $p = (x, y)$ or $p = (x, y, z)$, where $x, y, z \in \mathbb{R}$, whereas vectors in $\mathbb{R}^2$ or $\mathbb{R}^3$ are denoted by

$$v = \begin{bmatrix} x \\ y \end{bmatrix} \text{ or } v = \begin{bmatrix} x \\ y \\ z \end{bmatrix}$$

J. P. Fortney, *A Visual Introduction to Differential Forms and Calculus on Manifolds*,
https://doi.org/10.1007/978-3-319-96992-3_2

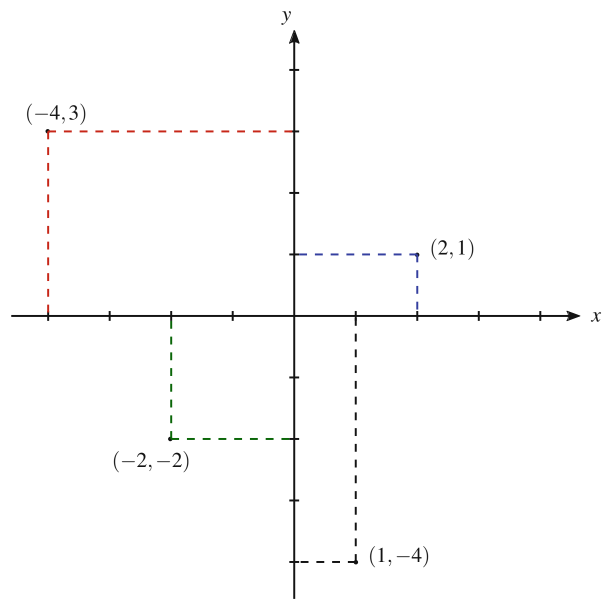


Fig. 2.1 The Cartesian coordinate system for $\mathbb{R}^2$ with the points (2, 1), (−4, 3), (−2, −2), and (1, −4) shown

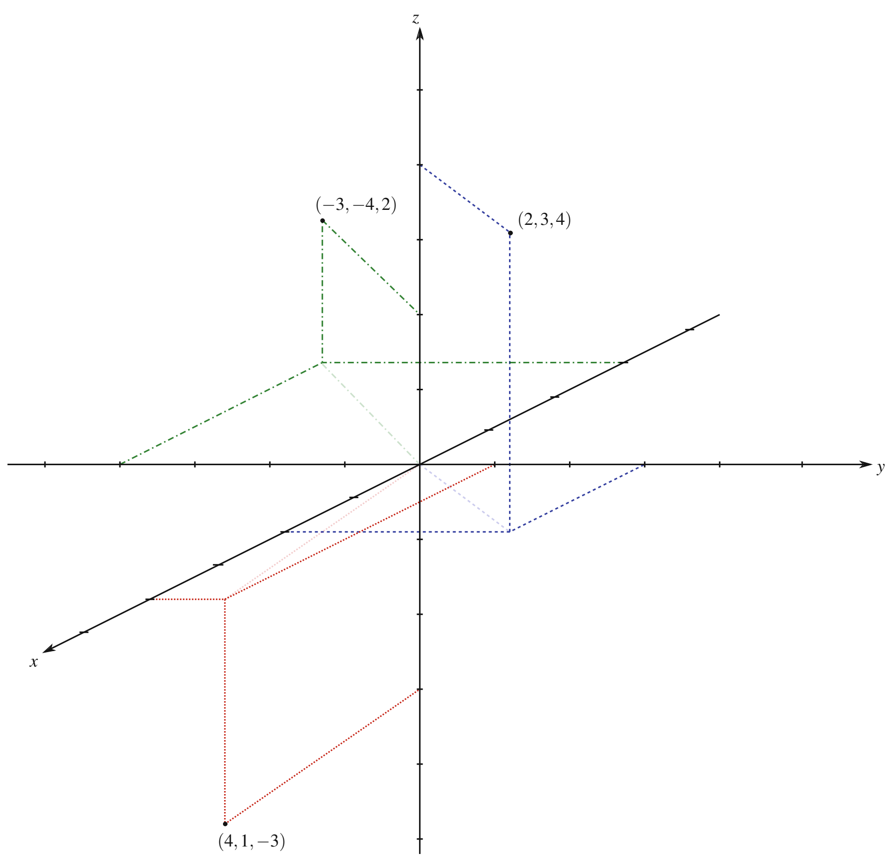


Fig. 2.2 The Cartesian coordinate system for $\mathbb{R}^3$ with the points (2, 3, 4), (−3, −4, 2), and (4, 1, −3) shown

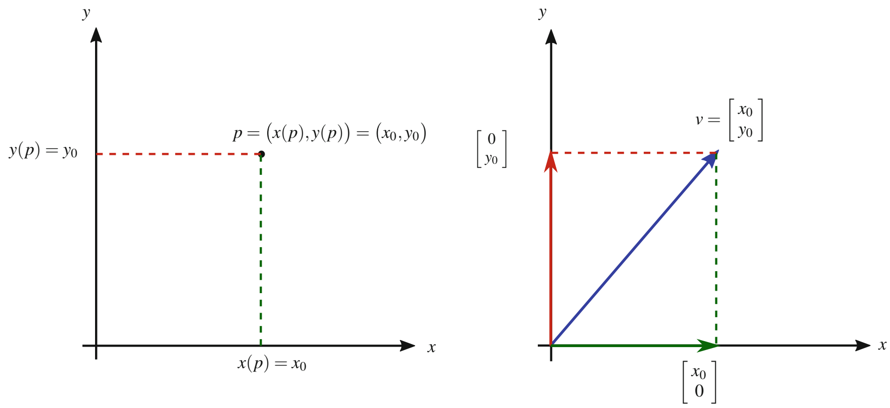


Fig. 2.3 The manifold $\mathbb{R}^2$ (left) and the vector space $\mathbb{R}^2$ (right)

where $x, y, z \in \mathbb{R}$. Points and vectors in $\mathbb{R}^n$ for arbitrary n are denoted similarly. Here comes a subtle distinction that is almost always glossed over in calculus classes, especially in multivariable or vector calculus, yet it is a distinction that is very important.

- We will call the collection of points $p = (x_1, x_2, \dots, x_n)$, with $x_1, x_2, \dots, x_n \in \mathbb{R}$, the *manifold* $\mathbb{R}^n$.
- We will call the collection of vectors $v = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$, where $x_1, x_2, \dots, x_n \in \mathbb{R}$, the *vector space* $\mathbb{R}^n$.

One of the reasons this distinction is glossed over in calculus classes is that most calculus students have not yet taken a course in linear algebra and do not yet know what a vector space is. Also, it is certainly possible to "do" vector calculus without knowing the technical definition of a vector space by being a little vague and imprecise mathematically speaking. The manifold $\mathbb{R}^2$ and the vector space $\mathbb{R}^2$ look exactly alike yet are totally different spaces. They have been pictured in Fig. 2.3. The manifold $\mathbb{R}^2$ contains points and the vector space $\mathbb{R}^2$ contains vectors. Of course, in the case of $\mathbb{R}^2$, $\mathbb{R}^3$, or $\mathbb{R}^n$ for that matter, these spaces can be naturally identified. That is, they are really the same. So why is making the distinction so important? It is important to make the distinction for a couple of reasons. First, keeping clear track of our spaces will help us understand the theory of differential forms better, and second, when we eventually get to more general manifolds, most manifolds will not have a vector space structure so we do not want to get into the bad habit of thinking of our manifolds as vector spaces too.

Now let us spend a moment reviewing good old-fashioned functions on $\mathbb{R}^2$ and $\mathbb{R}^3$. We will consider real-valued functions, which simply means that the range (or codomain) is the reals $\mathbb{R}$ or some subset of the reals. Real-valued functions from $\mathbb{R}^2$ are the easiest to visualize since they can be graphed in three dimensional space. Consider the real-valued functions $f : \mathbb{R}^2 \to \mathbb{R}$ defined by the following $f(x, y) = x^3 + y^2 - 3xy$. For example, $f(2, 3)$ is given by $(2)^3 + (3)^2 - 3(2)(3) = -1$, $f(5, -2)$ is given by $(5)^3 + (-2)^2 - 3(5)(-2) = 159$ and $f(-3, -1)$ is given by $(-3)^3 + (-1)^2 - 3(-3)(-1) = -35$. A function from $\mathbb{R}^2$ to $\mathbb{R}$ is graphed in $\mathbb{R}^3$ by the set of points $(x, y, f(x, y))$. The function $f(x, y) = x^3 + y^2 - 3xy$ is shown in Fig. 2.4. Sometimes we will write $f(p)$ where $p = (x, y)$. By doing this we are emphasizing that the input (x, y) is a point in the domain.

As a further illustration, the graphs of the functions $f(x, y) = xy$ and $f(x, y) = x^2 + y^2$ are shown in Fig. 2.5. As before, the set of points that are actually graphed are the points $(x, y, f(x, y))$ in $\mathbb{R}^3$, thus we can see that it requires three dimensions to draw the graph of a function $f : \mathbb{R}^2 \to \mathbb{R}$. That means that we can not accurately draw the graph of any function $f : \mathbb{R}^n \to \mathbb{R}$ for $n \geq 3$. However, we often continue to draw inaccurate cartoons similar to Figs. 2.4 and 2.5 when we want to try to visualize or depict functions $f : \mathbb{R}^n \to \mathbb{R}$ where $n \geq 3$.

We will use different name for "functions" on the manifold $\mathbb{R}^n$ and "functions" on the vector space $\mathbb{R}^n$. Functions on the manifold $\mathbb{R}^n$ will simply be called **functions**, or sometimes, if we want to emphasize that range is the set of real numbers, **real-valued functions**. However, our "functions" on the vector space $\mathbb{R}^n$ are called **functionals** when the range is the set of real values and **transformations** when the range is another vector space. We have already spent some time considering a

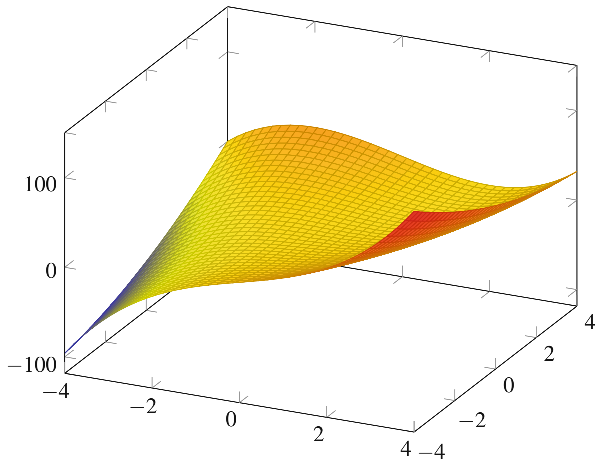


Fig. 2.4 The graph of function $f(x, y) = x^3 + y^2 - 3xy$ is displayed

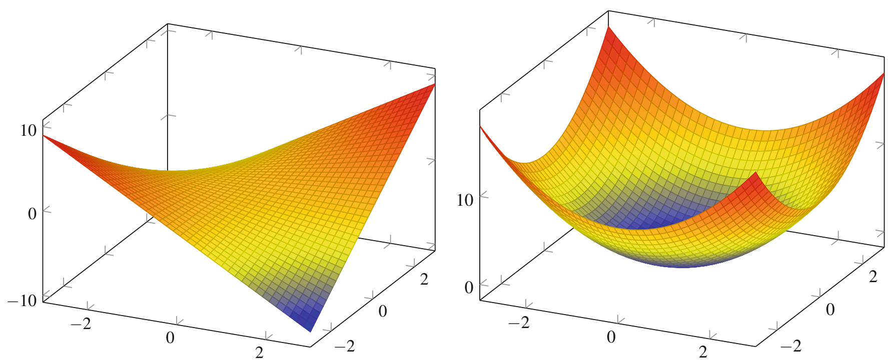


Fig. 2.5 The graphs of the function $f(x, y) = xy$ (left) and $f(x, y) = x^2 + y^2$ (right) as an illustration of the graphs of functions in two variables depicted in $\mathbb{R}^3$

certain class of functionals, the linear functionals that satisfy

$$f(v + w) = f(v) + f(w)$$
$$f(cv) = cf(v)$$

where v and w are vectors and c is a real number. So, the difference between manifolds $\mathbb{R}^n$ and vector spaces $\mathbb{R}^n$ is implicit in our language:

$$\text{function } f : (\text{manifold})\ \mathbb{R}^n \longrightarrow \mathbb{R}$$
$$\text{functional } f : (\text{vector space})\ \mathbb{R}^n \longrightarrow \mathbb{R}.$$

Now, we will consider a very special function from $\mathbb{R}^n$ to $\mathbb{R}$. We will start by using $\mathbb{R}^2$ as an example. The special functions we want to look at are called the **coordinate functions**. For the moment we will look at the **Cartesian coordinate functions**.

Consider a point p in the manifold $\mathbb{R}^2$ given by the Cartesian coordinates $(3, 5)$. Then the coordinate function

$$x : \mathbb{R}^2 \longrightarrow \mathbb{R}$$
$$p \longmapsto x(p)$$

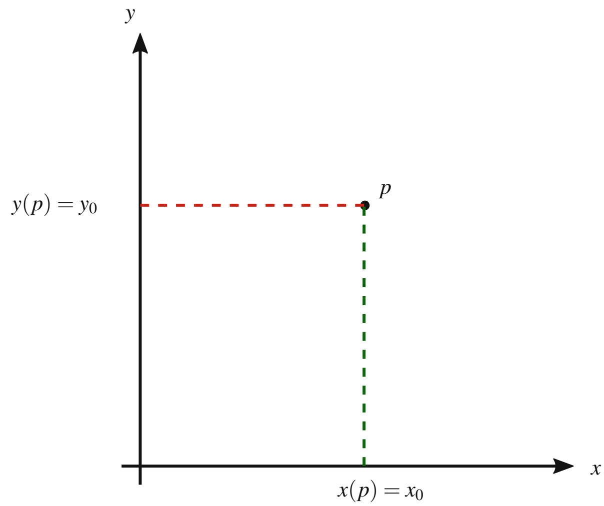


Fig. 2.6 The coordinate functions $x : \mathbb{R}^2 \to \mathbb{R}$ and $y : \mathbb{R}^2 \to \mathbb{R}$ picking off the coordinate values x_0 and y_0, respectively, of the point $p \in \mathbb{R}^2$

picks off the first coordinate, that is, $x(p) = x\big((3, 5)\big) = 3$, and the coordinate function

$$\begin{aligned} y : \mathbb{R}^2 &\longrightarrow \mathbb{R} \\ p &\longmapsto y(p) \end{aligned}$$

picks off the second coordinate, $y(p) = y\big((3, 5)\big) = 5$. This is illustrated in Fig. 2.6.

The fact of the matter is, even though you are used to thinking of (3, 5) as a point, a point is actually more abstract than that. A point exists independent of its coordinates. For example, you will sometimes see something like this,

$$p = \big(x(p), y(p)\big)$$

where, as an example, for the point we were given above we have $x(p) = 3$ and $y(p) = 5$.

There is some ambiguity that we need to get used to here. Often x and y are used to represent the outputs of the coordinate functions x and y. So, we could end up with an ambiguous statement like this

$$\begin{aligned} x(p) &= x \\ y(p) &= y \end{aligned}$$

where the x and the y on the left hand sides of the equations are coordinate functions from $\mathbb{R}^2$ to $\mathbb{R}$, and the x and y on the right hand side are real numbers, that is, elements of $\mathbb{R}$. Learning to recognize when x and y are coordinate functions and when they are real values requires a little practice, you have to look carefully at how they are being used. This distinction may sometimes be pointed out, but it often is glossed over.

Even though points are abstract and independent of coordinates, we can not actually talk about a particular point without giving it an "address." This is actually what coordinate functions do, they are used to give a point an "address."

Coordinate functions on $\mathbb{R}^3$ (and $\mathbb{R}^n$ for that matter) are completely analogous. For example, on $\mathbb{R}^3$ we have the three coordinate functions

$$\begin{aligned} x : \mathbb{R}^3 &\longrightarrow \mathbb{R} \\ p &\longmapsto x(p) \end{aligned} \qquad \begin{aligned} y : \mathbb{R}^3 &\longrightarrow \mathbb{R} \\ p &\longmapsto y(p) \end{aligned} \qquad \begin{aligned} z : \mathbb{R}^3 &\longrightarrow \mathbb{R} \\ p &\longmapsto z(p) \end{aligned}$$

which pick off the x, the y and the z coordinates of the point p as illustrated in Fig. 2.7.

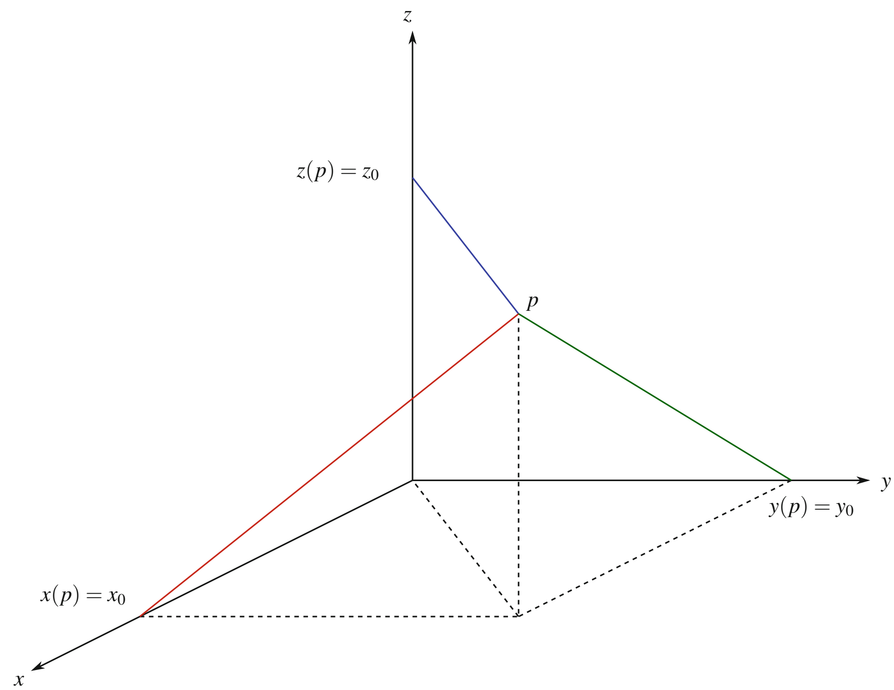


Fig. 2.7 The coordinate functions $x : \mathbb{R}^3 \to \mathbb{R}$, $y : \mathbb{R}^3 \to \mathbb{R}$, and $z : \mathbb{R}^3 \to \mathbb{R}$ picking off the coordinate values x_0, y_0, and z_0 respectively, of the point $p \in \mathbb{R}^3$

Of course Cartesian coordinate functions are not the only coordinate functions there are. Polar coordinate, spherical coordinate, and cylindrical coordinate functions are some other examples you have likely encountered before. We will talk about these coordinate systems in more detail later on but for the moment we consider one example. Consider the point $p = p\big(x(p), y(p)\big) = (3, 5)$. Polar coordinates r and θ are often given by $x = r\cos(\theta)$ and $y = r\sin(\theta)$. Suppose we want to find the "address" of point p in terms of polar coordinates. That is, we want to find $p = \big(r(p), \theta(p)\big)$.

Since we know p in terms of x and y coordinates, we can write $r = \sqrt{x^2 + y^2}$ and $\theta = \arctan\left(\frac{y}{x}\right)$. Then we have

$$
\begin{aligned}
p = \big(r(p), \theta(p)\big) &= \Big(r\big(x(p), y(p)\big), \theta\big(x(p), y(p)\big)\Big) \\
&= \left(\sqrt{x(p)^2 + y(p)^2}, \arctan\left(\frac{y(p)}{x(p)}\right)\right) \\
&= \left(\sqrt{3^2 + 5^2}, \arctan\left(\frac{5}{3}\right)\right) \\
&= \left(\underbrace{\sqrt{34}}_{r}, \underbrace{\arctan\left(\frac{5}{3}\right)}_{\theta}\right) \\
&\approx \big(5.831, 59.036\big).
\end{aligned}
$$

Again, the same ambiguity applies, r and θ can either be coordinate functions or the values given by the coordinate functions.

2.2 Tangent Spaces and Vector Fields

Let us recap what we have done so far. We have reviewed vector spaces and talked about $\mathbb{R}^n$ as a vector space, that is, as a collection of vectors of the from

$$v = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$$

where $x_1, \ldots, x_n \in \mathbb{R}$. Then we talked about $\mathbb{R}^n$ as a manifold, that is, as a collection of points $(x_1, \ldots, x_n)$ where $x_1, \ldots, x_n \in \mathbb{R}$. We have also reviewed linear functionals on the vector space $\mathbb{R}^n$ and functions on the manifold $\mathbb{R}^n$ and introduced coordinate functions. We will now see how the manifold $\mathbb{R}^n$ is related to vector spaces $\mathbb{R}^n$. Our motivation for this is our desire to do calculus. Well, for the moment to take directional derivatives anyway.

However, to help us picture what is going on a little better we are going to start by considering three manifolds, S^1, S^2, and T. The manifold S^1 is simply a circle, the manifold S^2 is the sphere, and the manifold T is the torus, which looks like a donut. For the moment we are going to be quite imprecise. We want you to walk away with a general feeling for what a manifold is and not overwhelm you with technical details. For the moment we will simply say a **manifold** is a space that is **locally Euclidian**. What we mean by that is if we look at a very small portion of the manifold that portion looks like $\mathbb{R}^n$ for some natural number n.

We will illustrate this idea for the three manifolds we have chosen to look at. The circle S^1 is locally like $\mathbb{R}^1$, which explains the superscript 1 on the S. In Fig. 2.8 we zoom in on a small portion of the one-sphere S^1 a couple of times to see that locally the manifold S^1 looks like $\mathbb{R}^1$. Similarly, in Fig. 2.9 we zoom in on a small portion of the two-sphere S^2 to see that it looks like $\mathbb{R}^2$ locally and in Fig. 2.10 we zoom in on a small portion of the torus T see that it also looks locally like $\mathbb{R}^2$.

The point with these examples is that even though these example manifolds *locally* look like $\mathbb{R}^1$ and $\mathbb{R}^2$, *globally* they do not. That is, their global behavior is more complex; these spaces somehow twist around and reconnect with themselves in a way that Euclidian space does not.

Now we will introduce the idea of a **tangent space**. Again, we want to give you an general feeling for what a tangent space is and not overwhelm you with the technical details, which will be presented in 10.2. From calculus you should recall what the tangent line to a curve at some point of the curve is. Each point of the curve has its own tangent line, as is shown in Fig. 2.11. We can see that the one dimensional curve has tangent spaces that are lines, that is, that are also one dimensional.

Similarly, a two dimensional surface has a tangent plane at each point of the surface, as is pictured in Fig. 2.12. The tangent space to a manifold at the point p is basically the set of all lines that are tangent to smooth curves that go through

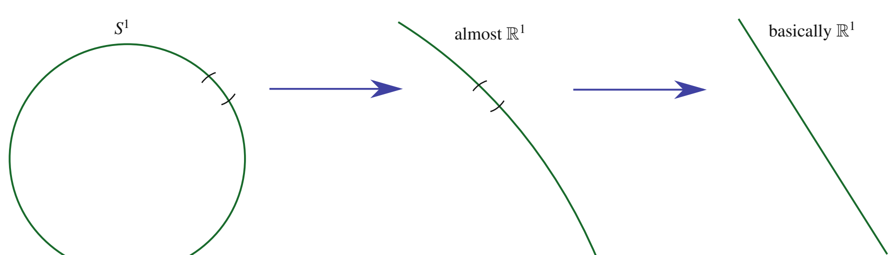


Fig. 2.8 Here a small portion of a one-sphere, or circle, S^1 is zoomed in on to show that locally the manifold S^1 looks like $\mathbb{R}^1$

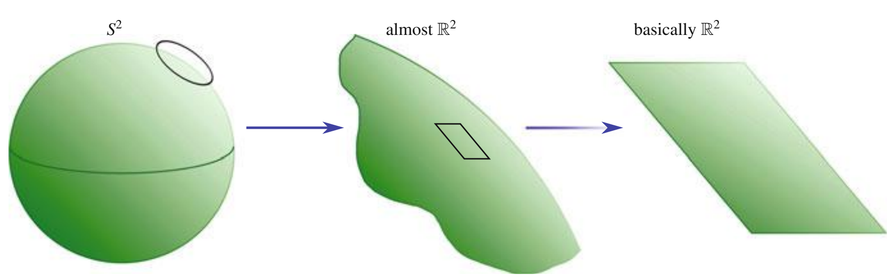


Fig. 2.9 Here a small portion of the two-sphere S^2 is zoomed in on to show that locally the manifold S^2 looks like $\mathbb{R}^2$

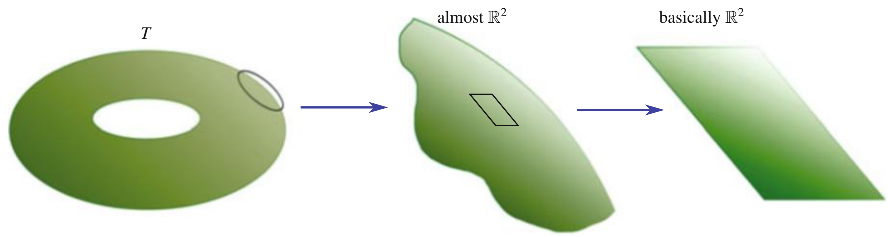


Fig. 2.10 Here a small portion of the torus T is zoomed in on to show that locally the manifold T looks like $\mathbb{R}^2$

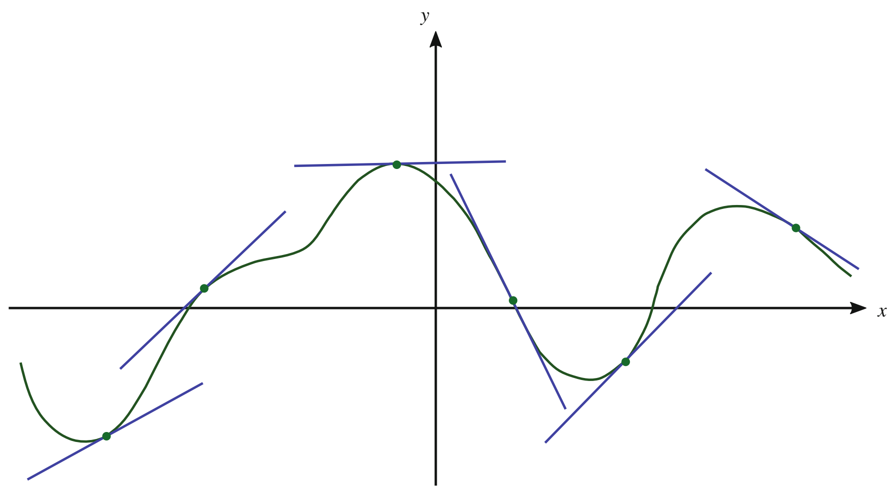


Fig. 2.11 Here a curve is pictured with tangent lines drown at several points

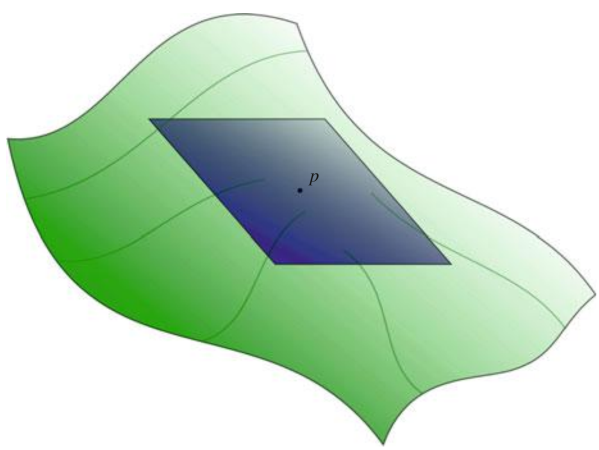


Fig. 2.12 Here a two-dimensional surface is pictured along with a tangent plane drown at one point

the manifold at the point p. This is illustrated in Fig. 2.13 where we have drawn a point p with five smooth curves going through it along with the five tangent lines to the curves at the point p. In both of these pictures the tangent space is a plane.

If you imagine a point in a three dimensional manifold and all the possible curves that can go through that point, and then imagine all the tangent lines to those curves at that point, you will see that the tangent space at that point is three dimensional. It turns out that the tangent space at a "nice" point is the same dimension as the manifold at that point. In this book we will not discuss manifolds with points that are not "nice" in this sense. In fact, the tangent spaces turn out to be nothing more than the vector spaces $\mathbb{R}^n$ for the correct value of n. Because the tangent spaces are vector spaces we think of the tangent space at a point p to be the set of all tangent vectors to the point p. We use special notation for the tangent space. Either T_pM or $T_p(M)$ denotes the tangent space to the manifold M at the point p.

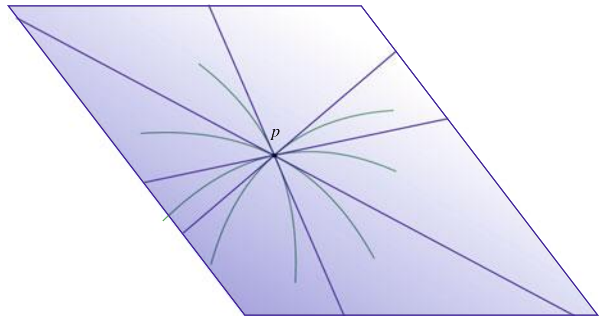


Fig. 2.13 The tangent plane to a two-dimensional surface at the point p is the set of all tangent lines at p to curves in that surface that go through p

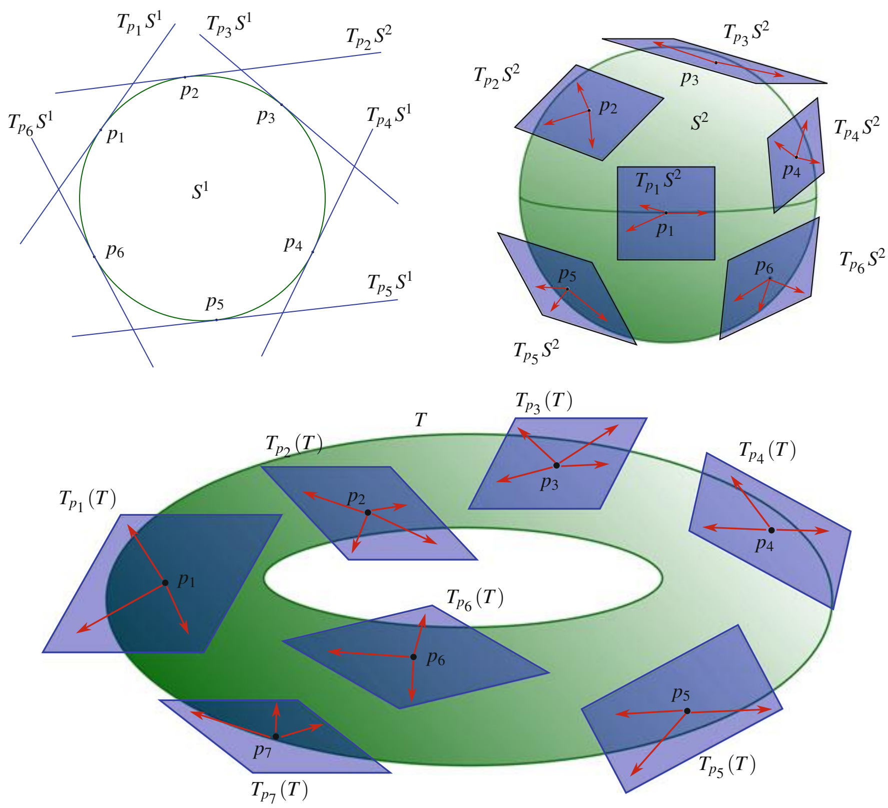


Fig. 2.14 The manifold S^1 along with a number of tangent lines depicted (top left). The manifold S^2 along with a number of tangent planes depicted (top right). The manifold T along with a number of tangent planes depicted (bottom). For both manifolds S^2 and T several vectors in each tangent plane are also depicted

To get a better idea of what the set of tangent spaces to a manifold look like, we have drawn a few tangent spaces to the three simple manifolds we had looked at earlier in Fig. 2.14, the circle S^1, the sphere S^2, and the torus T. For the two dimensional tangent spaces we have also drawn some elements of the tangent spaces, that is, vectors emanating from the point p. The tangent spaces of the circle are denoted by T_pS^1, the tangent spaces of the sphere are denoted by T_pS^2, and the tangent spaces of the torus are denoted by $T_p(T)$. We use the parenthesis here simply because T_pT looks a little odd.

Even though we are quite familiar with the manifolds $\mathbb{R}^2$ and $\mathbb{R}^3$, imagining the tangent spaces to these manifolds is a little strange, though is it something you have been implicitly doing since vector calculus. The tangent space of the manifold $\mathbb{R}^2$ at a point p is the set of all vectors based at the point p. In vector calculus we dealt with vectors with different base points a lot, but we always simply thought of these vectors as being *in* the manifold $\mathbb{R}^2$ or $\mathbb{R}^3$. Now we think of these vectors as belonging to a separate copy of the vector space $\mathbb{R}^2$ or $\mathbb{R}^3$ attached at the point and called the tangent space.

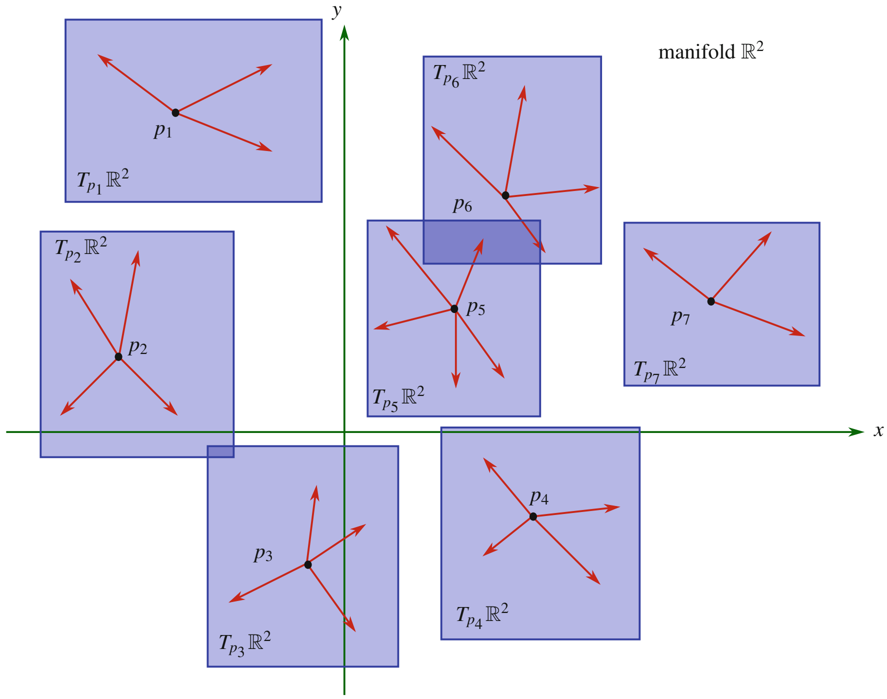


Fig. 2.15 The manifold $\mathbb{R}^2$ along with the tangent spaces $T_{p_i}\mathbb{R}^2$ at seven different points. A few vectors in each tangent space are shown

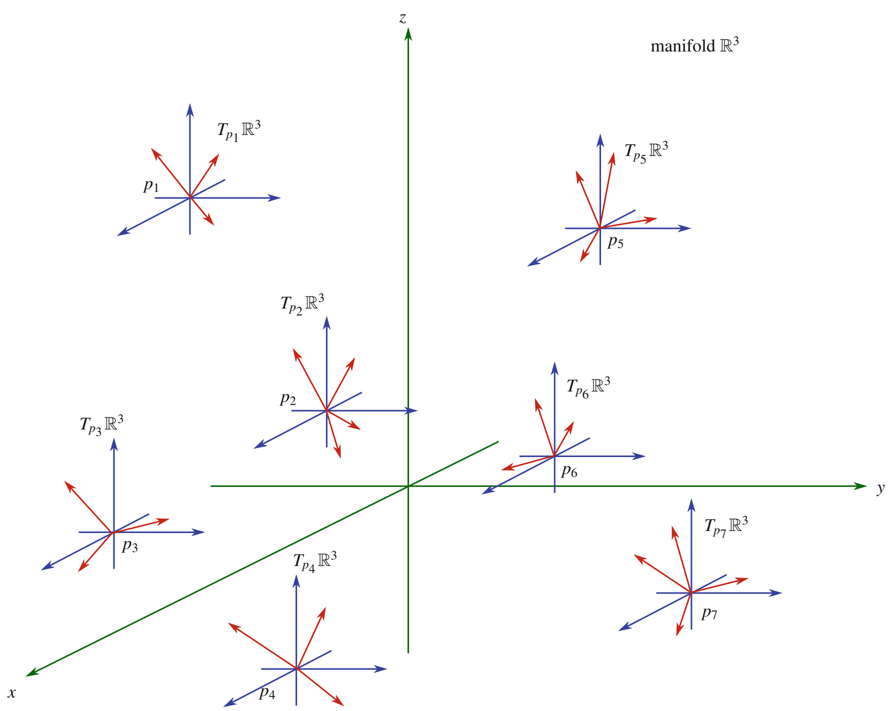


Fig. 2.16 The manifold $\mathbb{R}^3$ along with the tangent spaces $T_{p_i}\mathbb{R}^3$ at seven different points. A few vectors in each tangent space are shown

Figure 2.15 tries to help you imagine what is going on. The manifold $\mathbb{R}^2$ is pictured along with the tangent spaces $T_{p_i}\mathbb{R}^2$ at seven different points $p_1, \ldots, p_7$. Each of the tangent spaces $T_p\mathbb{R}^2$ is **isomorphic** to ("the same as") $\mathbb{R}^2$. Though we have drawn the tangent spaces at only seven points, there is in fact a separate tangent space $T_p\mathbb{R}^2$ for every single point $p \in \mathbb{R}^2$.

Figure 2.16 tries to help you imagine the tangent spaces to $\mathbb{R}^3$. Again, the manifold $\mathbb{R}^3$ is pictured along with the tangent spaces $T_{p_i}\mathbb{R}^3$ at seven different points $p_1, \ldots, p_7$. Each of the tangent spaces $T_p\mathbb{R}^3$ is naturally isomorphic to $\mathbb{R}^3$. Again, though we have drawn the tangent spaces at only seven points, there is in fact a separate tangent space $T_p\mathbb{R}^3$ for every single point $p \in \mathbb{R}^3$.

From now on we will often include the base point of a vector in our notation. For example, the vector v_p means vector v at the base point p of the manifold M. Clearly, the vector v_p belongs to, or is an element of, the tangent space T_pM. So, for example, we have

$$\begin{bmatrix} 1 \\ 1 \end{bmatrix}_{(-2,-3)} \in T_{(-2,-3)}\mathbb{R}^2, \qquad \begin{bmatrix} 1 \\ 1 \end{bmatrix}_{(-1,4)} \in T_{(-1,4)}\mathbb{R}^2, \qquad \begin{bmatrix} 1 \\ 1 \end{bmatrix}_{(2,4)} \in T_{(2,4)}\mathbb{R}^2,$$

and

$$\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}_{(-2,-3,-5)} \in T_{(-2,-3,-5)}\mathbb{R}^3, \qquad \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}_{(-1,4,3)} \in T_{(-1,4,3)}\mathbb{R}^3, \qquad \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}_{(2,4,-1)} \in T_{(2,4,-1)}\mathbb{R}^3.$$

The manifold M, along with a copy of the vector space T_pM attached at every point of the manifold is called the **tangent bundle** of the manifold M which is denoted as TM or $T(M)$. Often a cartoon version of the tangent bundle of a manifold M is drawn as in Fig. 2.17 using a one dimensional representation for both the manifold and the tangent spaces.

If M is an n dimensional manifold then TM is $2n$ dimensional. That makes intuitive sense since to specify exactly any element (vector) in TM you need to specify the vector part, which requires n numbers, and the base point, which also requires n numbers. Thus, $T\mathbb{R}^2$ has four dimensions and $T\mathbb{R}^3$ has six dimensions. In fact, the tangent bundle TM of a manifold M is itself also a manifold. Thus, if M is an n dimensional manifold then TM is a $2n$ dimensional manifold.

Now we will very briefly introduce the concept of a **vector field**. Doubtless you have been exposed to the idea of a vector field in vector calculus where it is usually defined on $\mathbb{R}^2$ or $\mathbb{R}^3$ as a vector-valued function that assigns to each point $(x, y) \in \mathbb{R}^2$ or $(x, y, z) \in \mathbb{R}^3$ a vector

$$\begin{bmatrix} v_1 \\ v_2 \end{bmatrix}_{(x,y)} \text{ or } \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix}_{(x,y,z)}.$$

An example of a vector field on $\mathbb{R}^2$ is shown in Fig. 2.18.

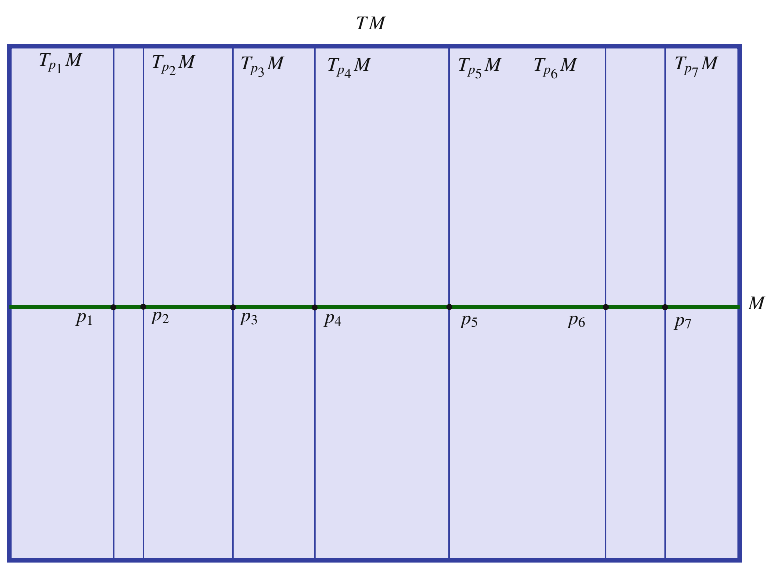


Fig. 2.17 A cartoon of a tangent bundle TM over a manifold M. Seven tangent spaces $T_{p_i}M$ are shown

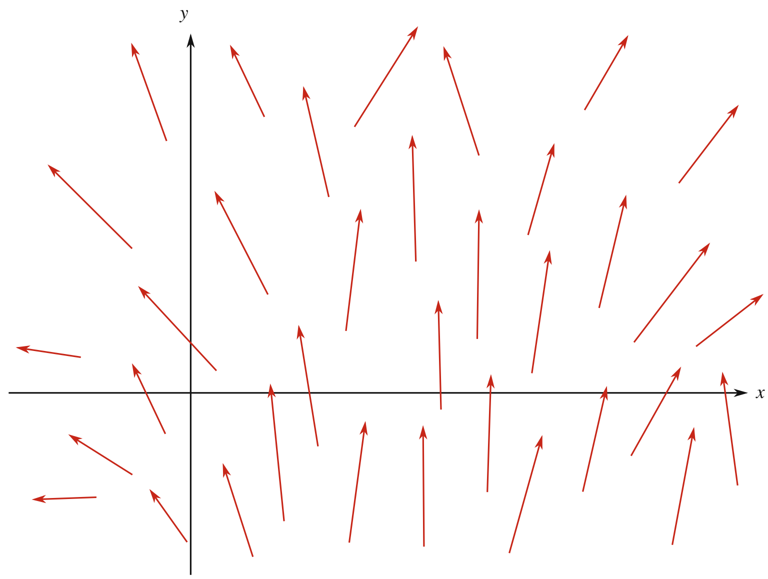


Fig. 2.18 A vector field drawn on the manifold $\mathbb{R}^2$. Recall, each of the vectors show are an element of a tangent space and not actually in the manifold $\mathbb{R}^2$

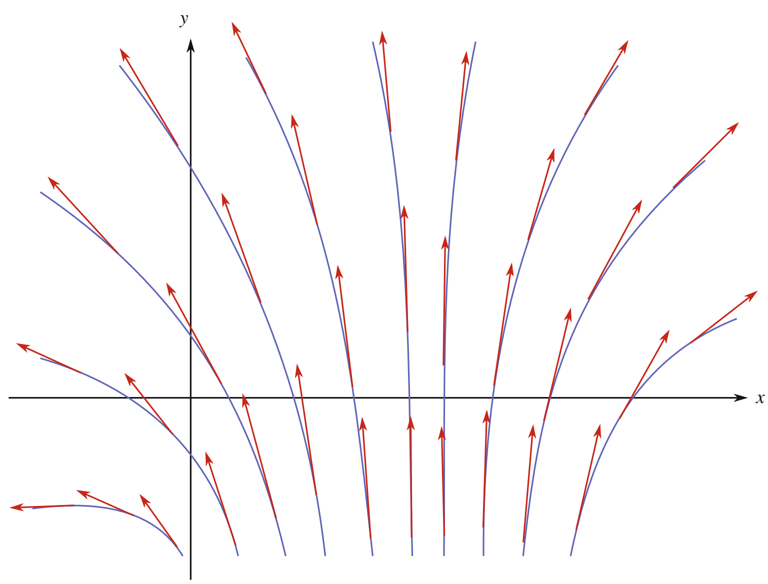


Fig. 2.19 A smooth vector field drawn on the manifold $\mathbb{R}^2$. Again, recall, each of the vectors shown are an element of a tangent space and not actually in the manifold $\mathbb{R}^2$

A vector field is called **smooth** if it is possible to find curves on the manifold such that all the vectors in the vector field are tangent to the curves. These curves are called the **integral curves** of the vector field, and finding them is essentially what differential equations are all about. Figure 2.19 shows several integral curves of a smooth vector field in $\mathbb{R}^3$.

While we certainly are not going to get into differential equations here, what we want to do is introduce a new way of thinking about vector fields. Notice, a vector field on a manifold such as $\mathbb{R}^2$ or $\mathbb{R}^3$ gives us a vector at each point of the manifold. But a vector v_p at a point p of a manifold M is an element of the tangent space at that point, T_pM. So being given a vector field is the same thing as being given an element of each tangent space of M. This is often called a **section** of the tangent bundle TM. Graphically, we often visualize sections of tangent bundles as in Fig. 2.20.

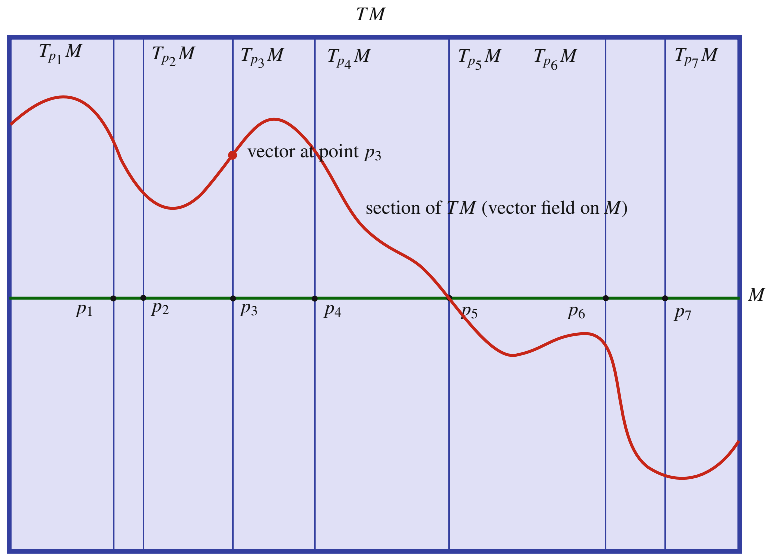


Fig. 2.20 A cartoon of a section of a tangent bundle over the manifold M. The section of TM is simply a vector field on M. The vector in the vector field on M at point p_3 is depicted as the point in $T_{p_3}M$ that is in the section. Notice, in this cartoon representation the vector at the point p_5 would be the zero vector

Question 2.1 Before we get into the next section let's take just a moment to review what derivatives were all about. Consider the function $f : \mathbb{R} \to \mathbb{R}$ given by $f(x) = \frac{x^2}{2} + 2$.

(a) Sketch the graph of the function $f(x) = \frac{x^2}{2} + 2$.
(b) Find the derivative of $f(x)$.
(c) What is $f(2)$? What is $f(3)$? Draw a line between the points $\big(2, f(2)\big)$ and $\big(3, f(3)\big)$.
(d) What is the slope of this secant line?
(e) Write down the limit of the difference quotient definition of derivative.
(f) Use this definition to find the derivative of f at $x = 2$. Does it agree with what you found above?
(g) The derivative of f at $x = 2$ is a number. Explain what this number represents.

From the above question you should now remember that derivatives are basically slopes of lines tangent to the graph of the function at a certain point.

2.3 Directional Derivatives

Now the time comes to explore why we might actually want to use all of this. Suppose we were given a real-valued function f on the manifold $\mathbb{R}^n$. That is, $f : \mathbb{R}^n \to \mathbb{R}$. How might we use a vector v_p at some given point p on the manifold? One natural thing we could do is take the **directional derivative** of f at the point p in the direction of v_p. Consider Fig. 2.21 where a real-valued function $f : \mathbb{R} \to \mathbb{R}$ is shown. Given a point p in the domain there are two directions that it is possible to move in, left or right. Similarly, if you had a vector based at that point it could either point left, as v_p does, or right, as w_p does. We could ask how the function varies in either of these directions. In Fig. 2.22 a real-valued function $f : \mathbb{R}^2 \to \mathbb{R}$ is depicted. Given a point $p = (x, y)$ in the domain one can go any direction in two dimensions, and a vector based at that point could point in any direction in two dimensions as well. And similarly, we could ask how the function varies in any of these directions. In Fig. 2.23 we choose one of those directions, the direction given by the vector

$$v_p = \begin{bmatrix} a \\ b \end{bmatrix}_{(x,y)}.$$

We are able to use this vector to find the directional derivative of f in the direction v_p at the point p. In other words, we are able to find how f is changing in the direction v_p at the point p.

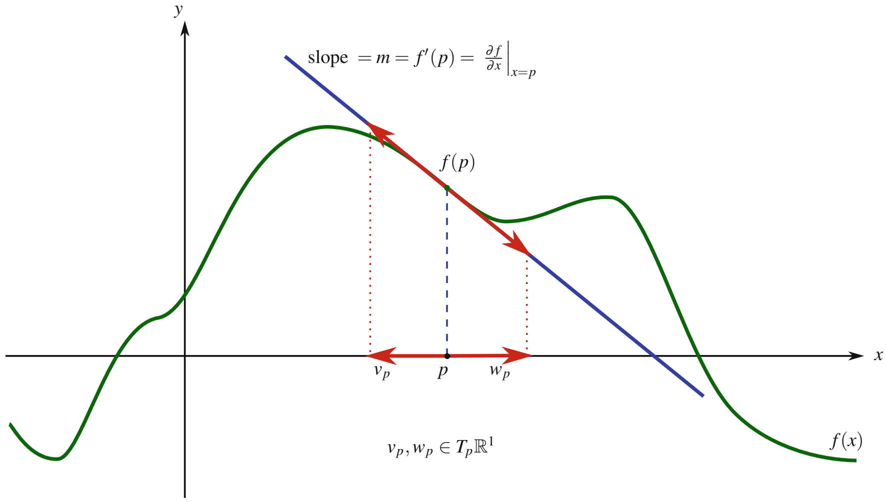


Fig. 2.21 Starting at $p \in \mathbb{R}$ one can go in any direction in one dimension. In other words, one can go either right or left

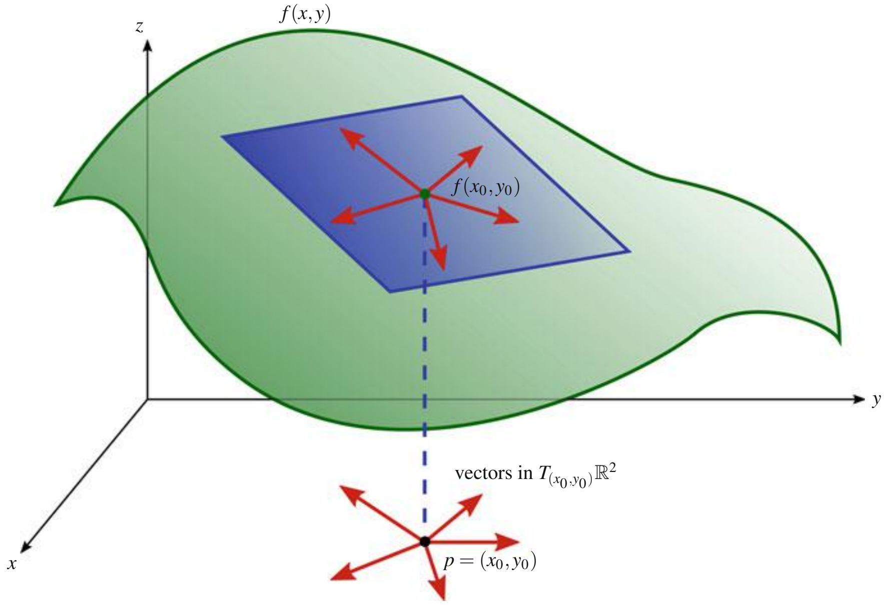


Fig. 2.22 Starting at $p = (x, y) \in \mathbb{R}^2$ one can go in any direction in two dimensions. A number of different directions are pictured

Notice, for the moment that we are going back to the vector calculus way of thinking, of thinking of the vector v_p as sitting inside the manifold $\mathbb{R}^2$ at the point p instead of in the tangent space $T_p\mathbb{R}^2$. Even though we just made a big deal about the vectors not being in the manifold but being in the tangent spaces instead, since we are reviewing the idea of directional derivatives from the viewpoint of vector calculus in this section and the next we will go back to thinking of vectors as being in the manifold based at a certain point. That way we can use the vector calculus notation and focus on the relevant ideas without being bogged down with notational nuance.

We start by considering how directional derivatives are generally introduced in most multivariable calculus courses. The reason we do this is that the definition of directional derivative that we will want to use will be slightly different than that used in most vector calculus classes and we do not want there to be any confusion. The definition that you probably saw was something like this.

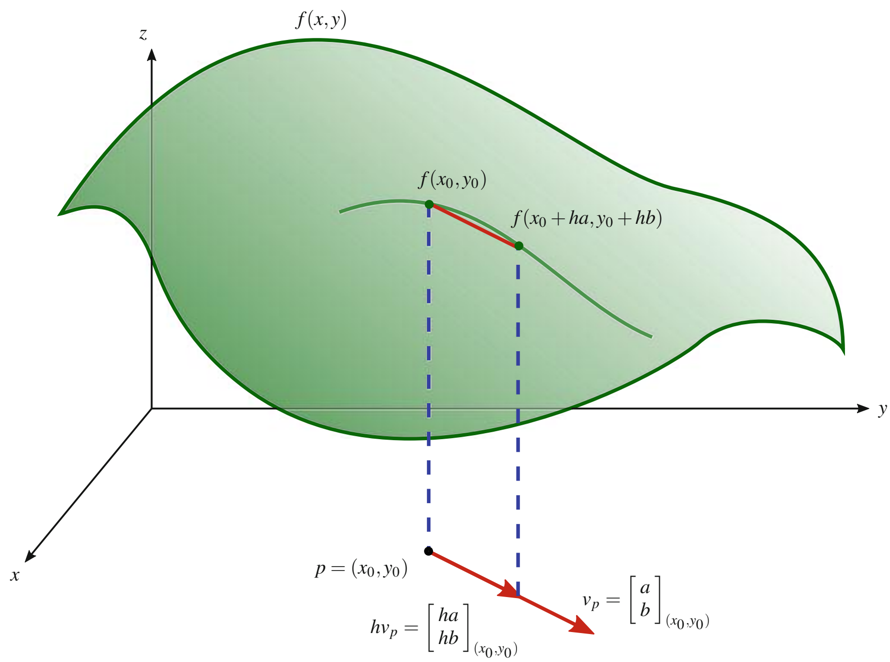


Fig. 2.23 Finding the directional derivative of a function $f : \mathbb{R}^2 \to \mathbb{R}$ requires a unit vector in some direction and the associated values for $f(t_0 + ha, y_0 + tb)$ and $f(x_0, y_0)$, which are necessary for the difference quotient

Definition 2.3.1 The **directional derivative** of $f : \mathbb{R}^2 \to \mathbb{R}$ at (x_0, y_0) in the direction of the unit vector $u = \begin{bmatrix} a \\ b \end{bmatrix}$ is

$$D_u f(x_0, y_0) = \lim_{t\to 0} \frac{f\big(x_0 + ta, y_0 + tb\big) - f\big(x_0, y_0\big)}{t}$$

if this limit exists.

Definition 2.3.2 The **directional derivative** of $f : \mathbb{R}^3 \to \mathbb{R}$ at (x_0, y_0, z_0) in the direction of the unit vector $u = \begin{bmatrix} a \\ b \\ c \end{bmatrix}$ is

$$D_u f(x_0, y_0, z_0) = \lim_{t\to 0} \frac{f\big(t_0 + ta, y_0 + tb, z_0 + tc\big) - f\big(x_0, y_0, z_0\big)}{t}$$

if this limit exists.

Often vector calculus textbooks will write vectors as row vectors, sometimes like $[a, b]$ or sometimes like $\langle a, b\rangle$. We will never do that in this book. We will always write vectors as column vectors with square brackets, $v = \begin{bmatrix} a \\ b \end{bmatrix}$. Also, we will never use the angle brackets $\langle\cdot,\cdot\rangle$ for vectors or points, which one sometimes sees. Finally, notice that in both of the definitions that the vector u had to be a unit vector, that means the length of the vector is one unit. In other words, for the first definition we have to have $\sqrt{a^2 + b^2} = 1$ and for the second we have to have $\sqrt{a^2 + b^2 + c^2} = 1$.

To remind ourselves of some other equivalent notations, notice that if we let $p = (x_0, y_0)$ then we can also write

$$\lim_{t\to 0} \frac{f\big(x_0 + ta, y_0 + tb\big) - f\big(x_0, y_0\big)}{t} = \frac{d}{dt}\Big(f\big(p + tu\big)\Big)\Big|_{t=0}$$

and if we let $p = (x_0, y_0, z_0)$ then we can also write

$$\lim_{t\to 0} \frac{f\big(x_0 + ta, y_0 + tb, z_0 + tc\big) - f\big(x_0, y_0, z_0\big)}{t} = \frac{d}{dt}\Big(f\big(p + tu\big)\Big)\Big|_{t=0}.$$

That means that we could also have written the above definitions as

$$D_u f(x_0, y_0) = \frac{d}{dt}\Big(f\big(p + tu\big)\Big)\Big|_{t=0}$$

and

$$D_u f(x_0, y_0, z_0) = \frac{d}{dt}\Big(f\big(p + tu\big)\Big)\Big|_{t=0}.$$

Suppose we had a function $f : \mathbb{R}^2 \to \mathbb{R}$ and the unit vectors

$$u = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, v = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

then it is easy to see that $D_u f = \frac{\partial f}{\partial x}$ and $D_v f = \frac{\partial f}{\partial y}$. Similarly, for a function $g : \mathbb{R}^3 \to \mathbb{R}$ and the unit vectors

$$u = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, v = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, w = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$$

we get $D_u g = \frac{\partial g}{\partial x}$, $D_v g = \frac{\partial g}{\partial y}$, and $D_w g = \frac{\partial g}{\partial z}$.

Question 2.2 Find the derivative of $f(x, y) = x^3 - 3xy + y^2$ at $p = (1, 2)$ in the direction of $u = \begin{bmatrix} \cos(\pi/6) \\ \sin(\pi/6) \end{bmatrix}$.

Proceeding as you probably did in multivariable calculus, after the above definitions of directional derivatives were made the following theorems were derived.

Theorem 2.1 *If $f : \mathbb{R}^2 \to \mathbb{R}$ is a differentiable function of x and y, then f has directional derivatives in the direction of any unit vector $u = \begin{bmatrix} a \\ b \end{bmatrix}$ and*

$$D_u f(x, y) = \frac{\partial f}{\partial x}(x, y) \cdot a + \frac{\partial f}{\partial y}(x, y) \cdot b.$$

Sketch of proof: First define a one variable function $g(h)$ by $g(h) = f(x_0 + ha, y_0 + hb)$ and use this to rewrite $D_u f$. Then use the chain rule to write the derivative of $g(h)$. Finally, let $h = 0$ and combine.

Theorem 2.2 *If $f : \mathbb{R}^3 \to \mathbb{R}$ is a differentiable function of x, y and z, then f has directional derivatives in the direction of any unit vector $u = \begin{bmatrix} a \\ b \\ c \end{bmatrix}$ and*

$$D_u f(x, y, z) = \frac{\partial f}{\partial x}(x, y, z) \cdot a + \frac{\partial f}{\partial y}(x, y, z) \cdot b + \frac{\partial f}{\partial z}(x, y, z) \cdot c.$$

Question 2.3 Let $p = (2, 0, -1)$ and $v = \frac{1}{\sqrt{14}}\begin{bmatrix} 2 \\ -1 \\ 3 \end{bmatrix}$. Find $D_{v_p} f$ if

(a) $f(x, y, z) = y^2 z$
(b) $f(x, y, z) = e^x \cos(y)$
(c) $f(x, y, z) = x^7$

Now we want to explore the reason why the stipulation that

$$u = \begin{bmatrix} a \\ b \end{bmatrix} \quad \text{or} \quad u = \begin{bmatrix} a \\ b \\ c \end{bmatrix}$$

be a unit vector is made. We begin by noting that the equation of a plane through the origin is given by

$$z = m_x x + m_y y$$

as shown in Fig. 2.24. From this it is straightforward to show that the equation of a plane through the point $(x_0, y_0, f(x_0, y_0))$ on the graph of f is given by

$$z - f(x_0, y_0) = m_x(x - x_0) + m_y(y - y_0).$$

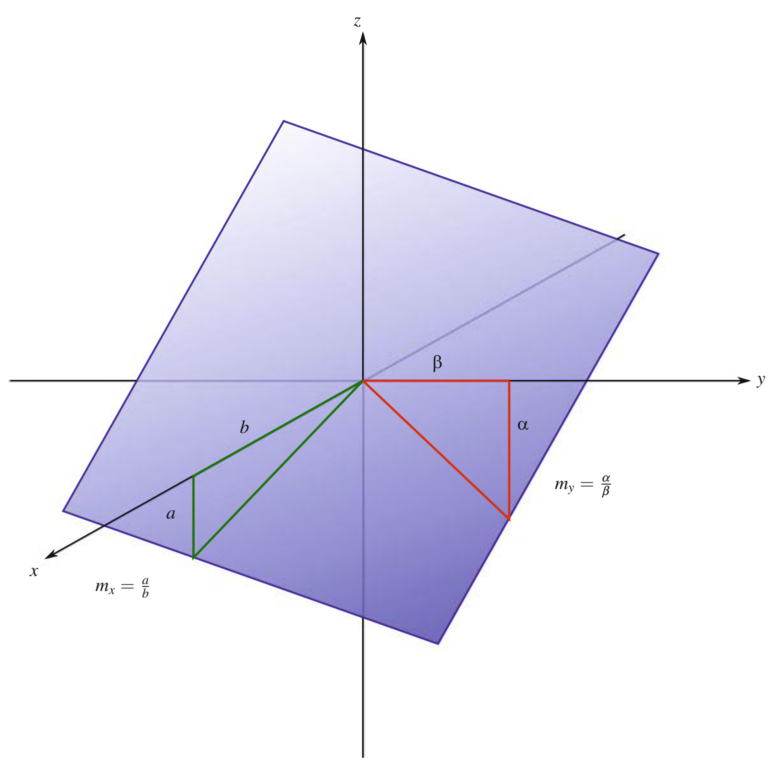


Fig. 2.24 The equation of a plane through the origin is given by $z = m_x x + m_y y$ where m_x is the slope of the plane along the x-axis and m_y is the slope of the plane along the y-axis

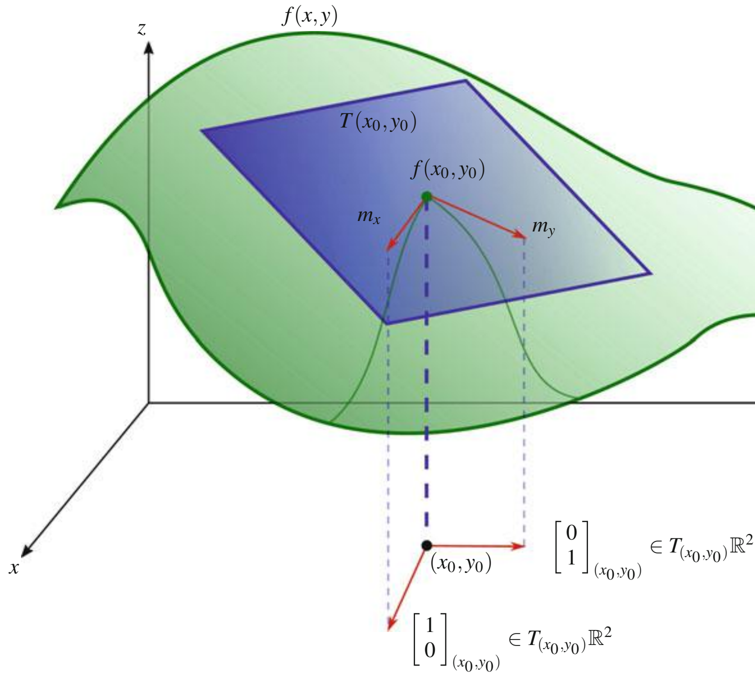


Fig. 2.25 The tangent plane $T(x_0, y_0)$ to $f(x, y)$ at point $(x_0, y_0, f(x_0, y_0))$

To get the equation of the tangent plane $T(x_0, y_0)$ to $f(x, y)$ at point $(x_0, y_0, f(x_0, y_0))$ we notice that

$$m_x = \frac{\partial f}{\partial x}, \quad m_y = \frac{\partial f}{\partial y}$$

$$\Rightarrow T(x, y) - f(x_0, y_0) = \left.\frac{\partial f}{\partial x}\right|_{(x_0, y_0)} \cdot (x - x_0) + \left.\frac{\partial f}{\partial y}\right|_{(x_0, y_0)} \cdot (y - y_0),$$

$$\Rightarrow T(x, y) = f(x_0, y_0) + \left.\frac{\partial f}{\partial x}\right|_{(x_0, y_0)} \cdot (x - x_0) + \left.\frac{\partial f}{\partial y}\right|_{(x_0, y_0)} \cdot (y - y_0),$$

which is shown in Fig. 2.25.

Consider the vector $u = \begin{bmatrix} a \\ b \end{bmatrix}$ at the point (x_0, y_0), that is, $u = \begin{bmatrix} a \\ b \end{bmatrix}_{(x_0, y_0)}$ as shown in Fig. 2.26. How would we find the slope of the line tangent to the graph of f through point $f(x_0, y_0)$? We would use $m = \frac{rise}{run}$, where clearly the "run" is the length of the vector $\begin{bmatrix} a \\ b \end{bmatrix}$. Our rise, or change in height z, can be calculated from the equation of the tangent plane to the point that we developed above,

$$\begin{aligned}
\frac{rise}{run} &= \frac{T(x_0 + a, y_0 + b) - T(x_0, y_0)}{\sqrt{a^2 + b^2}} \\
&= \frac{\left.\frac{\partial f}{\partial x}\right|_{(x_0, y_0)} \cdot a + \left.\frac{\partial f}{\partial y}\right|_{(x_0, y_0)} \cdot b}{\sqrt{a^2 + b^2}} \\
&= \frac{\frac{\partial f}{\partial x} \cdot a + \frac{\partial f}{\partial y} \cdot b}{\sqrt{a^2 + b^2}}.
\end{aligned}$$

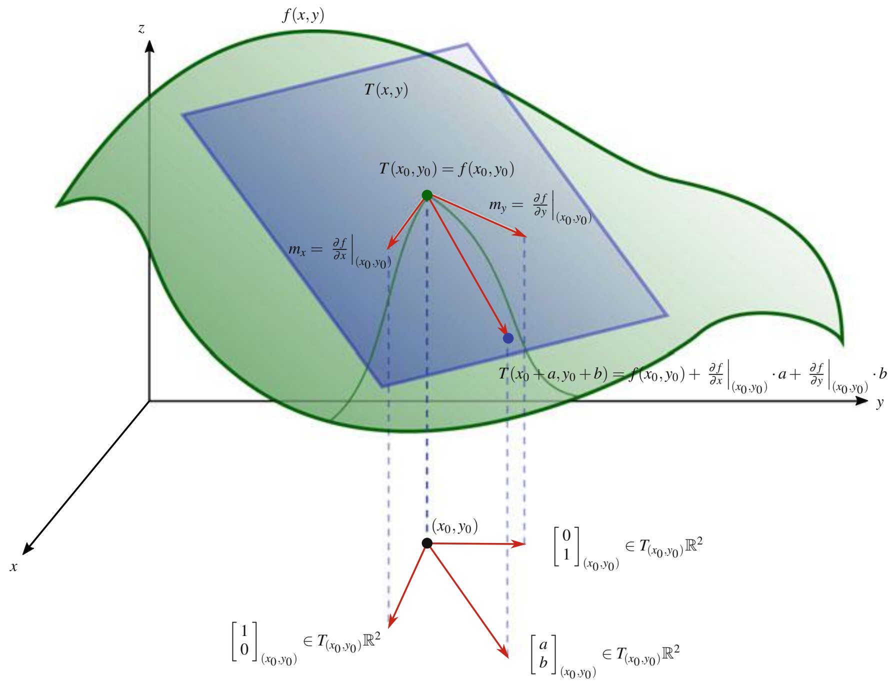


Fig. 2.26 The tangent plane to $f(x, y)$ at the point (x_0, y_0) shown with the tangent line in the direction $v = ae_1 + be_2$

Notice how close this is to the formula given for $D_u f$ in multivariable calculus,

$$D_u f = \frac{\partial f}{\partial x} \cdot a + \frac{\partial f}{\partial y} \cdot b.$$

As long as the vector $\begin{bmatrix} a \\ b \end{bmatrix}$ is a unit vector then these two formulas are the same. Thus if u is a unit vector the formula for $D_u f$ also gives the slope of the tangent line that lies in the tangent plane and is in the direction of $\begin{bmatrix} a \\ b \end{bmatrix}$. So by stipulating that $u = \begin{bmatrix} a \\ b \end{bmatrix}$ be a unit vector ***we basically maintain that the definition of derivative also be the slope of the tangent line while simultaneously retaining a nice simple formula for the derivative***.

We will now loosen this requirement and allow any vector, not just unit vectors. However, we want to keep the nice formula. This means that we need *let go* of the idea that the derivative is the slope of the tangent line. So we will define the directional derivative exactly as before but we will drop the stipulation that u must be a unit vector. In order to obtain the slope of the tangent line we will have to divide the value given by our new definition of directional derivative by the length of the vector. So what exactly will our new definition of directional derivative actually represent? It will represent the "rise" portion of the above equation - how much "rise" the tangent line has over the length of the vector u.

Consider Fig. 2.26. By taking away the requirement that u be a unit vector the directional derivative

$$D_u f = \frac{\partial f}{\partial x} \cdot a + \frac{\partial f}{\partial y} \cdot b.$$

tells us how much the tangent plane to the graph of the function f at the point (x_0, y_0) "rises" as we move along u. Thus the directional derivative actually tells us something about the tangent plane to the function at the point. We also take a moment to point out that the tangent plane to the function f at the point (x_0, y_0) is actually the closest linear approximation of the graph of the function f at (x_0, y_0).

Now take a moment to consider our review of how directional derivatives are generally introduced in multivariable calculus classes. At no point was there any mention of the tangent space of the manifolds $\mathbb{R}^2$ or $\mathbb{R}^3$. One can, and does, mix-up, or merge, the tangent spaces with the underlying manifold in multivariable calculus because they are essentially the same space. However, eventually when we start to deal with other kinds of manifolds we will not be able to do this. That is why we make such an effort to always make the distinction.

Definition 2.3.3 Let $f : \mathbb{R}^n \to \mathbb{R}$ be a real-valued function on the manifold $\mathbb{R}^n$ and let v_p be a vector tangent to manifold $\mathbb{R}^n$, that is, $v_p \in T_p(\mathbb{R}^n)$. The number

$$v_p[f] \equiv \frac{d}{dt}\Big(f\big(p + tv_p\big)\Big)\bigg|_{t=0}$$

is called the **directional derivative** of f with respect to v_p, if it exists.

This is exactly the same definition as in the multivariable calculus case, only without the requirement that v_p be a unit vector. Thus, the number that we get from this definition of directional derivative will *not* be the slope of the tangent line if v_p is not a unit vector but will instead be the "rise" of the tangent line to the graph of f at the point p as we move the length of v_p.

Also notice the new notation, $v_p[f]$. This notation is very similar to the functional notation $f(x)$ where the function f "takes in" an input value x and gives out a number. In the notation $v_p[f]$ the vector v_p "takes in" a function f and gives out a number; the directional derivative of f in the direction v_p (at the point p.) So, the vector v_p becomes what is called an **operator** on the function f. Basically, an operator is a function that takes as inputs other functions.

Question 2.4 Let $v = \begin{bmatrix} 2 \\ -1 \\ 3 \end{bmatrix}$ and $p = (2, 0, -1)$.

(a) Find $v_p[f]$ where
   (i) $f(x) = x$,
   (ii) $f(x) = x^2 - x$,
   (iii) $f(x) = \cos(x)$.

(b) For each of the functions above, find $w_p[f]$ where
   (i) $w_p = 2v_p$,
   (ii) $w_p = \frac{1}{2}v_p$,
   (i) $w_p = -5v_p$,
   (ii) $w_p = \frac{-1}{5}v_p$.

Suppose that we have a function $f : \mathbb{R}^3 \to \mathbb{R}$ and a vector $v_p = \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix}_p \in T_p(\mathbb{R}^3)$ at the point $p = (p_1, p_2, p_3)$ and we want to find an expression for $v_p[f]$. First, write $p + tv_p$ as $p + tv_p = (p_1 + tv_1, p_2 + tv_2, p_3 + tv_3)$, which then gives us $f(p + tv_p) = f\big(p_1 + tv_1, p_2 + tv_2, p_3 + tv_3\big)$. Next we notice that if we define the functions $x_i(t) = p_i + tv_i$, for $i = 1, 2, 3$, then we have $\frac{dx_i}{dt} = \frac{d}{dt}\big(p_i + tv_i\big) = v_i$. Putting all of this together and using the chain rule we find

$$\begin{aligned}
v_p[f] &= \frac{d}{dt} f(p + tv)\bigg|_{t=0} \\
&= \frac{d}{dt} f\big(\underbrace{p_1 + tv_1}_{x_1(t)}, \underbrace{p_2 + tv_2}_{x_2(t)}, \underbrace{p_3 + tv_3}_{x_3(t)}\big)\bigg|_{t=0} \\
&= \frac{\partial f}{\partial x_1}\bigg|_{x_1(0)} \cdot \frac{dx_1}{dt}\bigg|_{t=0} + \frac{\partial f}{\partial x_2}\bigg|_{x_2(0)} \cdot \frac{dx_2}{dt}\bigg|_{t=0} + \frac{\partial f}{\partial x_3}\bigg|_{x_3(0)} \cdot \frac{dx_3}{dt}\bigg|_{t=0} \\
&= \frac{\partial f}{\partial x_1}\bigg|_{p_1} \cdot v_1 + \frac{\partial f}{\partial x_2}\bigg|_{p_2} \cdot v_2 + \frac{\partial f}{\partial x_3}\bigg|_{p_3} \cdot v_3 \\
&= \sum_{i=1}^{3} v_i \cdot \frac{\partial f}{\partial x_i}\bigg|_p .
\end{aligned}$$

In summary, in this example we have just found that

$$v_p[f] = \sum_{i=1}^{3} v_i \cdot \left. \frac{\partial f}{\partial x_i} \right|_p .$$

In particular, suppose we had $f : \mathbb{R}^3 \to \mathbb{R}$ a real-valued function on the manifold $\mathbb{R}^3$ along with the standard Euclidian vectors

$$e_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \; e_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \; e_3 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}.$$

For the moment we will suppress the base point from the notation since it plays no particular role. Since $v[f] = \sum_{i=1}^{3} v_i \cdot \frac{\partial f}{\partial x_i}$ we have $e_1[f] = \frac{\partial f}{\partial x_1}$, $e_2[f] = \frac{\partial f}{\partial x_2}$, and $e_3[f] = \frac{\partial f}{\partial x_3}$ We have actually just shown something very interesting. With a very slight change in notation it becomes even more obvious

$$e_1[f] = \frac{\partial}{\partial x_1}\Big(f\Big), \; e_2[f] = \frac{\partial}{\partial x_2}\Big(f\Big), \; e_3[f] = \frac{\partial}{\partial x_3}\Big(f\Big).$$

We have just equated the operator $\frac{\partial}{\partial x_i}$ with the euclidian vector e_i. In a single short paragraph we have shown one of the most important ideas in this book.

$$\boxed{\text{The Euclidian vectors } e_i \text{ can be identified with the partial differential operators } \frac{\partial}{\partial x_i}}$$

In other words, we can think of the Euclidian vector e_i as actually being the partial differential operator $\frac{\partial}{\partial x_i}$.

Question 2.5 Redo the above calculation putting the base point p into the notation in order to convince yourself that the base point p does not play any role in the calculation.

Question 2.6 Suppose that $a, b \in \mathbb{R}$, p a point on manifold $\mathbb{R}^3$, $v_p, w_p \in T_p(\mathbb{R}^3)$ and $f, g : \mathbb{R}^3 \to \mathbb{R}$.

(a) Show that $(av_p + bw_p)[f] = av_p[f] + bw_p[f]$.
(b) Show that $v_p[af + bg] = av_p[f] + bv_p[f]$.

Besides the identities in the above question that hold, something else, called the **Leibnitz rule**, holds. The Leibnitz rule says that $v_p[fg] = v_p[f] \cdot g(p) + f(p) \cdot v_p[g]$. In essence the Leibnitz rule is the product rule, which is actually used in proving this identity,

$$\begin{aligned}
v_p[fg] &= \sum_{i=1}^{3} v_i \cdot \left. \frac{\partial fg}{\partial x_i} \right|_{p_i} \\
&= \sum_{i=1}^{3} v_i \cdot \left(\left. \frac{\partial f}{\partial x_i} \right|_p \cdot g(p) + f(p) \cdot \left. \frac{\partial g}{\partial x_i} \right|_p \right) \\
&= \left(\sum_{i=1}^{3} v_i \left. \frac{\partial f}{\partial x_i} \right|_p \right) g(p) + f(p) \left(\sum_{i=1}^{3} v_i \left. \frac{\partial g}{\partial x_i} \right|_p \right) \\
&= v_p[f] \cdot g(p) + f(p) \cdot v_p[g].
\end{aligned}$$

First, we take a moment to point out a notational ambiguity. When we are giving a vector v we usually use the subscript to indicate the base point the vector is at, so v_p means that the vector v is based at point p. But here, for the Euclidian unit vectors e_1, e_2, e_3 the subscript clearly does not refer to a point, it refers to which of the Cartesian coordinates are non-zero (in fact, are one.) The Euclidian unit vectors are so common and the notation so useful we usually omit the base point p when using them. One could of course write e_{1_p}, e_{2_p}, and e_{3_p} but that starts getting a little silly after a while. Usually it should be clear from the context what the base point actually is.

Now that we have that little notational comment made, we can bask in the glow of what we have just accomplished, and try to wrap our heads around it at the same time. This last few pages, culminating in the last example, is actually a lot more profound than it seems, so we will take a moment to reiterate what has just been done.

First, you have seen vectors before and are probably fairly comfortable with them, at least vectors in n-dimensional Euclidian space. Of course, now we are viewing vectors as being elements of the tangent space, but they are still just good old-fashioned vectors. Second, you have seen differentiation before and are probably comfortable with differential operators, things like $\frac{d}{dx}, \frac{d}{dt}, \frac{\partial}{\partial x}, \frac{\partial}{\partial y}$, though it is also possible you have not seen the word operator applied to them before.

Given a function $f : \mathbb{R}^3 \to \mathbb{R}$ we have used the act of taking the directional derivative of f at the point $p \in \mathbb{R}^3$ in the direction of $v_p \in T_p\mathbb{R}^3$ to turn the vector $v_p \in T_p\mathbb{R}^3$ into an operator on f. This has given the following formula

$$v_p[f] = \sum_{i=1}^{3} v_i \cdot \left. \frac{\partial f}{\partial x_i} \right|_{p_i}.$$

This formula was then used to see exactly what the operator for the Euclidian unit vectors e_1, e_2, e_3 would be at the point p and we determined that we can make the following identifications

$$\begin{aligned} e_1 &\equiv \frac{\partial}{\partial x_1}, \\ e_2 &\equiv \frac{\partial}{\partial x_2}, \\ e_3 &\equiv \frac{\partial}{\partial x_3}. \end{aligned}$$

This is, if you have not seen it before, really a surprising identification. It doubtless feels a bit like a notational slight-of-hand. But this is one of the powers of just the right notation, it helps you understand relationships that may be difficult to grasp otherwise. By viewing a vector in the tangent space as acting on a function via the directional derivative we have equated vectors with partial differential operators. See Fig. 2.27. This is an identification that we will continue to make for the rest of the book.

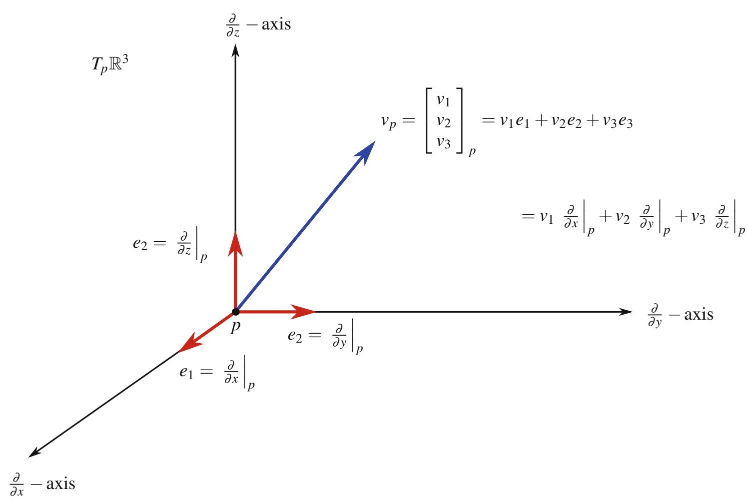


Fig. 2.27 Identifying the euclidian vectors e_i with the partial differential operators $\partial/\partial x_i$

Putting this identification into action consider the vector

$$v = \begin{bmatrix} 2 \\ -3 \\ 1 \end{bmatrix}.$$

When v is written using the Euclidian unit vectors e_1, e_2, e_2 we have

$$v = 2e_1 - 3e_2 + e_3$$

and when it is written using the differential operators we have

$$v = 2\frac{\partial}{\partial x_1} - 3\frac{\partial}{\partial x_2} + \frac{\partial}{\partial x_3}.$$

From now on you we shall make little or no distinction between e_1, e_2, e_3 and $\frac{\partial}{\partial x_1}, \frac{\partial}{\partial x_2}, \frac{\partial}{\partial x_3}$, using either notation as warranted. You should also note, often you may see $\frac{\partial}{\partial x_1}, \frac{\partial}{\partial x_2}, \frac{\partial}{\partial x_3}$ written as $\partial_1, \partial_2, \partial_3$, or $\partial_{x_1}, \partial_{x_2}, \partial_{x_3}$, or even $\partial_x, \partial_y, \partial_z$. If we want to take into account the base point, say

$$v_p = \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix}_p \in T_p(\mathbb{R}^3)$$

then we would write

$$v_p = v_1 \left.\frac{\partial}{\partial x_1}\right|_p + v_2 \left.\frac{\partial}{\partial x_2}\right|_p + v_3 \left.\frac{\partial}{\partial x_3}\right|_p .$$

2.4 Differential One-Forms

Let us recap what we have done. First we reviewed vector spaces and dual spaces. After that we made a distinction between manifolds (manifold $\mathbb{R}^3$) and vector spaces (vector space $\mathbb{R}^3$) and introduced the concept of the tangent space $T_p(\mathbb{R}^3)$, which is basically a vector space $\mathbb{R}^3$ attached to each point p of manifold $\mathbb{R}^3$. After that we reviewed directional derivatives and used them to discover an identity between vectors and partial differential operators. We discovered that each vector $v_p \in T_p(\mathbb{R}^3)$ was exactly a differential operator

$$v_p = \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix}_p = v_1 \left.\frac{\partial}{\partial x_1}\right|_p + v_2 \left.\frac{\partial}{\partial x_2}\right|_p + v_3 \left.\frac{\partial}{\partial x_3}\right|_p .$$

Recognizing that each tangent space $T_p(\mathbb{R}^n)$ is itself a vector space it should be obvious that each tangent space $T_p(\mathbb{R}^n)$ has its own dual space, which is denoted $T_p^*(\mathbb{R}^n)$. It is this space that we are now going to look at.

We are now ready for the definition we have all been waiting for, the definition of a differential one-form.

Definition 2.4.1 A **differential one-form** α on manifold $\mathbb{R}^n$ is a linear functional on the set of tangent vectors to the manifold $\mathbb{R}^n$. That is, at each point p of manifold $\mathbb{R}^n$, $\alpha : T_p(\mathbb{R}^n) \to \mathbb{R}$ and

$$\alpha(v_p + w_p) = \alpha(v_p) + \alpha(w_p),$$
$$\alpha(av_p) = a\alpha(v_p)$$

for all $v_p, w_p \in T_p(\mathbb{R}^n)$ and $a \in \mathbb{R}$.

There are a number of comments that need to be made to help you understand this definition better. First of all, we have already defined linear functionals when we discussed vector spaces. The set of all linear functionals of a vector space is called the **dual space** of that vector space. In the definition of one-forms the vector space was $T_p(\mathbb{R}^n)$. The dual space of $T_p(\mathbb{R}^n)$ is generally denoted as $T_p^*(\mathbb{R}^n)$. Thus, we have $\alpha \in T_p^*(\mathbb{R}^n)$.

The next thing to notice is that in the first required identity in the definition the addition on the left of the equal sign takes place in the vector space $T_p(\mathbb{R}^n)$. In other words, $v_p + w_p \in T_p(\mathbb{R}^n)$. The addition on the right of the equal sign takes place in the reals $\mathbb{R}$. That is, $\alpha(v_p) + \alpha(w_p) \in \mathbb{R}$. There is also a slight peculiarity with the terminology; α is called a **differential one-form on the manifold** $\mathbb{R}^n$ even though its inputs are vectors, elements of the manifold's tangent space $T_p(\mathbb{R}^n)$ at some point p in the manifold $\mathbb{R}^n$. So, even though α eats elements of the tangent spaces $T_p(\mathbb{R}^n)$ of manifold $\mathbb{R}^n$ it is still called a differential one-form **on** manifold $\mathbb{R}^n$. If we want to specify the point the differential form is at, we will use either α_p or $\alpha(p)$. Clearly we have $\alpha_p \in T_p^*(\mathbb{R}^n)$, the dual space of $T_p(\mathbb{R}^n)$.

Now, why is it called a differential one-form? The one- is added because it takes as its input only one tangent vector. Later on we will meet two-forms, three-forms, and general k-forms that take as inputs two, three, or k tangent vectors. The word differential is used because of the intimate way forms are related to the idea of exterior differentiation, which we will study later. Finally, the word form is a somewhat generic mathematical term that gets applied to a fairly wide range of objects (bilinear forms, quadratic forms, multilinear forms, the first and second fundamental forms, etc.) Thus, a **differential** k**-form** is a mathematical object whose input is k vectors at a point and which has something to do with an idea about differentiation. Very often the word differential is dropped and differential k-forms are simply referred to as k-**forms**. So far this definition is a little abstract so in order to gain a better understanding we will now consider some concrete but simple examples of one-forms on the manifold $\mathbb{R}^3$. But before we can write down concrete examples of one-forms on manifold $\mathbb{R}^3$ we need to decide on a basis for $T_p^*(\mathbb{R}^3)$.

Based on our review of vector spaces and the notation we used there, the **dual basis** is was written as the vector space basis with superscripts instead of subscripts. Suppose we wrote the basis of $T_p(\mathbb{R}^3)$ as $\left\{e_{1_p}, e_{2_p}, e_{3_p}\right\}$ then we could possibly try to write the basis of $T_p^*(\mathbb{R}^3)$ as $\left\{e_p^1, e_p^2, e_p^3\right\}$. But recalling the identification between the standard Euclidian unit vectors and the differential operators, we could also write the bases of $T_p(\mathbb{R}^3)$ as

$$\left\{ \left.\frac{\partial}{\partial x_1}\right|_p, \left.\frac{\partial}{\partial x_2}\right|_p, \left.\frac{\partial}{\partial x_3}\right|_p \right\}.$$

But how would we write the dual basis now? We will write the dual basis, that is, the basis of the dual space $T_p^*(\mathbb{R}^3)$, as

$$\left\{dx_{1_p}, dx_{2_p}, dx_{3_p}\right\}.$$

Yes, this basis looks like the differentials that you are familiar with from calculus. There is a reason for that. Once we have worked a few problems and you are a little more comfortable with the notation we will explore the reasons for the notation. Dropping the base point p from the notation we will write

$$\left\{dx_1, dx_2, dx_3\right\}$$

as the basis of $T^*(\mathbb{R}^3)$ dual to the basis of $T(\mathbb{R}^3)$,

$$\left\{\frac{\partial}{\partial x_1}, \frac{\partial}{\partial x_2}, \frac{\partial}{\partial x_3}\right\}.$$

Figure 2.28 is an attempt to pictorially draw the manifold $\mathbb{R}^3$, a tangent space at a point p, $T_p\mathbb{R}^3$, and its dual space $T_p^*\mathbb{R}^3$. Often we think of the tangent space $T_p\mathbb{R}^3$ and the dual space $T_p^*(\mathbb{R}^3)$ as being "attached" to the manifold $\mathbb{R}^3$ at the point p, but we have drawn the dual space $T_p^*\mathbb{R}^3$, directly above the tangent space $T_p\mathbb{R}^3$. The dual space $T_p^*\mathbb{R}^3$ is very often called the **cotangent space** at p.

Since any element of $T_p^*\mathbb{R}^3$ can be written as a linear combination of $\{dx_1, dx_2, dx_3\}$ and elements of $T_p^*\mathbb{R}^3$ are one-forms this of course implies that the one-forms on $\mathbb{R}^3$ can be written as linear combinations of $\{dx_1, dx_2, dx_3\}$. In particular, the basis elements dx_1, dx_2, and dx_3 are themselves one-forms on $\mathbb{R}^3$.

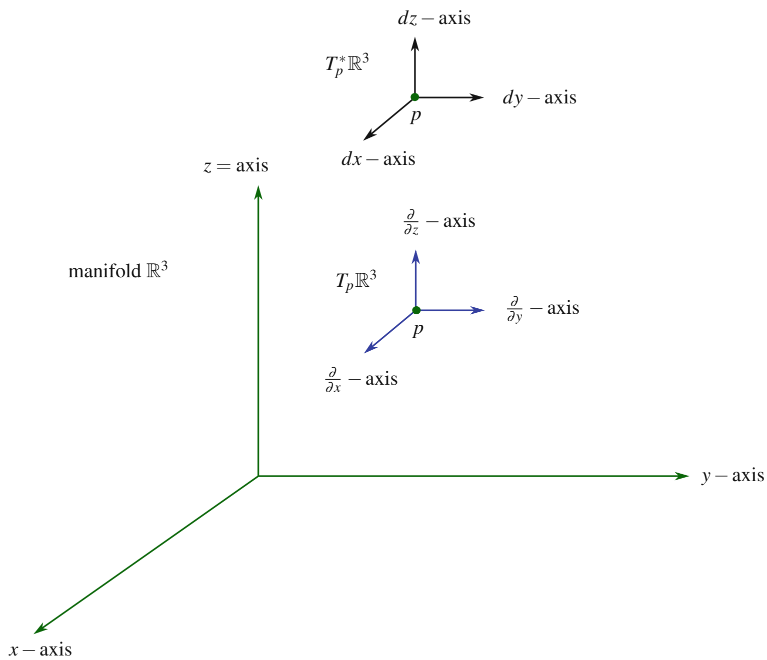


Fig. 2.28 An illustration of the manifold $\mathbb{R}^3$ along with the tangent space $T_p\mathbb{R}^3$ "attached" to the manifold at point p. The dual space $T_p^*\mathbb{R}^3$ is drawn above the tangent space which it is dual to. Notice the different ways the axis are labeled

Now we will see exactly how the dual basis, that is, the one-forms, $\{dx_1, dx_2, dx_3\}$ act on the tangent space basis elements $\left\{\frac{\partial}{\partial x_1}, \frac{\partial}{\partial x_2}, \frac{\partial}{\partial x_3}\right\}$. From the definition of dual basis, we have

$$\begin{aligned}
dx_1\left(\frac{\partial}{\partial x_1}\right) &= 1, \quad dx_2\left(\frac{\partial}{\partial x_1}\right) = 0, \quad dx_3\left(\frac{\partial}{\partial x_1}\right) = 0,\\
dx_1\left(\frac{\partial}{\partial x_2}\right) &= 0, \quad dx_2\left(\frac{\partial}{\partial x_2}\right) = 1, \quad dx_3\left(\frac{\partial}{\partial x_2}\right) = 0,\\
dx_1\left(\frac{\partial}{\partial x_3}\right) &= 0, \quad dx_2\left(\frac{\partial}{\partial x_3}\right) = 0, \quad dx_3\left(\frac{\partial}{\partial x_3}\right) = 1.
\end{aligned}$$

Now let us consider a vector, say

$$\begin{bmatrix} -1 \\ 3 \\ -4 \end{bmatrix} = -\frac{\partial}{\partial x_1} + 3\frac{\partial}{\partial x_2} - 4\frac{\partial}{\partial x_3},$$

and see how the differential forms dx_1, dx_2 and dx_3 act on it. First we see how dx_1 acts on the vector:

$$\begin{aligned}
&dx_1\left(-\frac{\partial}{\partial x_1} + 3\frac{\partial}{\partial x_2} - 4\frac{\partial}{\partial x_3}\right)\\
&= dx_1\left(-\frac{\partial}{\partial x_1}\right) + dx_1\left(3\frac{\partial}{\partial x_2}\right) + dx_1\left(-4\frac{\partial}{\partial x_3}\right)\\
&= -dx_1\left(\frac{\partial}{\partial x_1}\right) + 3dx_1\left(\frac{\partial}{\partial x_2}\right) - 4dx_1\left(\frac{\partial}{\partial x_3}\right)\\
&= -1(1) + 3(0) - 4(0)\\
&= -1.
\end{aligned}$$

Similarly, we have

$$\begin{aligned}
&dx_2\left(-\frac{\partial}{\partial x_1}+3\frac{\partial}{\partial x_2}-4\frac{\partial}{\partial x_3}\right)\\
&= dx_2\left(-\frac{\partial}{\partial x_1}\right)+dx_2\left(3\frac{\partial}{\partial x_2}\right)+dx_2\left(-4\frac{\partial}{\partial x_3}\right)\\
&= -dx_2\left(\frac{\partial}{\partial x_1}\right)+3dx_2\left(\frac{\partial}{\partial x_2}\right)-4dx_2\left(\frac{\partial}{\partial x_3}\right)\\
&= -1(0)+3(1)-4(0)\\
&= 3
\end{aligned}$$

and

$$\begin{aligned}
&dx_3\left(-\frac{\partial}{\partial x_1}+3\frac{\partial}{\partial x_2}-4\frac{\partial}{\partial x_3}\right)\\
&= dx_3\left(-\frac{\partial}{\partial x_1}\right)+dx_3\left(3\frac{\partial}{\partial x_2}\right)+dx_3\left(-4\frac{\partial}{\partial x_3}\right)\\
&= -dx_3\left(\frac{\partial}{\partial x_1}\right)+3dx_3\left(\frac{\partial}{\partial x_2}\right)-4dx_3\left(\frac{\partial}{\partial x_3}\right)\\
&= -1(0)+3(0)-4(1)\\
&= -4.
\end{aligned}$$

In essence, the differential one-forms dx_1, dx_2, and dx_3 find the projections of the vector onto the appropriate axis. Reverting to more traditional x, y, z notation, the one-form dx finds the projection of v_p onto the $\frac{\partial}{\partial x}$ axis, the one-form dy finds the projection of v_p onto the $\frac{\partial}{\partial y}$ axis, and the one-form dz finds the projection of v_p onto the $\frac{\partial}{\partial z}$ axis. Pictorially this is shown by Fig. 2.29.

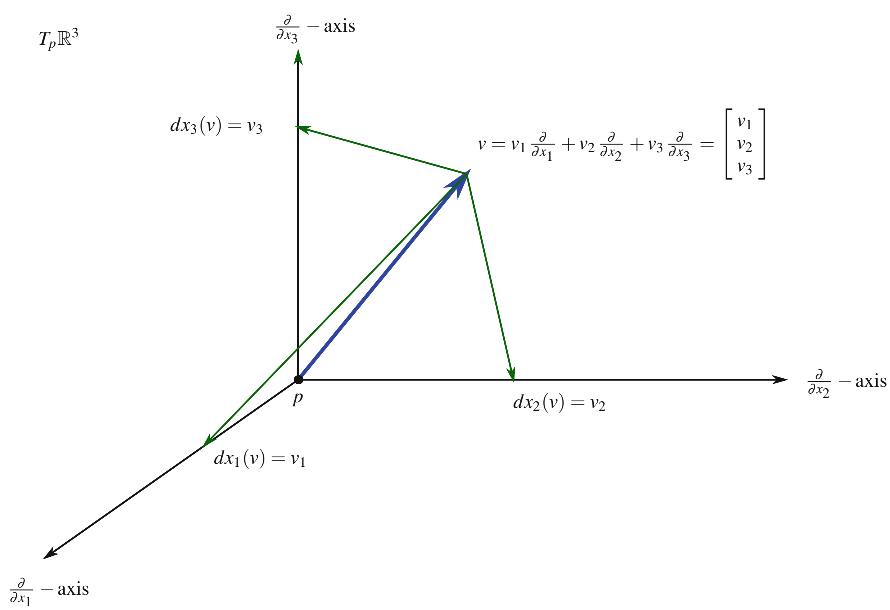


Fig. 2.29 An illustration of how the dual basis elements act on a vector v. The diagram does not include the point p in the notation

Question 2.7 Find the following:

(a) $dx_1\Big(2\frac{\partial}{\partial x_1} + 3\frac{\partial}{\partial x_1}\Big)$
(b) $dx_2\Big(-\frac{\partial}{\partial x_1} + 3\frac{\partial}{\partial x_2} - 4\frac{\partial}{\partial x_3}\Big)$
(c) $dx_3\Big(7\frac{\partial}{\partial x_2} - 5\frac{\partial}{\partial x_3}\Big)$

Since the set $\{dx, dy, dz\}$ is the basis for $T_p^*\mathbb{R}^3$ then any element $\alpha \in T_p^*\mathbb{R}^3$ can be written in the form $adx + bdy + cdz$ where $a, b, c \in \mathbb{R}$. These one-forms behave exactly as one would expect them too. Consider, for example, how the one form $2dx - 3dy + 5dz$ acts on the vector $-\frac{\partial}{\partial x_1} + 3\frac{\partial}{\partial x_2} - 4\frac{\partial}{\partial x_3}$:

$$
\begin{aligned}
&(2dx - 3dy + 5dz)\left(-\frac{\partial}{\partial x} + 3\frac{\partial}{\partial y} - 4\frac{\partial}{\partial z}\right)\\
&= 2dx\left(-\frac{\partial}{\partial x} + 3\frac{\partial}{\partial y} - 4\frac{\partial}{\partial z}\right)\\
&\quad - 3dy\left(-\frac{\partial}{\partial x} + 3\frac{\partial}{\partial y} - 4\frac{\partial}{\partial z}\right)\\
&\quad + 5dz\left(-\frac{\partial}{\partial x} + 3\frac{\partial}{\partial y} - 4\frac{\partial}{\partial z}\right)\\
&= 2dx\left(-\frac{\partial}{\partial x}\right) + 2dx\Big(3\frac{\partial}{\partial y}\Big) + 2dx\Big(-4\frac{\partial}{\partial z}\Big)\\
&\quad - 3dy\left(-\frac{\partial}{\partial x}\right) - 3dy\left(3\frac{\partial}{\partial y}\right) - 3dy\left(-4\frac{\partial}{\partial z}\right)\\
&\quad 5dz\left(-\frac{\partial}{\partial x}\right) + 5dz\left(3\frac{\partial}{\partial y}\right) + 5dz\left(-4\frac{\partial}{\partial z}\right)\\
&= (2)(-1)dx\left(\frac{\partial}{\partial x}\right) + (2)(3)dx\left(\frac{\partial}{\partial y}\right) + (2)(-4)dx\Big(\frac{\partial}{\partial z}\Big)\\
&\quad (-3)(-1)dy\left(\frac{\partial}{\partial x}\right) + (-3)(3)dy\left(\frac{\partial}{\partial y}\right) + (-3)(-4)dy\left(\frac{\partial}{\partial z}\right)\\
&\quad (5)(-1)dz\left(\frac{\partial}{\partial x}\right) + (5)(3)dz\left(\frac{\partial}{\partial y}\right) + (5)(-4)dz\left(\frac{\partial}{\partial z}\right)\\
&= (2)(-1)(1) + (2)(3)(0) + (2)(-4)(0)\\
&\quad (-3)(-1)(0) + (-3)(3)(1) + (-3)(-4)(0)\\
&\quad (5)(-1)(0) + (5)(3)(0) + (5)(-4)(1)\\
&= -31.
\end{aligned}
$$

We have gone to extra care to show what happens at every step of this computation, something we certainly will not always do.

Question 2.8 Find the following:

(a) $\Big(dx_1 + dx_2\Big)\Big(5\frac{\partial}{\partial x_1}\Big)$
(b) $\Big(2dx_2 - 3dx_3\Big)\Big(3\frac{\partial}{\partial x_1} - 7\frac{\partial}{\partial x_2} + 5\frac{\partial}{\partial x_3}\Big)$
(c) $\Big(-dx_1 - 4dx_2 + 6dx_3\Big)\Big(-2\frac{\partial}{\partial x_1} + 5\frac{\partial}{\partial x_2} - 3\frac{\partial}{\partial x_3}\Big)$

In order to make the computations simpler, one-forms are often written as row vectors, exactly as we wrote elements of the dual space as row vectors in our review of vector spaces. For example, the one-form $4dx_1 - 2dx_2 + 5dx_3$ can be written as the row vector $[4, -2, 5]$. Recall, we said the dual space at p is often called the cotangent space at p. Thus, when a one-form, which is an element of the cotangent space, is written as a row vector it is often called a **co-vector**. This allows us to do the above computation as a matrix multiplication. Let us redo the last computation to see how,

$$\begin{aligned}
&(2dx - 3dy + 5dz)\left(-\frac{\partial}{\partial x} + 3\frac{\partial}{\partial y} - 4\frac{\partial}{\partial z}\right) \\
&= [2, -3, 5]\begin{bmatrix} -1 \\ 3 \\ -4 \end{bmatrix} \\
&= (2)(-1) + (-3)(3) + (5)(-4) \\
&= -31.
\end{aligned}$$

This is certainly a more straight forward computation.

Question 2.9 Do the following computations as matrix multiplication by first writing the one-form as a row vector (co-vector) and the vector element as a column vector.

(a) $\Big(-7dx_1 + dx_2\Big)\Big(2\frac{\partial}{\partial x_2}\Big)$
(b) $dx_3\Big(7\frac{\partial}{\partial x_2} - 5\frac{\partial}{\partial x_3}\Big)$
(c) $\Big(2dx_2 - 3dx_3\Big)\Big(3\frac{\partial}{\partial x_1} - 7\frac{\partial}{\partial x_2} + 5\frac{\partial}{\partial x_3}\Big)$
(d) $\Big(-dx_1 - 4dx_2 + 6dx_3\Big)\Big(-2\frac{\partial}{\partial x_1} + 5\frac{\partial}{\partial x_2} - 3\frac{\partial}{\partial x_3}\Big)$

Notice that we now have two different mental images that we can use to try to picture the differential one-form. Consider Fig. 2.30. The top is an image of a differential one-form as a row vector, which is the easiest image to have. Differential one-forms of the form $adx + bdy + cdz$, $a, b, c \in \mathbb{R}$, are elements in the cotangent space at some point, $T_p^*\mathbb{R}^3$, and so imagining them as a row vector (called a co-vector) is natural. The second image is to view the one-forms dx, dy, dz as ways to find the projection of a vector v_p on different axes. In that case the one-form $adx + bdy + cdz$ can be viewed as scaling the respective projections by the factors a, b, and c and then summing these different scalings.

In the last few questions we implicitly assumed that the one-forms and vectors were all at the same point of the manifold. But just as vectors gave us vector fields, we can think of one-form "fields" on the manifold $\mathbb{R}^3$. Actually though, we generally won't call them fields, we simply refer to them as one-forms. Remember that slightly odd terminology, "one-forms *on* the manifold $\mathbb{R}^3$"? A one-form *on* a manifold is actually a **one-form field** that gives a particular one-form at each point p of the manifold, which acts on tangent vectors v_p that are based at that point.

Let us consider an example. Given the following real-valued functions on the manifold $\mathbb{R}^3$, $f(x, y, z) = x^2y$, $g(x, y, z) = \frac{x}{2} + yz$, and $h(x, y, z) = x + y + 3$ define, at each point at each point $p = (x, y, z) \in \mathbb{R}^3$, the one-form ϕ on manifold $\mathbb{R}^3$ by $\phi_{(x,y,z)} = f(x, y, z)dx + g(x, y, z)dy + h(x, y, z)dz$. The one-form ϕ on manifold $\mathbb{R}^3$ defines a different one-form at each point $p \in \mathbb{R}^3$. For example, at the point $(1, 2, 3)$ the one-form $\phi_{(1,2,3)}$ is given by

$$\begin{aligned}
\phi_{(1,2,3)} &= f(1, 2, 3)dx + g(1, 2, 3)dy + h(1, 2, 3)dz \\
&= 2dx + \frac{13}{2}dy + 6dz.
\end{aligned}$$

If we were given the vector

$$v_p = \begin{bmatrix} 2 \\ -1 \\ -2 \end{bmatrix}_p$$

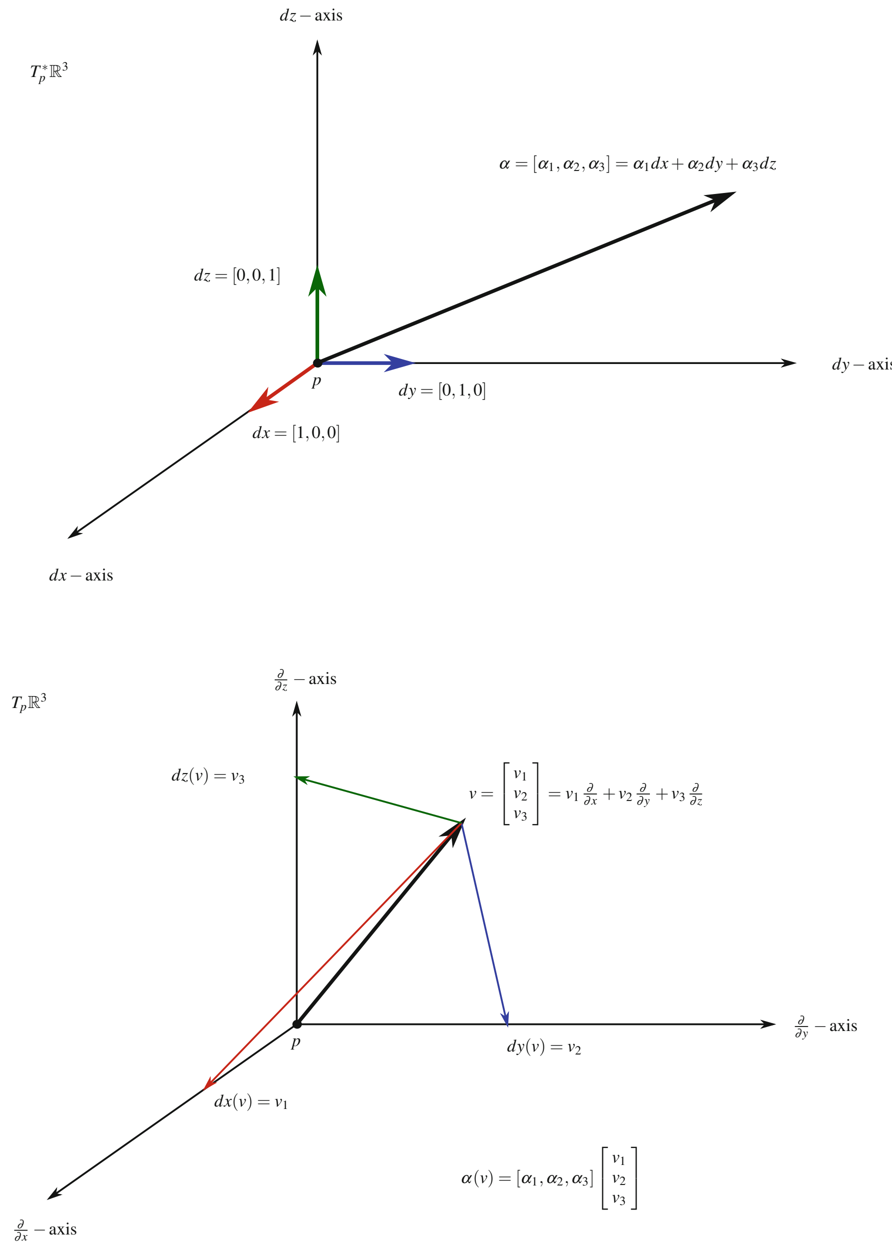


Fig. 2.30 Two different ways to imagine or visualize a differential one-form. As co-vectors in $T_p^*\mathbb{R}^3$ (top) or as a linear combination of the projections onto the axes in $T_p\mathbb{R}^3$ (bottom)

at $p = (1, 2, 3)$ we would have

$$\begin{aligned}
\phi_p(v_p) &= \left(2dx + \frac{13}{2}dy + 6dz\right)\left(2\frac{\partial}{\partial x} - 1\frac{\partial}{\partial y} - 2\frac{\partial}{\partial z}\right) \\
&= 2dx\left(2\frac{\partial}{\partial x} - 1\frac{\partial}{\partial y} - 2\frac{\partial}{\partial z}\right) + \frac{13}{2}dy\left(2\frac{\partial}{\partial x} - 1\frac{\partial}{\partial y} - 2\frac{\partial}{\partial z}\right) + 6dz\left(2\frac{\partial}{\partial x} - 1\frac{\partial}{\partial y} - 2\frac{\partial}{\partial z}\right) \\
&= \left[2, \frac{13}{2}, 6\right]_{(1,2,3)} \begin{bmatrix} 2 \\ -1 \\ -2 \end{bmatrix}_{(1,2,3)}
\end{aligned}$$

$$
\begin{aligned}
&= 2(2) + \frac{13}{2}(-1) + 6(-2) \\
&= \frac{-29}{2}.
\end{aligned}
$$

Thus, given a one-form ϕ on the manifold $\mathbb{R}^3$, at each point $p \in \mathbb{R}^3$ we have a mapping $\phi_p : T_p(\mathbb{R}^3) \to \mathbb{R}$. If we were given a vector field v on manifold $\mathbb{R}^3$, that is, a section on the tangent bundle $T\mathbb{R}^3$, then for each $p \in \mathbb{R}^3$ we have $\phi_p(v_p) \in \mathbb{R}$. So what would $\phi(v)$ be? Notice we have not put in a point. We could consider $\phi(v)$ to be a function on manifold $\mathbb{R}^3$. That is, its inputs are points p on the manifold and its outputs are real numbers, like so

$$
\begin{aligned}
\phi(v) : \mathbb{R}^3 &\longrightarrow \mathbb{R} \\
p &\longmapsto \phi_p(v_p).
\end{aligned}
$$

Question 2.10 For the one-form $\phi_{(x,y,z)} = f(x,y,z)dx + g(x,y,z)dy + h(x,y,z)dz$ where $f(x,y,z) = xy^2$, $g(x,y,z) = \frac{xy}{3} + x^2$, and $h(x,y,z) = xy + yz + xz$, and the vector $v_p = \begin{bmatrix} x \\ x^2y \\ xz \end{bmatrix}_p$ find

(a) (i) $\phi_{(-1,2,1)}$
 (ii) $\phi_{(0,-1,2)}$
 (iii) $\phi_{(8,4,-3)}$
(b) Find $\phi(v)$
(c) (i) Find $\phi_p(v_p)$ for $p = (-1,2,1)$
 (ii) Find $\phi_p(v_p)$ for $p = (0,-1,2)$
 (iii) Find $\phi_p(v_p)$ for $p = (8,4,-3)$

Now that we are a little more comfortable with what one-forms actually do, we are ready to try to understand the notation better. In essence, we want to understand our reason for denoting dx, dy, dz as the basis of $T_p^*(\mathbb{R}^3)$ dual to the basis $\frac{\partial}{\partial x}, \frac{\partial}{\partial y}, \frac{\partial}{\partial z}$ of $T_p(\mathbb{R}^3)$. First we make another definition that, at first glance, may seem a little circular to you.

Definition 2.4.2 Let $f : \mathbb{R}^n \to \mathbb{R}$ be a function on the manifold $\mathbb{R}^n$. The **differential** df of f is defined to be the one-form on $\mathbb{R}^n$ such that for all vectors v_p we have

$$
df(v_p) = v_p[f].
$$

Let's just take a moment to try to unpack and digest all this considering the manifold $\mathbb{R}^3$. First of all, given a function $f : \mathbb{R}^3 \to \mathbb{R}$ on manifold $\mathbb{R}^3$ and a vector $v_p \in T_p(\mathbb{R})$ at point p in manifold $\mathbb{R}^3$ we know how to take the directional derivative of this function at the point p the direction v_p

$$
\lim_{t \to 0} \frac{f\big(p + tv_p\big) - f\big(p\big)}{t},
$$

which can also be written as

$$
\frac{d}{dt}\Big(f\big(p + tv_p\big)\Big)\Big|_{t=0}.
$$

We then thought of this as using the vector v_p to perform an operation on the given function f, and so introduced a slightly new notation for this

$$
v_p[f] = \frac{d}{dt}\Big(f\big(p + tv_p\big)\Big)\Big|_{t=0},
$$

which in turn we have just used to define the differential of f, written as df, by $df(v_p) = v_p[f]$. Letting $p = (x_0, y_0, z_0)$ we put this all together to get

$$
\begin{aligned}
df(v_p) &= v_p[f] \\
&= \frac{d}{dt}\Big(f\big(p + tv_p\big)\Big)\bigg|_{t=0} \\
&= \frac{d}{dt} f\big(\underbrace{x_0 + tv_1}_{x(t)}, \underbrace{y_0 + tv_2}_{y(t)}, \underbrace{z_0 + tv_3}_{z(t)}\big)\Big|_{t=0} \\
&= \frac{\partial f}{\partial x}\bigg|_p \cdot \frac{dx(t)}{dt}\bigg|_{t=0} + \frac{\partial f}{\partial y}\bigg|_p \cdot \frac{dy(t)}{dt}\bigg|_{t=0} + \frac{\partial f}{\partial z}\bigg|_p \cdot \frac{dz(t)}{dt}\bigg|_{t=0} \\
&= \frac{\partial f}{\partial x}\bigg|_p \cdot v_1 + \frac{\partial f}{\partial y}\bigg|_p \cdot v_2 + \frac{\partial f}{\partial z}\bigg|_p \cdot v_3.
\end{aligned}
$$

In summary we have

$$
df(v_p) = \frac{\partial f}{\partial x}\bigg|_p \cdot v_1 + \frac{\partial f}{\partial y}\bigg|_p \cdot v_2 + \frac{\partial f}{\partial z}\bigg|_p \cdot v_3.
$$

Since this formula works for any function $f : \mathbb{R}^3 \to \mathbb{R}$ it works for the special Cartesian coordinate functions $x, y, z : \mathbb{R}^3 \to \mathbb{R}$. Recall the Cartesian coordinate functions; if $p \in \mathbb{R}^3$, then $x(p)$ is the Cartesian x-coordinate value of p, $y(p)$ is the Cartesian y-coordinate value of p, and $z(p)$ is the Cartesian z-coordinate value of p. For example, if we could write the point p as $(3, 2, 1)$ in Cartesian coordinates, then $x(p) = 3$, $y(p) = 2$, and $z(p) = 1$. Now let us see exactly what happens to each of these Cartesian coordinate functions. Suppose that

$$
v_p = \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix}_p = v_1 \frac{\partial}{\partial x}\bigg|_p + v_2 \frac{\partial}{\partial y}\bigg|_p + v_3 \frac{\partial}{\partial z}\bigg|_p .
$$

Then for the Cartesian coordinate function $x : \mathbb{R}^3 \to \mathbb{R}$ we have that

$$
\begin{aligned}
dx(v_p) &= \frac{\partial x}{\partial x}\bigg|_p \cdot v_1 + \frac{\partial x}{\partial y}\bigg|_p \cdot v_2 + \frac{\partial x}{\partial z}\bigg|_p \cdot v_3 \\
&= (1) \cdot v_1 + (0) \cdot v_2 + (0) \cdot v_3 \\
&= v_1.
\end{aligned}
$$

But what do we mean when we write $\frac{\partial x}{\partial x}$? The x in the numerator is the Cartesian coordinate function while the x in the denominator represents the variable we are differentiating with respect to. Thus $\frac{\partial x}{\partial x}$ gives the rate of change of the Cartesian coordinate function x is in the x-direction.

Question 2.11 Explain that the rate of change of the Cartesian coordinate function x in the x-direction is indeed one, thereby showing that $\frac{\partial x}{\partial x} = 1$. It may be helpful to consider how $x(p)$ changes as the point p moves in the x-direction. What is the rate of change of the Cartesian coordinate function x in the y and z-directions? Explain how your answer shows that $\frac{\partial x}{\partial y} = 0$ and $\frac{\partial x}{\partial z} = 0$.

Similarly, we have

$$
\begin{aligned}
dy(v_p) &= \frac{\partial y}{\partial x}\bigg|_p \cdot v_1 + \frac{\partial y}{\partial y}\bigg|_p \cdot v_2 + \frac{\partial y}{\partial z}\bigg|_p \cdot v_3 \\
&= (0) \cdot v_1 + (1) \cdot v_2 + (0) \cdot v_3 \\
&= v_2
\end{aligned}
$$

and

$$
\begin{aligned}
dz(v_p) &= \left.\frac{\partial z}{\partial x}\right|_p \cdot v_1 + \left.\frac{\partial z}{\partial y}\right|_p \cdot v_2 + \left.\frac{\partial z}{\partial z}\right|_p \cdot v_3 \\
&= (0) \cdot v_1 + (0) \cdot v_2 + (1) \cdot v_3 \\
&= v_3.
\end{aligned}
$$

Question 2.12 Find the rate of change of the Cartesian coordinate function y in the x, y, and z-directions and use this to show that $\frac{\partial y}{\partial x} = 0$, $\frac{\partial y}{\partial y} = 1$, and $\frac{\partial y}{\partial z} = 0$. Then find the rate of change of the Cartesian coordinate function z in the x, y, and z-directions to show that $\frac{\partial z}{\partial x} = 0$, $\frac{\partial z}{\partial y} = 0$, and $\frac{\partial z}{\partial z} = 1$.

Question 2.13 Writing $e_1 = \frac{\partial}{\partial x}$, $e_2 = \frac{\partial}{\partial y}$ and $e_3 = \frac{\partial}{\partial z}$ show that the differentials of the Cartesian coordinate functions, dx, dy, and dz give

$$
\begin{aligned}
&dx\left(\frac{\partial}{\partial x}\right) = 1, \quad dy\left(\frac{\partial}{\partial x}\right) = 0, \quad dz\left(\frac{\partial}{\partial x}\right) = 0, \\
&dx\left(\frac{\partial}{\partial y}\right) = 0, \quad dy\left(\frac{\partial}{\partial y}\right) = 1, \quad dz\left(\frac{\partial}{\partial y}\right) = 0, \\
&dx\left(\frac{\partial}{\partial z}\right) = 0, \quad dy\left(\frac{\partial}{\partial z}\right) = 0, \quad dz\left(\frac{\partial}{\partial z}\right) = 1.
\end{aligned}
$$

Now we can compare the behavior of the differentials dx, dy, and dz of the Cartesian coordinate functions x, y, and z with the behavior of the dual basis elements, that is, the duals of the basis elements $\frac{\partial}{\partial x}, \frac{\partial}{\partial y}, \frac{\partial}{\partial z}$ of $T_p(\mathbb{R}^3)$. We see that they behave exactly the same. This explains the choice of notation for the dual basis elements at the beginning of the section. ***The dual basis elements are exactly the differentials of the Cartesian coordinate functions.***

Given a specific function f we would like to know how to actually write the differential df. Suppose $f : \mathbb{R}^3 \to \mathbb{R}$ is a function on manifold $\mathbb{R}^3$ and v_p is an arbitrary vector at an arbitrary point. Since

$$
df(v_p) = \left.\frac{\partial f}{\partial x}\right|_p \cdot v_1 + \left.\frac{\partial f}{\partial y}\right|_p \cdot v_2 + \left.\frac{\partial f}{\partial z}\right|_p \cdot v_3
$$

and

$$
v_1 = dx(v_p), \;\; v_2 = dy(v_p), \;\; v_3 = dz(v_p).
$$

We put this together to get

$$
\begin{aligned}
df(v_p) &= \left.\frac{\partial f}{\partial x}\right|_p \cdot v_1 + \left.\frac{\partial f}{\partial y}\right|_p \cdot v_2 + \left.\frac{\partial f}{\partial z}\right|_p \cdot v_3 \\
&= \left.\frac{\partial f}{\partial x}\right|_p \cdot dx(v_p) + \left.\frac{\partial f}{\partial y}\right|_p \cdot dy(v_p) + \left.\frac{\partial f}{\partial z}\right|_p \cdot dz(v_p) \\
&= \left(\left.\frac{\partial f}{\partial x}\right|_p \cdot dx + \left.\frac{\partial f}{\partial y}\right|_p \cdot dy + \left.\frac{\partial f}{\partial z}\right|_p \cdot dz \right)(v_p)
\end{aligned}
$$

and so we have

$$
df = \left.\frac{\partial f}{\partial x}\right|_p dx + \left.\frac{\partial f}{\partial y}\right|_p dy + \left.\frac{\partial f}{\partial z}\right|_p dz.
$$

We have written the differential of the function f as a linear combination of the dual basis elements, thereby showing that $df \in T_p^*(\mathbb{R}^3)$. The differential of a function f is a one-form.

The final thing we want to do is to consider another way to think of the differential df of f. In order to draw the necessary pictures we will consider the manifold $\mathbb{R}^2$ since we can picture the graph of a function $f : \mathbb{R}^2 \to \mathbb{R}$ in three dimensions, while picturing the graph of a function $f : \mathbb{R}^3 \to \mathbb{R}$ properly would require four dimensions, but the underlying idea for all functions $f : \mathbb{R}^n \to \mathbb{R}$ is the same. Letting $v = \begin{bmatrix} v_1 \\ v_2 \end{bmatrix}$ we have

$$\begin{aligned} v[f] &= \left(v_1 \frac{\partial}{\partial x} + v_2 \frac{\partial}{\partial y} \right) [f] \\ &= \frac{\partial f}{\partial x} v_1 + \frac{\partial f}{\partial y} v_2 \\ &= \frac{\partial f}{\partial x} dx(v) + \frac{\partial f}{\partial y} dy(v) \\ &= \left(\frac{\partial f}{\partial x} dx + \frac{\partial f}{\partial y} dy \right) (v) \\ &= df(v). \end{aligned}$$

Question 2.14 Repeat the above calculation putting in the base point. The differential of f at p can be written either as df_p or as $df(p)$.

But now for the real question, what exactly is this differential of f, df? The key lies in Fig. 2.31. The differential df of f takes in a vector v_p, which is at some point p of the manifold $\mathbb{R}^2$, and gives out a number that is the "rise" of the tangent plane to the graph of the function f as one moves from p along v_p. This output can be viewed as also being a point on the tangent plane. Thus df in a sense "encodes" the "rises" that occur in the tangent plane to the function f as one moves along different vectors with base point p. This should not be too surprising. The formula for the tangent plane T to $f(x, y)$ at the point (x_0, y_0) is given by

$$T(x, y) = f(x_0, y_0) + \left. \frac{\partial f}{\partial x} \right|_{(x_0, y_0)} \cdot (x - x_0) + \left. \frac{\partial f}{\partial y} \right|_{(x_0, y_0)} \cdot (y - y_0).$$

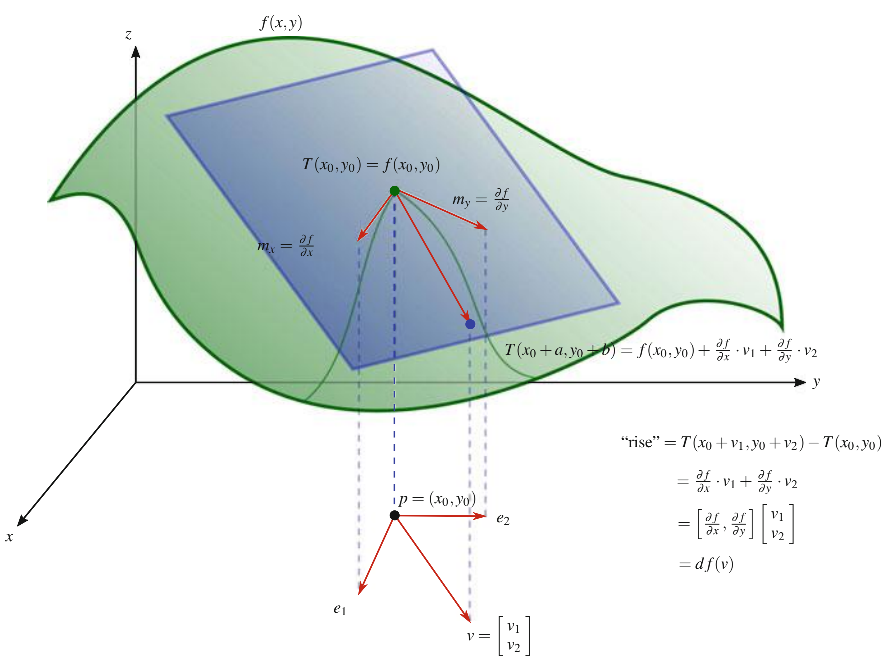


Fig. 2.31 The differential df_p is the linear approximation of the function f at the point p. In other words, the differential df_p "encodes" the tangent plane of f at p

As the picture indicates, given the vector $v = \begin{bmatrix} v_1 \\ v_2 \end{bmatrix}_p$, this vector's endpoint (in $\mathbb{R}^2$) is $(x_0 + v_1, y_0 + v_2)$. The point in the tangent plane to f (at the point $p = (x_0, y_0)$) that lies above $(x_0 + v_1, y_0 + v_2)$ is given by

$$\begin{aligned} T(x_0 + v_1, y_0 + v_2) &= f(x_0, y_0) + \left.\frac{\partial f}{\partial x}\right|_{(x_0,y_0)} \cdot (x_0 + v_1 - x_0) + \left.\frac{\partial f}{\partial y}\right|_{(x_0,y_0)} \cdot (y_0 + v_2 - y_0) \\ &= f(x_0, y_0) + \left.\frac{\partial f}{\partial x}\right|_{(x_0,y_0)} \cdot v_1 + \left.\frac{\partial f}{\partial y}\right|_{(x_0,y_0)} \cdot v_2 \end{aligned}$$

so the "rise" in the tangent plane from $f(x_0, y_0)$ is given by

$$\begin{aligned} \text{"rise"} &= T(x_0 + v_1, y_0 + v_2) - T(x_0, y_0) \\ &= f(x_0, y_0) + \left.\frac{\partial f}{\partial x}\right|_{(x_0,y_0)} \cdot v_1 + \left.\frac{\partial f}{\partial y}\right|_{(x_0,y_0)} \cdot v_2 - f(x_0, y_0) \\ &= \left.\frac{\partial f}{\partial x}\right|_{(x_0,y_0)} \cdot v_1 + \left.\frac{\partial f}{\partial y}\right|_{(x_0,y_0)} \cdot v_2 \\ &= df_p(v_p). \end{aligned}$$

Thus the differential of f at p, df_p, which gives the "rise" of the tangent plane to f at p as you move along the vector v_p, essentially encodes how this tangent plane behaves. The tangent plane is the closest linear approximation of f at p, so in essence *df_p can be thought of as the linear approximation of the function f at the point p.*

As one last comment, does the differential one-form df written as a row vector

$$\begin{aligned} df &= \frac{\partial f}{\partial x} dx + \frac{\partial f}{\partial y} dy \\ &= \left[\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y} \right] \end{aligned}$$

in the $\mathbb{R}^2$ case or

$$\begin{aligned} df &= \frac{\partial f}{\partial x} dx + \frac{\partial f}{\partial y} dy + \frac{\partial f}{\partial z} dz \\ &= \left[\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}, \frac{\partial f}{\partial z} \right] \end{aligned}$$

in the $\mathbb{R}^3$ case remind you of anything from vector calculus? It should remind you of $grad(f)$ or $\nabla(f)$. In fact, df and $grad(f)$ are closely related. We will explore this relationship in Chap. 9.

We now consider a couple of examples. Let $f : \mathbb{R}^3 \to \mathbb{R}$ be given by $f(x, y, z) = x^2 y^3 z$ and consider the vector $v_p = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}_p$ at each point $p = (x, y, z) \in \mathbb{R}$. First let us find $df(v_p)$. Since v_p is the same at every point p we can drop the p from the notation. We have

$$\begin{aligned} df(v) &= v[f] \\ &= \left(\frac{\partial}{\partial x} + 2\frac{\partial}{\partial y} + 3\frac{\partial}{\partial z} \right) (x^2 y^3 z) \\ &= \frac{\partial (x^2 y^3 z)}{\partial x} + 2\frac{\partial (x^2 y^3 z)}{\partial y} + 3\frac{\partial (x^2 y^3 z)}{\partial z} \\ &= 2xy^3 z + 2(3x^2 y^2 z) + 3(x^2 y^3). \end{aligned}$$

Now suppose we were given the specific point $p = (-1, 2, -2)$. Finding $df(v_p)$ at $p = (-1, 2, -2)$ simply requires us to substitute the numbers in

$$df(v_p) = 2(-1)^2(2)^3(-2) + 2(3(-1)^2(2)^2(-2)) + 3(-1)^2(2)^3 = 8.$$

Question 2.15 Suppose $f(x, y, z) = (x^2 - 1)y + (y^2 + 2)z$.

(i) Find df.

(ii) Let $v_p = \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix}_{(p_1, p_2, p_3)}$. Find $df(v_p)$.

(iii) Let $v_p = \begin{bmatrix} -2 \\ 1 \\ 3 \end{bmatrix}_{(-1,0,2)}$. Find $df(v_p)$.

Question 2.16 Let f and g be functions on $\mathbb{R}^3$. Show that $d(fg) = g\,df + f\,dg$.

Question 2.17 Suppose $f : \mathbb{R}^3 \to \mathbb{R}$, $h : \mathbb{R} \to \mathbb{R}$, and $h(f) : \mathbb{R}^3 \to \mathbb{R}$ are all defined. Show that $d(h(f)) = h'(f)df$.

2.5 Summary, References, and Problems

2.5.1 Summary

Because the distinction is important for more abstract spaces we first made a big deal between the collection of points in $\mathbb{R}^n$ as being the manifold $\mathbb{R}^n$ and the collection of vectors in $\mathbb{R}^n$ as being the vector space $\mathbb{R}^n$. Then we defined the Cartesian functions $x_i : \mathbb{R}^n \to \mathbb{R}$ that take a point p in the manifold $\mathbb{R}^n$ to its ith-coordinate value.

The tangent space of an n-dimensional manifold M at a point $p \in M$ was introduced as the space of all tangent vectors to M at the point p and was denoted as T_pM. The tangent space T_pM could essentially be viewed as a copy of the vector space $\mathbb{R}^n$ attached to M at the point p. The collection of all tangent spaces for a manifold was called the tangent bundle and was denoted by TM. A vector field was defined as a function v that assigns to each point p of M an element v_p of T_pM. A vector field is sometimes called a section of the tangent bundle.

The directional derivative of a function $f : M \to \mathbb{R}$ at the point p in the direction v_p was defined as

$$v_p[f] \equiv \frac{d}{dt}\Big(f\big(p + tv_p\big)\Big)\bigg|_{t=0},$$

which is exactly the same definition from vector calculus only without the requirement that v_p be a unit vector. Dropping this requirement slightly changes the geometrical meaning of the directional derivative but allows us to retain the simple formula

$$v_p[f] = \sum_{i=1}^{n} v_i \cdot \frac{\partial f}{\partial x_i}\bigg|_{p_i} \qquad \text{where} \qquad v_p = \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix}.$$

This formula is then used to make one of the must useful and unexpected identifications,

$$\boxed{\text{The Euclidian vectors } e_i \text{ can be identified with the partial differential operators } \frac{\partial}{\partial x_i}.}$$

In other words, we think of the Euclidian vector e_i as being the partial differential operator $\frac{\partial}{\partial x_i}$.

This directional derivative is then used to define a particular linear functional associated with the function f,

$$df(v_p) = v_p[f].$$

This linear functional is called the differential of f. Since $v_p \in T_pM$, which is a vector space, the linear functional df is in the dual space of T_pM, which is called the cotangent space. We write $df \in T_p^*M$. Furthermore, one can use the Cartesian coordinate functions x_i to define the differentials dx_i, which turn out to be the basis of T_p^*M dual to the Euclidian basis of T_pM. Thus, using the identification above, we have $\{\frac{\partial}{\partial x_1}, \dots, \frac{\partial}{\partial x_n}\}$ as the standard basis for T_pM and $\{dx_1, \dots, dx_n\}$ as the standard dual basis of T_p^*M. Elements of T_p^*M are also called differential one-forms. The differential one-forms dx_i act similar to the Cartesian coordinate functions and pick off the ith-component of the vector v_p, that is, $dx_i(v_p) = v_i$. Geometrically we can think of the dx_i as finding the length of the vector which is the projection of v_p onto the ith axis.

2.5.2 *References and Further Reading*

Vectors and vector fields show up all over multivariable calculus, see for example Stewart [43] and even Marsden and Hoffmen [31], but tangent spaces and tangent bundles are generally not encountered or made explicit at this level. That is usually left for more advanced courses, often in differential geometry. There are quite a number of interesting and rigorous introductions to this material. Naming only a few, Munkres [35], O'Neill [36], Tu [46], Spivak [41], and Thorpe [45] are all good introductions to analysis on manifolds and differential geometry and thus contain good expositions on the tangent space and tangent bundles. Again, Stewart [43] and Marsden and Hoffmen [31] are excellent references to directional derivatives and multivariable calculus. This chapter also begins a gentle introduction to differential forms, that will take several chapters to complete, by looking at differential one-forms. The material on differential one-forms in this chapter has generally followed the spirit of Bachman [4] and O'Neill [36], though Munkres [35], Edwards [18], and Martin [33] were also consulted.

2.5.3 *Problems*

Question 2.18 Let $p_1 = (2, 9, -2)$, $p_2 = (1, 0, -3)$, and $p_3 = (12, -7, 4)$ be points in $\mathbb{R}^3$ and let x,y, and z be the Cartesian coordinate functions on $\mathbb{R}^3$. Find

$a)\ x(p_1)$,	$d)\ x(p_2)$,	$g)\ x(p_3)$,
$b)\ y(p_1)$,	$e)\ y(p_2)$,	$h)\ y(p_3)$,
$c)\ z(p_1)$,	$f)\ z(p_2)$,	$i)\ z(p_3)$.

Question 2.19 Consider the manifold $\mathbb{R}^3$ and the vector field $v = xye_1 + \left(z - \frac{x}{3}\right)e_2 - (x + z - 4)e_3$, where e_1, e_2, e_3 are the Euclidian vectors. If $p_1 = (1, -3, 2)$, $p_2 = (0, 1, 4)$, and $p_3 = (-5, -2, 2)$ find the vectors v_{p_1}, v_{p_2}, and v_{p_3}. What space is each of these vectors in?

Question 2.20 Find the directional derivative of $f(x, y) = 5x^2y - 4xy^3$ in the directions $v_1 = \begin{bmatrix} 5 \\ 12 \end{bmatrix}$ and $v_2 = \frac{1}{13}\begin{bmatrix} 5 \\ 12 \end{bmatrix}$ at the point $(1, 2)$. Explain what the difference between $D_{v_1}f$ and $D_{v_2}f$ is.

Question 2.21 Find the directional derivative of $f(x, y) = x\ln(y)$ in the directions $v_1 = \begin{bmatrix} -4 \\ 3 \end{bmatrix}$ and $v_2 = \frac{1}{5}\begin{bmatrix} -4 \\ 3 \end{bmatrix}$ at the point $(1, -3)$. Explain what the difference between $D_{v_1}f$ and $D_{v_2}f$ is.

Question 2.22 Find the directional derivative of $f(x, y, z) = xe^{yz}$ in the directions $v_1 = \begin{bmatrix} 2 \\ -2 \\ 1 \end{bmatrix}$ and $v_2 = \frac{1}{3}\begin{bmatrix} 2 \\ -2 \\ 1 \end{bmatrix}$ at the point $(3, 0, 2)$. Explain what the difference between $D_{v_1}f$ and $D_{v_2}f$ is.

Question 2.23 Given $f(x, y, z) = \sqrt{x + yz}$ and $v = 2\frac{\partial}{\partial x} + 3\frac{\partial}{\partial y} + 6\frac{\partial}{\partial z}$ find $v[f]$ at the point $(1, 3, 1)$.

Question 2.24 Given $f(x, y, z) = xe^y + ye^z + ze^x$ and $v = 5\frac{\partial}{\partial x} + 1\frac{\partial}{\partial y} - 2\frac{\partial}{\partial z}$ find $v[f]$ at the point $(0, 0, 0)$.

Question 2.25 Given $f(x, y, z) = \sqrt{xyz}$ and $v = -1\frac{\partial}{\partial x} - 2\frac{\partial}{\partial y} + 2\frac{\partial}{\partial z}$ find $v[f]$ at the point $(3, 2, 6)$.

Question 2.26 Given $f(x, y, z) = x + y^2 + z^3$ and $v = -3\frac{\partial}{\partial x} + 4\frac{\partial}{\partial y} - 2\frac{\partial}{\partial z}$ find $v[f]$ at the point $(2, 5, -1)$.

Question 2.27 Given the one-forms $\alpha = (x^3 + 2yz^2)dx + (xy - 2)dy + x^4 dz$ on the manifold $\mathbb{R}^3$ find the one-forms α_{p_i} for $p_1 = (-1, 3, 2)$, $p_2 = (2, -2, 1)$, and $p_3 = (3, -5, -2)$.

Question 2.28 Given the one-forms $\alpha = (x^3 - 2yz^2)dx + dy + x^4 dz$ on the manifold $\mathbb{R}^3$ find the one-forms α_{p_i} for $p_1 = (-1, 3, 2)$, $p_2 = (2, -2, 1)$, and $p_3 = (3, -5, -2)$.

Question 2.29 Given the one-forms $\alpha = (x^3 + 2yz^2)dx + (xy - 2)dy + x^4 dz$ and $\beta = (x^3 - 2yz^2)dx + dy + x^4 dz$ on $\mathbb{R}^3$ find the one-forms $\alpha + \beta$ and $\alpha - \beta$. Then find $(\alpha + \beta)_{p_i}$ and $(\alpha - \beta)_{p_i}$ for the points p_i in the last question.

Question 2.30 Given the one-form $\alpha = x^3 dx - xyz dy + y^2 dz$ find the one-forms 2α and $\frac{1}{3}\alpha$.

Question 2.31 Given the one-form $\omega = \frac{xy}{2}dx - \sqrt{yz}dy + (x + z)^{\frac{2}{3}}dz$ and $p = (2, -1, 3)$ find $\alpha_p(v_p)$ for the following vectors:

$a)\ \begin{bmatrix} 2 \\ -2 \\ 1 \end{bmatrix}$ $\quad d)\ 4e_1 + 7e_2 - 3e_3$ $\quad g)\ 2\frac{\partial}{\partial x} - 3\frac{\partial}{\partial y} + 4\frac{\partial}{\partial z}$

$b)\ \begin{bmatrix} 1 \\ 3 \\ -5 \end{bmatrix}$ $\quad e)\ \frac{1}{2}e_1 + \frac{7}{4}e_2 - \frac{3}{5}e_3$ $\quad h)\ -2\frac{\partial}{\partial x} + \frac{5}{3}\frac{\partial}{\partial y} + 3\frac{\partial}{\partial z}$

$c)\ \begin{bmatrix} x + y \\ z - 2 \\ xz \end{bmatrix}$ $\quad f)\ -x^y\, e_1 + \sqrt{xz}\, e_2 - 3y\, e_3$ $\quad i)\ \sqrt{x + y + z}\frac{\partial}{\partial x} - x\ln(y)\frac{\partial}{\partial y} + z^y\frac{\partial}{\partial z}$

Question 2.32 Find df, the differential of f, for the following functions:

$a)\ f(x, y) = x^2y^3$ $\quad d)\ f(x, y, z) = x^2 + y^3 + z^4$ $\quad g)\ f(x, y, z) = \sqrt{x + yz}$

$b)\ f(x, y) = \sqrt{x^2 + y^2}$ $\quad e)\ f(x, y, z) = \sqrt{xyz}$ $\quad h)\ f(x, y, z) = x\ln(y) + y\ln(z) + z\ln(x)$

$c)\ f(x, y) = x + y^3$ $\quad f)\ f(x, y, z) = xe^y + ye^z + ze^x$ $\quad i)\ f(x, y, z) = \left(x^3 + y^3 + z^3\right)^{\frac{2}{3}}$

Question 2.33 Find $df[v_p]$ for the differentials you found in the previous problem and the vectors in the problem before that at the point $p = (2, -1, 3)$.

Chapter 3
The Wedgeproduct

In this chapter we introduce a way to "multiply" one-forms which is called the wedgeproduct. By wedgeproducting two one-forms together we get two-forms, by wedgeproducting three one-forms together we get three-forms, and so on. The geometrical meaning of the wedgeproduct and how it is computed is explained in section one. In section two general two-, three-, and k-forms are introduced and the geometry behind them is also explored.

The algebraic properties and several different formulas for the wedgeproduct are explored in depth in section three. Different books introduce the wedgeproduct in different ways. Often some of these formulas are given as definitions of the wedgeproduct. Doing this, however, obscures the geometric meaning behind the wedgeproduct, which is why we have taken a different approach in this book.

The fourth section is rather short and simply introduces something called the interior product and proves a couple of identities relating the interior product and the wedgeproduct. We will need these identities later on, but as they rely on the wedgeproduct this was the appropriate time to introduce and prove them.

3.1 Area and Volume with the Wedgeproduct

We begin this chapter by emphasizing an important point. Consider a manifold $\mathbb{R}^n$. At each point p of the manifold $\mathbb{R}^n$ there is a tangent space $T_p\mathbb{R}^n$ that is a vector space which is isomorphic to the vector space $\mathbb{R}^n$. Two vector spaces V and W are called **isomorphic**, denoted by $V \simeq W$, if there is a one-to-one and onto mapping $\phi : V \to W$ such that for $v_1, v_2, v \in V$ and $c \in \mathbb{R}$ we have

1. $\phi(v_1 + v_2) = \phi(v_1) + \phi(v_2)$,
2. $\phi(cv) = c\phi(v)$.

This is not a course in abstract algebra so we do not want to make too much of isomorphisms other than to emphasize that if two vector spaces are isomorphic we can think of them as essentially two copies of the same vector space. In the case of $\mathbb{R}^n$ the underlying manifold is also a vector space, which is isomorphic to the tangent space at each point.

Making the distinction between the underlying manifold and the vector spaces is extremely important when we eventually come to more general manifolds instead of manifolds given by the vector spaces $\mathbb{R}^n$. This will also eventually help us place all of vector calculus into a broader context. Vector calculus is a powerful subject, but it is implicitly built on $\mathbb{R}^3$ and for the manifold $\mathbb{R}^3$ we have

$$\text{manifold } \mathbb{R}^3 \simeq T_p\mathbb{R}^3 \simeq T_p^*\mathbb{R}^3 \simeq \text{ vector space } \mathbb{R}^3.$$

These isomorphisms, which were never made explicit in vector calculus, are what allowed you to think of vectors as being inside the manifold $\mathbb{R}^3$, thereby allowing you to take directional derivatives. Because of these isomorphisms all of these spaces are rather sloppily and imprecisely lumped together and combined. Moving to more manifolds requires us to tease out and understand all of the differences, which is a major goal of this book.

J. P. Fortney, *A Visual Introduction to Differential Forms and Calculus on Manifolds*,
https://doi.org/10.1007/978-3-319-96992-3_3

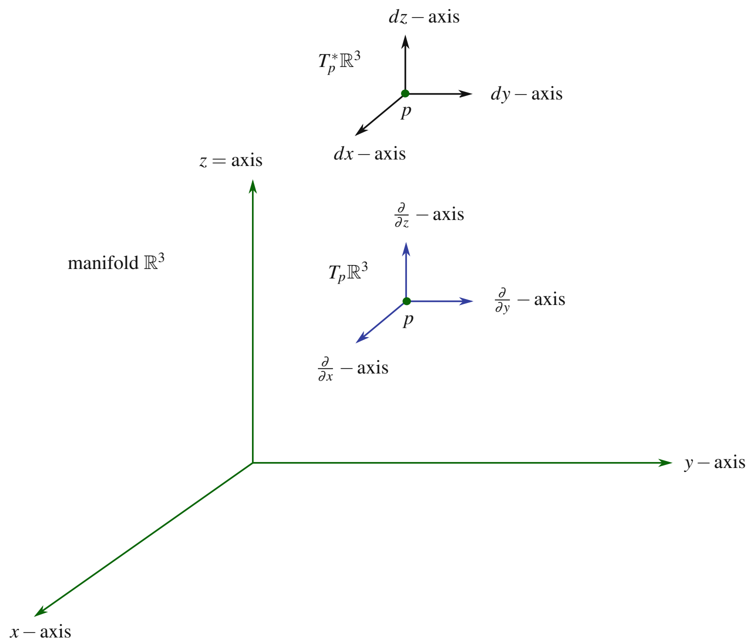


Fig. 3.1 The manifold $\mathbb{R}^3$, the tangent space $T_p\mathbb{R}^3$, and the cotangent space $T_p^*\mathbb{R}^3$, all shown together. Even though the cotangent space $T_p^*\mathbb{R}^3$ is actually attached to the manifold at the same point p that the tangent space $T_p\mathbb{R}^3$ is attached, it is shown above the tangent space. We will generally follow this convention in this book

In the last chapter we also discovered that the tangent space $T_p\mathbb{R}^n$ at each point p has a basis given by

$$\left\{ \left.\frac{\partial}{\partial x_1}\right|_p, \left.\frac{\partial}{\partial x_2}\right|_p, \dots, \left.\frac{\partial}{\partial x_n}\right|_p \right\}$$

and at each point p there is a cotangent space $T_p^*\mathbb{R}^n$, which is the dual space to the vector space $T_p\mathbb{R}^n$, which has a basis given by

$$\left\{ dx_1(p), dx_2(p), \dots, dx_n(p) \right\}$$

or simply $\{dx_1, dx_2, \dots, dx_n\}$ if we suppress the base point p in the notation. The picture we have built up so far is shown in Fig. 3.1 where we have drawn manifold $\mathbb{R}^3$ with the tangent space $T_p\mathbb{R}^3$ superimposed. The cotangent space $T_p^*\mathbb{R}^3$ is pictured above the tangent space.

The one-forms $dx_1(p), dx_2(p), \dots, dx_n(p)$ are exactly the linear functionals on the vector space $T_p\mathbb{R}^n$. And in line with the above, we have $T_p\mathbb{R}^n \simeq \mathbb{R}^n$, with elements $v_p \in T_p\mathbb{R}^n$ usually written as column vectors, and $T_p^*\mathbb{R}^n \simeq \mathbb{R}^n$, with elements $\alpha_p \in T_p^*\mathbb{R}^n$ often written as row vectors and called co-vectors. Thus we can use matrix multiplication (a row vector multiplied by a column vector) for the one-form α_p acting on the vector v_p, $\alpha_p(v_p)$. Right now we will introduce one additional bit of notation. This is also sometimes written as

$$\begin{aligned} \alpha_p(v_p) &= \langle \alpha_p, v_p \rangle \\ &= \langle \alpha, v \rangle_p. \end{aligned}$$

Thus, the angle brackets $\langle \cdot, \cdot \rangle$ denotes the **canonical pairing** between differential forms and vectors. Sometimes authors do not even worry about the order in the canonical pairing and you will even see $\langle v_p, \alpha_p \rangle$ on occasion. Some calculus textbooks use the angle brackets to denote row vectors; we will not use that notation here.

A differential one-form at the point p is simply an element of $T_p^*\mathbb{R}^n$. Our goal is to figure out how to multiply one-forms to give two-, three-, and k-forms. We want to be able to multiply our differential one-forms in such a way that certain volume related properties will be preserved. (This will become clearer in a moment.)

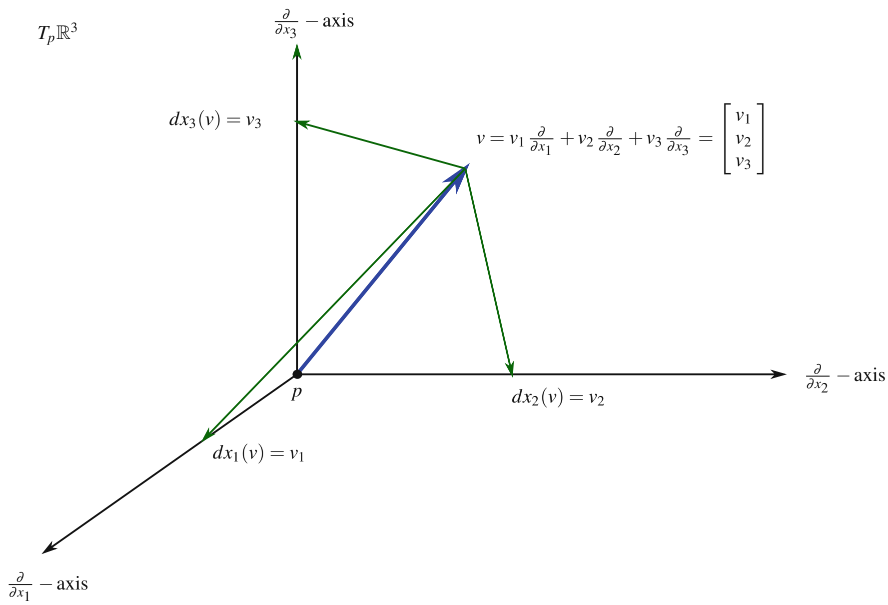


Fig. 3.2 An illustration of how the dual basis elements act on a vector v. The diagram does not include the point p in the notation

For the moment we will work with manifold $\mathbb{R}^3$ since it is easier to draw the relevant pictures. Then we will move to $\mathbb{R}^2$ and even $\mathbb{R}$ before moving to the more general $\mathbb{R}^n$ case. First we recall what we learned in the last chapter. Figure 3.2 shows what the linear functionals, or dual basis elements, dx_1, dx_2, dx_3, (dx, dy, dz), do to the vector $v_p \in T_p\mathbb{R}^3$; they find the projection of v_p onto the coordinate axes of $T_p\mathbb{R}^3$. In general, the base point p is suppressed from the notation most of the time.

Figure 3.3 top depicts the one-form $\alpha = \alpha_1 dx + \alpha_2 dy + \alpha_3 dz$ as an element of the cotangent space $T_p^*\mathbb{R}^3$. The image below shows the result of the one-form acting on the vector $v_p \in T_p\mathbb{R}^3$. Each basis element dx, dy, and dz finds the projection of v onto the appropriate axis. The projection of v onto the x-axis is found as $dx(v) = v_1$ which is then scaled, or multiplied, by α_1. The projection of v onto the y-axis is found as $dy(v) = v_2$, which is then multiplied by α_2. The projection of v onto the z-axis is found as $dz(v) = v_3$, which is then multiplied by α_3. These three scaled terms, $\alpha_1 v_1$, $\alpha_2 v_2$, and $\alpha_3 v_3$ are then added together to obtain $\alpha(v) = \alpha_1 v_1 + \alpha_2 v_2 + \alpha_3 v_3$.

For example, since

$$dx_1(v_p) = dx_1\left(\begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix}\right) = v_1$$

we can view dx_1 as finding the projection of the vector v_p onto the $\frac{\partial}{\partial x_1}\big|_p$ coordinate axis of $T_p\mathbb{R}^3$. Similarly for dx_2 and dx_3. In other words, dx_1 finds a length along the $\frac{\partial}{\partial x_1}\big|_p$ coordinate axis. And what is length but a one-dimensional "volume"? We will want the product of two one-forms to do something similar, to find a volume of a two-dimensional projection. In the end this will mean that the product of two one-forms will no longer be a one-form, or even a linear functional. That means it will no longer be an element of $T_p^*\mathbb{R}^3$ but will be something different, something we will call a two-form. We will denote the space of two-forms on $\mathbb{R}^3$ as $\bigwedge^2(\mathbb{R}^3)$.

Here we are trying to be very precise about notation and about keeping track of what spaces objects are in. It is easy to see how confusion can arise. Using Fig. 3.4 as a guide consider the vector

$$\begin{bmatrix} dx_1(v_p) \\ dx_2(v_p) \\ 0 \end{bmatrix}_p .$$

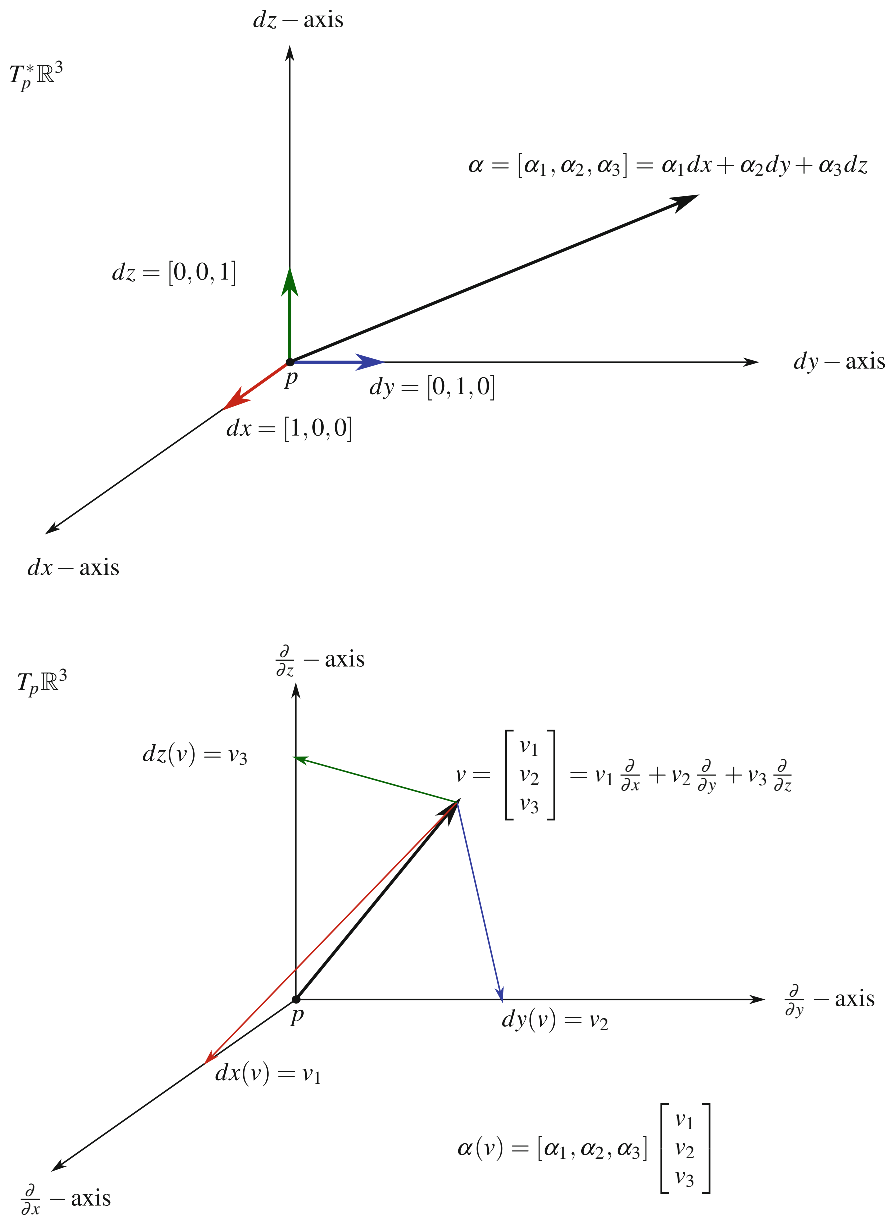


Fig. 3.3 Above the one-form $\alpha = \alpha_1 dx + \alpha_2 dy + \alpha_3 dz$ is shown as an element of the cotangent space $T_p^*\mathbb{R}^3$. Below the result of the one-form acting on the vector $v_p \in T_p\mathbb{R}^3$ is shown

This vector is in the tangent space of $\mathbb{R}^3$ at the point p. In other words, it is in $T_p\mathbb{R}^3$. But this vector is also the projection of v_p onto the $\frac{\partial}{\partial x_1}\big|_p \frac{\partial}{\partial x_2}\big|_p$-plane of $T_p\mathbb{R}^3$. Thus we can identify it with the vector

$$\begin{bmatrix} dx_1(v_p) \\ dx_2(v_p) \end{bmatrix}_p \in span\left\{ \frac{\partial}{\partial x_1}\bigg|_p, \frac{\partial}{\partial x_2}\bigg|_p \right\} \subset T_p\mathbb{R}^3.$$

This vector is generally called the projection of v_p onto the $\frac{\partial}{\partial x_1}\big|_p \frac{\partial}{\partial x_2}\big|_p$-plane of $T_p\mathbb{R}^3$.

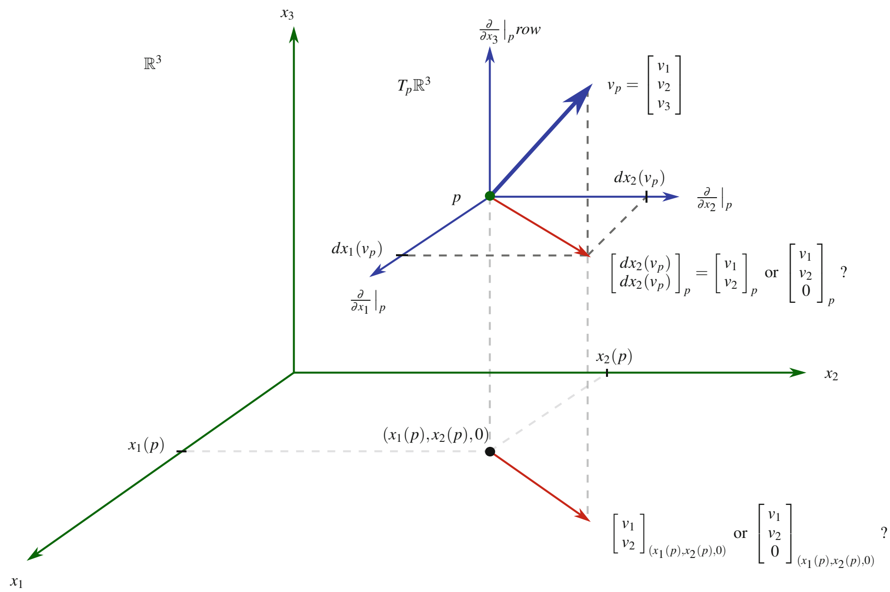


Fig. 3.4 Mixing things up

Looking again at Fig. 3.4 to guide you. Notice that if we are being imprecise how easy it is to confuse the vector

$$\begin{bmatrix} dx_1(v_p) \\ dx_2(v_p) \\ 0 \end{bmatrix}_p \in T_p\mathbb{R}^3$$

in the tangent space with the vector

$$\begin{bmatrix} dx_1(v_p) \\ dx_2(v_p) \end{bmatrix} \in \left.\frac{\partial}{\partial x_1}\right|_p \left.\frac{\partial}{\partial x_2}\right|_p \text{-plane of } T_p\mathbb{R}^3$$

with the vector

$$\begin{bmatrix} dx_1(v_p) \\ dx_2(v_p) \\ 0 \end{bmatrix}_p \in \mathbb{R}^3$$

at the point p in manifold $\mathbb{R}^3$ or even with the vector

$$\begin{bmatrix} dx_1(v_p) \\ dx_2(v_p) \\ 0 \end{bmatrix}_{(x_1(p),x_2(p),0)} \in \mathbb{R}^3$$

at the point $(x_1(p), x_2(p), 0)$ in manifold $\mathbb{R}^3$, or with the vector

$$\begin{bmatrix} dx_1(v_p) \\ dx_2(v_p) \end{bmatrix}_{(x_1(p),x_2(p),0)} \in xy\text{-plane of } \mathbb{R}^3$$

at the point $(x_1(p), x_2(p), 0)$ in the xy-plane of $\mathbb{R}^3$? In the case of $\mathbb{R}^3$ or $\mathbb{R}^n$ this may muddle things up in our minds but generally it does not result in any computational problems, which is the reason precise distinctions are genially not made in vector calculus. But in general we simply can not do this.

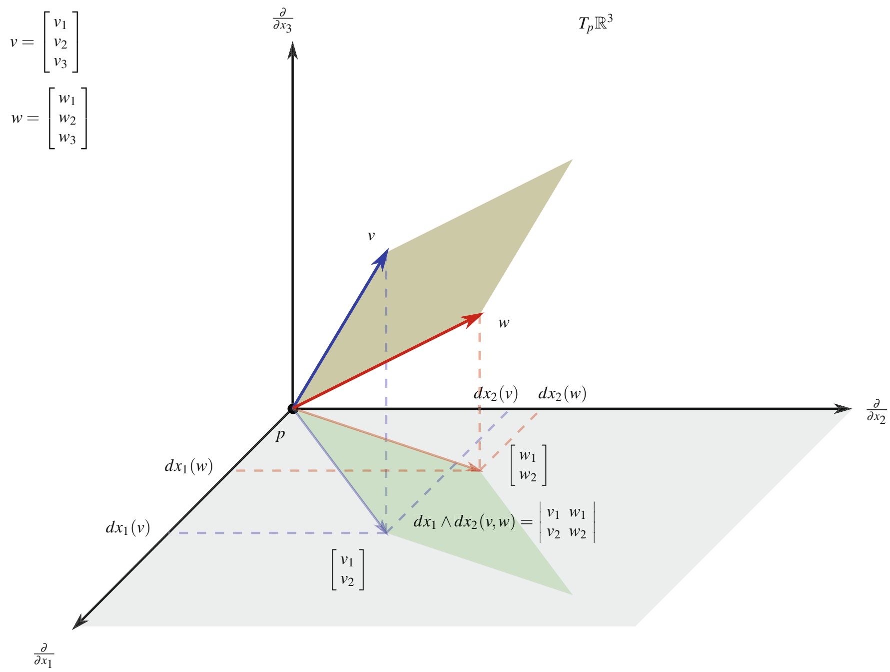


Fig. 3.5 The parallelepiped spanned by v and w (brown) is projected onto the $\partial_{x_1}\partial_{x_2}$-plane in $T_p\mathbb{R}^3$ (green). We want $dx_1 \wedge dx_2$ to find the volume of this projected area

Now that we have taken some effort to understand the spaces that various vectors are in, we turn our attention back our original problem of finding a product of two one-forms that does something similar to what a one-form does, that is, find some sort of a volume. Just like the one-form dx_1 took one vector v_p as input and give a one-dimensional volume as output we want the product of two one-forms, say dx_1 and dx_2, to take as input two vectors v_p and w_p and give as output a two-dimensional volume. What volume should it give? Consider Fig. 3.5.

We have shown the projection of vectors

$$v_p = \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix} \text{ and } w_p = \begin{bmatrix} w_1 \\ w_2 \\ w_3 \end{bmatrix}$$

onto the $\partial_{x_1}\partial_{x_2}$-plane in $T_p\mathbb{R}^3$. Notice we changed notation and wrote $\left.\frac{\partial}{\partial x_i}\right|_p$ as ∂_{x_i}. The projection of v_p onto the $\partial_{x_1}\partial_{x_2}$-plane is given by the vector

$$\begin{bmatrix} dx_1(v_p) \\ dx_2(v_p) \end{bmatrix} = \begin{bmatrix} v_1 \\ v_2 \end{bmatrix}.$$

Similarly, the projection of w_p onto the same plane is given by

$$\begin{bmatrix} dx_1(w_p) \\ dx_2(w_p) \end{bmatrix} = \begin{bmatrix} w_1 \\ w_2 \end{bmatrix}.$$

The most natural volume that we may want is the volume of the parallelepiped spanned by the projected vectors

$$\begin{bmatrix} dx_1(v_p) \\ dx_2(v_p) \end{bmatrix} \text{ and } \begin{bmatrix} dx_1(w_p) \\ dx_2(w_p) \end{bmatrix}$$

in the $\partial_{x_1}\partial_{x_2}$-plane in $T_p\mathbb{R}^3$. So, we want this "multiplication" or "product" of the two one-forms dx_1 and dx_2 to somehow give us this projected volume when we input the two vectors v_p and w_p.

To remind ourselves that this multiplication is actually different from anything we are used to so far we will use a special notation for it, called a wedge, $\wedge$, and call it something special as well, the **wedgeproduct**. So, this is what we want

$$dx_1 \wedge dx_2(v_p, w_p) = \begin{array}{c}\text{Volume of parallelepiped spanned} \\ \text{by the projection of} \\ v_p \text{ and } w_p \text{ onto } \partial_{x_1}\partial_{x_2}\text{-plane.}\end{array}$$

or,

$$dx_1 \wedge dx_2(v_p, w_p) = \begin{array}{c}\text{Volume of parallelepiped spanned} \\ \text{by } \begin{bmatrix} dx_1(v_p) \\ dx_2(v_p) \end{bmatrix} \text{ and } \begin{bmatrix} dx_1(w_p) \\ dx_2(w_p) \end{bmatrix}.\end{array}$$

But this is exactly how we derived the definition of determinant in the determinant section of chapter one. The volume of the parallelepiped spanned by the vectors

$$\begin{bmatrix} dx_1(v_p) \\ dx_2(v_p) \end{bmatrix} \text{ and } \begin{bmatrix} dx_1(w_p) \\ dx_2(w_p) \end{bmatrix}.$$

is given by the formula for the determinant of a matrix which has these vectors as columns. Thus we can use the determinant to help us define the wedgeproduct,

$$dx_1 \wedge dx_2(v_p, w_p) \equiv \begin{vmatrix} dx_1(v_p) & dx_1(w_p) \\ dx_2(v_p) & dx_2(w_p) \end{vmatrix}.$$

In summary, **the wedgeproduct of two one-forms is defined in terms of the determinant of the appropriate vector projections**. So we see that volumes, determinants, and projections (via the one-forms) are all mixed together and intimately related in the definition of the wedgeproduct. More generally, the wedgeproduct of two one-forms dx_i and dx_j is defined by

Wedgeproduct of two one-forms	$dx_i \wedge dx_j(v_p, w_p) \equiv \begin{vmatrix} dx_i(v_p) & dx_i(w_p) \\ dx_j(v_p) & dx_j(w_p) \end{vmatrix}.$

Sticking with the manifold $\mathbb{R}^3$ let us try to get a better picture of what is going on. Refer back to Fig. 3.5 where two vectors v and w are drawn in a tangent space $T_p\mathbb{R}^3$ at some arbitrary point p. (The point p is henceforth omitted from the notation.) The projections of v and w onto the $\partial_{x_1}\partial_{x_2}$-plane are given by

$$\begin{bmatrix} dx_1(v_p) \\ dx_2(v_p) \end{bmatrix} = \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} \text{ and } \begin{bmatrix} dx_1(w_p) \\ dx_2(w_p) \end{bmatrix} = \begin{bmatrix} w_1 \\ w_2 \end{bmatrix}.$$

The area of the parallelepiped spanned by these projections is given by $dx_1 \wedge dx_2(v, w)$. As an example, suppose that

$$v = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} \text{ and } w = \begin{bmatrix} 4 \\ 5 \\ 6 \end{bmatrix}.$$

First we find $dx_1 \wedge dx_2(v, w)$ as

$$dx_1 \wedge dx_2\left(\begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}, \begin{bmatrix} 4 \\ 5 \\ 6 \end{bmatrix}\right) = \begin{vmatrix} 1 & 4 \\ 2 & 5 \end{vmatrix} = (1)(5) - (4)(2) = -3.$$

Notice we have an area of -3. This shouldn't be surprising since we know that areas are really signed. If instead we found $dx_2 \wedge dx_1(v, w)$ as

$$dx_2 \wedge dx_1 \left(\begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}, \begin{bmatrix} 4 \\ 5 \\ 6 \end{bmatrix} \right) = \begin{vmatrix} 2 & 5 \\ 1 & 4 \end{vmatrix} = (2)(4) - (5)(1) = 3$$

we have an area of 3. A word of caution, when calculating $dx_1 \wedge dx_2$ our projected vectors are

$$\begin{bmatrix} dx_1(v_p) \\ dx_2(v_p) \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \end{bmatrix} \text{ and } \begin{bmatrix} dx_1(w_p) \\ dx_2(w_p) \end{bmatrix} = \begin{bmatrix} 4 \\ 5 \end{bmatrix}$$

while when calculating $dx_2 \wedge dx_1$ our projected vectors are

$$\begin{bmatrix} dx_2(v_p) \\ dx_1(v_p) \end{bmatrix} = \begin{bmatrix} 2 \\ 1 \end{bmatrix} \text{ and } \begin{bmatrix} dx_2(w_p) \\ dx_1(w_p) \end{bmatrix} = \begin{bmatrix} 5 \\ 4 \end{bmatrix},$$

which may look like different vectors. However, they are not different vectors. We have just written the vector components in a different order, that is all. Instead of writing the x_1 component first as you typically would we have written the x_2 component first.

In Fig. 3.6 the same two vectors v and w are drawn in the same tangent space $T_p\mathbb{R}^3$ at some arbitrary point p. (Again,the point p is henceforth omitted from the notation.) The projections of v and w onto the $\partial_{x_2}\partial_{x_3}$-plane are given by

$$\begin{bmatrix} dx_2(v_p) \\ dx_3(v_p) \end{bmatrix} = \begin{bmatrix} v_2 \\ v_3 \end{bmatrix} \text{ and } \begin{bmatrix} dx_2(w_p) \\ dx_3(w_p) \end{bmatrix} = \begin{bmatrix} w_2 \\ w_3 \end{bmatrix}.$$

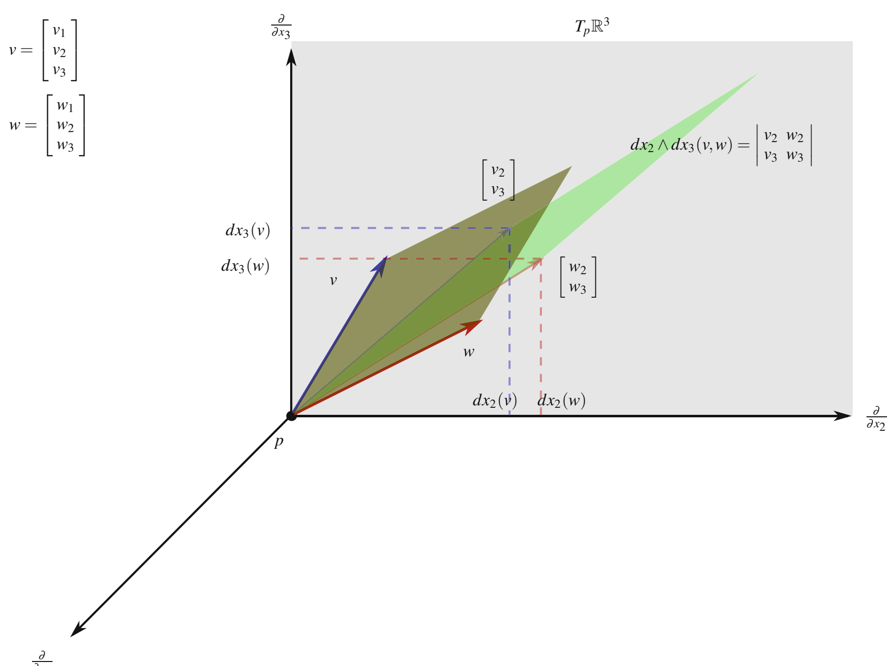


Fig. 3.6 The parallelepiped spanned by v and w (brown) is projected onto the $\partial_{x_2}\partial_{x_3}$-plane in $T_p\mathbb{R}^3$ (green). The wedgeproduct $dx_2 \wedge dx_3$ will find the volume of this projected area when the vectors v and w are its input

The area of the parallelepiped spanned by these projections is given by $dx_2 \wedge dx_3(v, w)$. Continuing with the same example, we find $dx_2 \wedge dx_3(v, w)$ as

$$dx_2 \wedge dx_3\left(\begin{bmatrix}1\\2\\3\end{bmatrix}, \begin{bmatrix}4\\5\\6\end{bmatrix}\right) = \begin{vmatrix}2 & 5\\3 & 6\end{vmatrix} = (2)(6) - (3)(5) = -3.$$

whereas $dx_3 \wedge dx_2(v, w)$ gives

$$dx_3 \wedge dx_2\left(\begin{bmatrix}1\\2\\3\end{bmatrix}, \begin{bmatrix}4\\5\\6\end{bmatrix}\right) = \begin{vmatrix}3 & 6\\2 & 5\end{vmatrix} = (3)(5) - (6)(2) = 3.$$

Again, the same two vectors v and w are drawn in the same tangent space $T_p\mathbb{R}^3$ and the projections of v and w onto the $\partial_{x_1}\partial_{x_3}$-plane are given by

$$\begin{bmatrix}dx_1(v_p)\\dx_3(v_p)\end{bmatrix} = \begin{bmatrix}v_1\\v_3\end{bmatrix} \text{ and } \begin{bmatrix}dx_1(w_p)\\dx_3(w_p)\end{bmatrix} = \begin{bmatrix}w_1\\w_3\end{bmatrix}.$$

This is shown in Fig. 3.7. The area of the parallelepiped spanned by these projections is given by $dx_1 \wedge dx_3(v, w)$.

Using the same example we find $dx_1 \wedge dx_3(v, w)$ as

$$dx_1 \wedge dx_3\left(\begin{bmatrix}1\\2\\3\end{bmatrix}, \begin{bmatrix}4\\5\\6\end{bmatrix}\right) = \begin{vmatrix}1 & 4\\3 & 6\end{vmatrix} = (1)(6) - (3)(4) = -6.$$

whereas $dx_3 \wedge dx_1(v, w)$ gives

$$dx_3 \wedge dx_1\left(\begin{bmatrix}1\\2\\3\end{bmatrix}, \begin{bmatrix}4\\5\\6\end{bmatrix}\right) = \begin{vmatrix}3 & 6\\1 & 4\end{vmatrix} = (3)(4) - (1)(6) = 6.$$

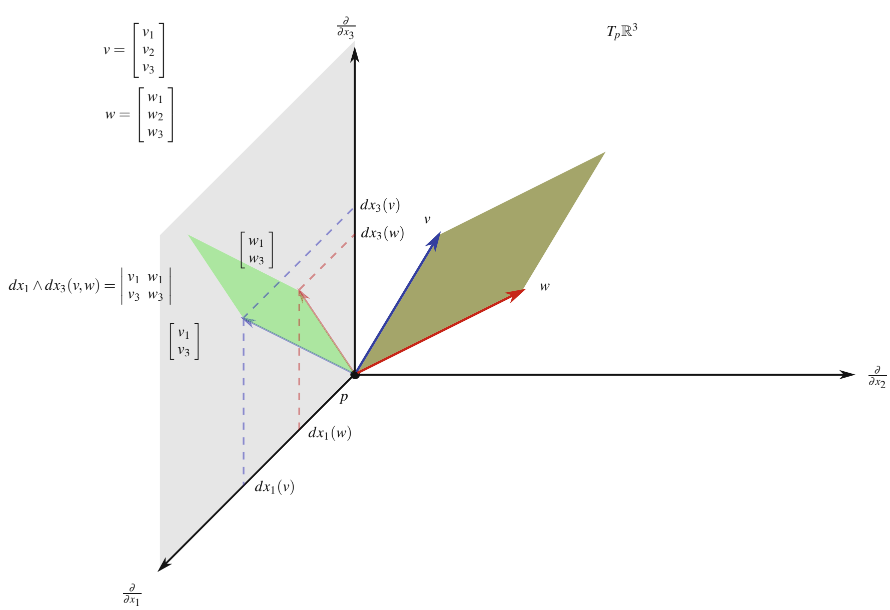


Fig. 3.7 The parallelepiped spanned by v and w (brown) is projected onto the $\partial_{x_1}\partial_{x_3}$-plane in $T_p\mathbb{R}^3$ (green). The wedgeproduct $dx_1 \wedge dx_3$ will find the volume of this projected area when the vectors v and w are its input

Based on these examples it appears that $dx_i \wedge dx_j = -dx_j \wedge dx_i$.Using the definition of wedgeproduct given above, it is simple to see that indeed $dx_i \wedge dx_j = -dx_j \wedge dx_i$. This property is called **skew symmetry** and it follows from the properties of the determinant. Recall, if you switch two rows in the determinant the sign of the determinant changes. That is essentially what is happening here. This implies that the order in which we "multiply" (via wedgeproduct) two one-forms matters. If we switch the order our answer changes by a sign.

Sticking with the same vectors $v = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$ and $w = \begin{bmatrix} 4 \\ 5 \\ 6 \end{bmatrix}$ let us find $dx_1 \wedge dx_1(v_p, w_p)$,

$$dx_1 \wedge dx_1 \left(\begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}, \begin{bmatrix} 4 \\ 5 \\ 6 \end{bmatrix} \right) = \begin{vmatrix} 1 & 4 \\ 1 & 4 \end{vmatrix} = (1)(4) - (4)(1) = 0.$$

Similarly we have $dx_2 \wedge dx_2(v, w) = 0$ and $dx_3 \wedge dx_3(v, w) = 0$. This should not be too surprising since in general we would have

$$dx_i \wedge dx_i(v, w) = \begin{vmatrix} dx_i(v) & dx_i(w) \\ dx_i(v) & dx_i(w) \end{vmatrix} = dx_i(v) \cdot dx_i(w) - dx_i(v) \cdot dx_i(w) = 0.$$

Another way to see this is to note that since $dx_i \wedge dx_j = -dx_j \wedge dx_i$ that implies $dx_i \wedge dx_i = -dx_i \wedge dx_i$, which can only happen if $dx_i \wedge dx_i = 0$. Unfortunately there is no accurate picture that shows what is going on here. It is best to simply recognize that the parallelepiped spanned by the projected vectors is degenerate and thus has area zero.

So, given two vectors v_p and w_p in $T_p\mathbb{R}^3$ we now considered every possible wedgeproduct of two one-forms. We have considered

$$\begin{aligned} &dx_1 \wedge dx_2, \\ &dx_2 \wedge dx_3, \\ &dx_1 \wedge dx_3. \end{aligned}$$

We have also discovered that

$$\begin{aligned} dx_1 \wedge dx_2 &= -dx_2 \wedge dx_1, \\ dx_2 \wedge dx_3 &= -dx_3 \wedge dx_2, \\ dx_1 \wedge dx_3 &= -dx_3 \wedge dx_1, \end{aligned}$$

so once we know what the left hand side does to two vectors we automatically know what the right hand side does to these same vectors. And finally, we found that

$$\begin{aligned} dx_1 \wedge dx_1 &= 0, \\ dx_2 \wedge dx_2 &= 0, \\ dx_3 \wedge dx_3 &= 0. \end{aligned}$$

That is all there is. The wedgeproducts $dx_1 \wedge dx_2$, $dx_2 \wedge dx_3$, and $dx_1 \wedge dx_3$ are all called two-forms. In fact, every single two-form can be written as linear combination of these three two-forms. For example, for any $a, b, c \in \mathbb{R}$ we have that

$$a\ dx_1 \wedge dx_2 + b\ dx_2 \wedge dx_3 + c\ dx_1 \wedge dx_3$$

is a two-form. For this reason, the two-forms $dx_1 \wedge dx_2$, $dx_2 \wedge dx_3$, and $dx_1 \wedge dx_3$ are called a basis of the space of two-form on $\mathbb{R}^3$, which is denoted as $\bigwedge^2(\mathbb{R}^3)$. Actually, there is a habit, or convention, that we use the two-forms $dx_1 \wedge dx_2$, $dx_2 \wedge dx_3$, and $dx_3 \wedge dx_1$ as the basis of $\bigwedge^2(\mathbb{R}^3)$. We have just substituted $dx_3 \wedge dx_1$ for $dx_1 \wedge dx_3$. The reason we do that is to maintain

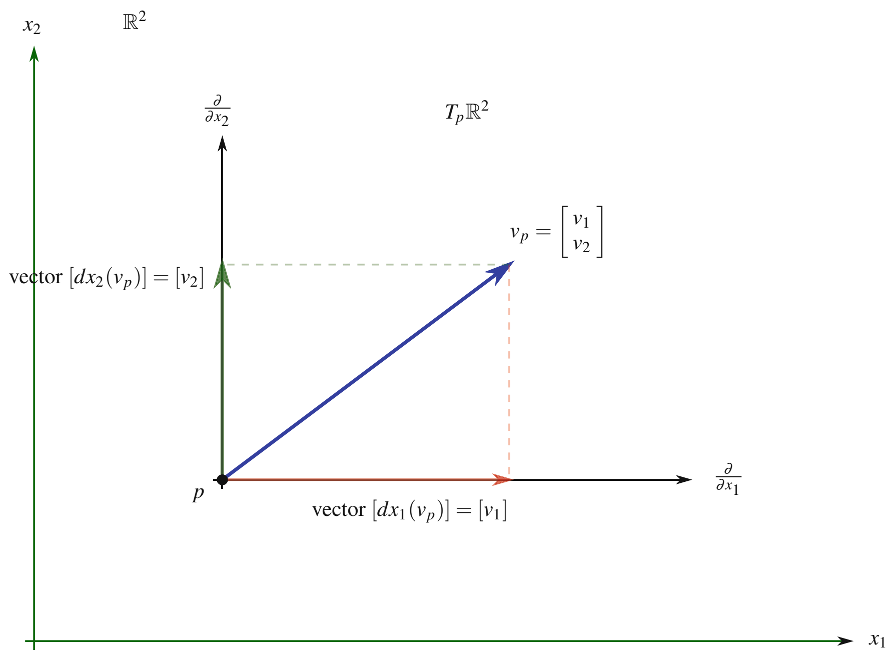


Fig. 3.8 The manifold $\mathbb{R}^2$ pictured with the tangent space $T_p\mathbb{R}^2$ superimposed on it. The vector v_p is an element of the tangent space $T_p\mathbb{R}^2$

a "cyclic ordering" when we write the basis down, something we can only do in three dimensions. But more about this in the next section.

Up to now we have been working with $\mathbb{R}^3$ because it is easy to draw picture and to see what we mean by projections. Now we step back and think about $\mathbb{R}^2$ as pictured in Fig. 3.8 for a few moments. Projections of $v_p = \begin{bmatrix} v_1 \\ v_2 \end{bmatrix}_p$ onto the one-dimensional subspace ∂_{x_1} of $T_p\mathbb{R}^2$ is given by the vector $[dx_1(v_p)] = [v_1]$ and span and the projection of v_p onto the subspace ∂_{x_2} is given by the vector $[dx_2(v_p)] = [v_2]$ as pictured.

But now consider Fig. 3.9. It is similar to Fig. 3.8 except now there are two vectors, v_p and w_p in $T_p\mathbb{R}^2$, that form a parallelepiped. What does $dx_1 \wedge dx_2(v_p, w_p)$ represent here? What space is v_p and w_p being projected on in this case? We have

$$dx_1 \wedge dx_2(v_p, w_p) = \begin{vmatrix} dx_1(v_p) & dx_1(w_p) \\ dx_2(v_p) & dx_2(w_p) \end{vmatrix} = \begin{vmatrix} v_1 & w_1 \\ v_2 & w_2 \end{vmatrix} = v_1 w_2 - w_1 v_2$$

which is exactly the area of the parallelepiped spanned by

$$\begin{bmatrix} dx_1(v_p) \\ dx_2(v_p) \end{bmatrix}_p = \begin{bmatrix} v_1 \\ v_2 \end{bmatrix}_p = v_p$$

and

$$\begin{bmatrix} dx_1(w_p) \\ dx_2(w_p) \end{bmatrix}_p = \begin{bmatrix} w_1 \\ w_2 \end{bmatrix}_p = w_p$$

so the space that v_p and w_p are getting "projected" onto is the whole space $T_p\mathbb{R}^2$.

Question 3.1 Find $dx_2 \wedge dx_1(v_p, w_p)$. How does it relate to $dx_1 \wedge dx_2(v_p, w_p)$. What would the basis of $\bigwedge_p^2 \left(\mathbb{R}^2\right)$ be? What would a general element of $\bigwedge_p^2 \left(\mathbb{R}^2\right)$ look like?

Question 3.2 For the moment we will remain with our $\mathbb{R}^2$ example.

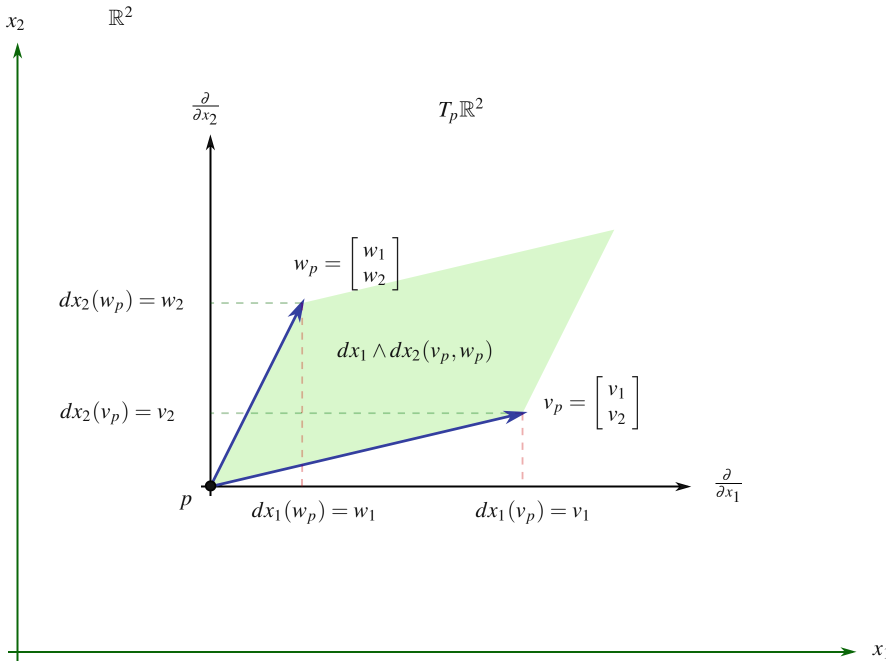


Fig. 3.9 The manifold $\mathbb{R}^2$ pictured with the tangent space $T_p\mathbb{R}^2$ superimposed on it. The parallelepiped spanned by v_p and w_p is shown. But what space does this get projected to when $dx_1 \wedge dx_2$ acts on the vectors v_p and w_p? It gets projected to the exact same space that it is already in, namely, $T_p\mathbb{R}^2$

(a) Consider $dx_1 \wedge dx_2 \wedge dx_1$. How many vectors do you think this would take as input?
(b) What would the three-dimensional volume of the parallelepiped spanned by three vectors be in a two-dimensional space?
(c) Suppose $u = \begin{bmatrix} u_1 \\ u_2 \end{bmatrix}$, $v = \begin{bmatrix} v_1 \\ v_2 \end{bmatrix}$, and $s = \begin{bmatrix} w_1 \\ w_2 \end{bmatrix}$ and suppose that

$$dx_1 \wedge dx_2 \wedge dx_1(u, v, w) \equiv \begin{vmatrix} dx_1(u) & dx_1(v) & dx_1(w) \\ dx_2(u) & dx_2(v) & dx_2(w) \\ dx_1(u) & dx_1(v) & dx_1(w) \end{vmatrix}.$$

Explain why this is always zero.
(d) Based on this, what do you think the space $\bigwedge_p^3 \left(\mathbb{R}^2\right)$ will be?
(e) Based on this, what do you think the space $\bigwedge_p^n \left(\mathbb{R}^2\right)$ for $n > 3$ will be?

We will define the wedgeproduct of three one-forms as follows,

Wedgeproduct of three one-forms

$$dx_i \wedge dx_j \wedge dx_k(u, v, w) \equiv \begin{vmatrix} dx_i(u) & dx_i(v) & dx_i(w) \\ dx_j(u) & dx_j(v) & dx_j(w) \\ dx_k(u) & dx_k(v) & dx_k(w) \end{vmatrix}.$$

We will call the wedgeproduct of three one-forms $dx_i \wedge dx_j \wedge dx_k$ a three-form. What does this three-form $dx_i \wedge dx_j \wedge dx_k$ find? Suppose we are on manifold $\mathbb{R}^n$. Given vectors u, v, w at a point p we first find the projection of these vectors onto the $\partial_{x_i} \partial_{x_i} \partial_{x_k}$-subspace of $T_p\mathbb{R}^n$. These projections are given by the vectors

$$\begin{bmatrix} dx_i(u) \\ dx_j(u) \\ dx_k(u) \end{bmatrix}, \begin{bmatrix} dx_i(v) \\ dx_j(v) \\ dx_k(v) \end{bmatrix}, \begin{bmatrix} dx_i(w) \\ dx_j(w) \\ dx_k(w) \end{bmatrix}.$$

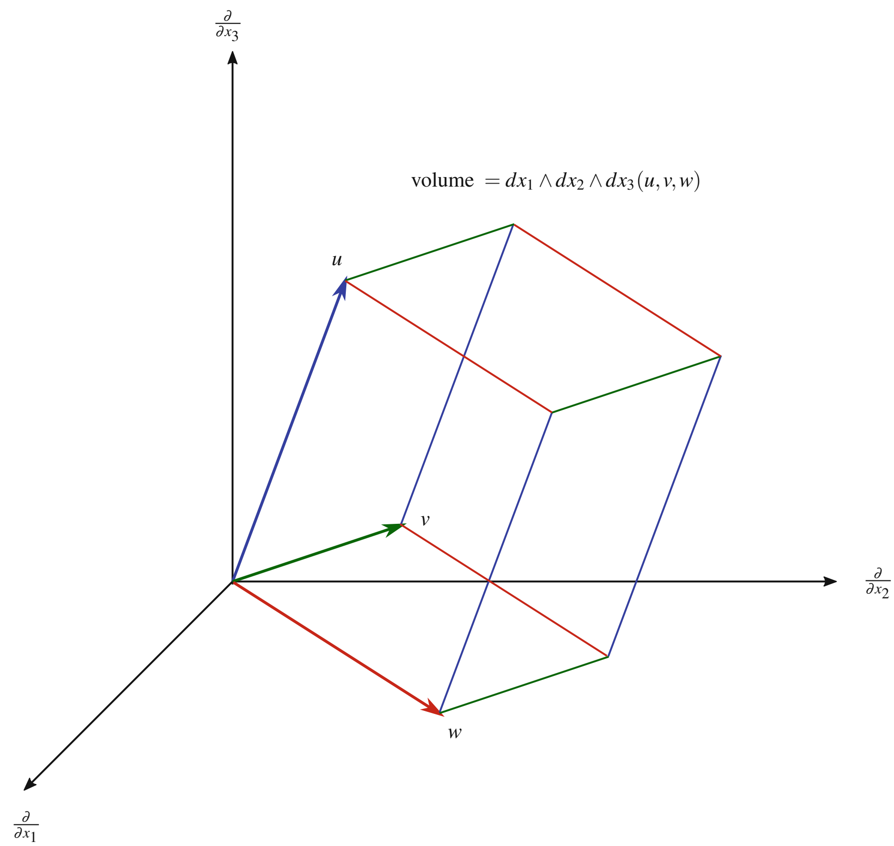


Fig. 3.10 The tangent space $T_p\mathbb{R}^3$ with the parallelepiped spanned by u, v and w shown. The three-form $dx_1 \wedge dx_2 \wedge dx_3$, shown here as $dx \wedge dy \wedge dz$, simply finds the volume of this parallelepiped

Next, we find the volume of the parallelepiped spanned by these projected vectors by using the determinant. Returning to $\mathbb{R}^3$ again we will consider the three-form $dx_1 \wedge dx_2 \wedge dx_3$. What subspace of $T_p\mathbb{R}^3$ does the three-form $dx_1 \wedge dx_2 \wedge dx_3$ project onto? The three-form $dx_1 \wedge dx_2 \wedge dx_3$ projects onto the whole space $T_p\mathbb{R}^3$. Thus the three-form $dx_1 \wedge dx_2 \wedge dx_3$ simply finds the volume of the parallelepiped spanned by the three input vectors, see Fig. 3.10.

Question 3.3 Consider the manifold $\mathbb{R}^3$ and the three-form $dx_1 \wedge dx_2 \wedge dx_3$.

(a) How does $dx_1 \wedge dx_2 \wedge dx_3$ relate to $dx_2 \wedge dx_1 \wedge dx_3$? How about to $dx_2 \wedge dx_3 \wedge dx_1$? Or $dx_3 \wedge dx_2 \wedge dx_1$? Do you notice a pattern?
(b) What is the basis for $\bigwedge_p^3(\mathbb{R}^3)$?
(c) How do you think $dx_1 \wedge dx_2 \wedge dx_3 \wedge dx_1(u, v, w, x)$ would be defined? What do you think it will be? Why? (Here the input x represents a fourth vector.)
(d) What is $\bigwedge_p^n(\mathbb{R}^3)$ for $n \geq 4$?

In general, letting $v_1, v_2, \ldots, v_n$ represent vectors, we will define the wedgeproduct of n one-forms as follows:

Wedgeproduct of n one-forms
$$dx_{i_1} \wedge dx_{i_2} \wedge \cdots \wedge dx_{i_n}\big(v_1, v_2, \ldots v_n\big) \equiv \begin{vmatrix} dx_{i_1}(v_1) & dx_{i_1}(v_2) & \cdots & dx_{i_1}(v_n) \\ dx_{i_2}(v_1) & dx_{i_2}(v_2) & \cdots & dx_{i_2}(v_n) \\ \vdots & \vdots & \ddots & \vdots \\ dx_{i_n}(v_1) & dx_{i_n}(v_2) & \cdots & dx_{i_n}(v_n) \end{vmatrix}$$

Notice, if we have any two of the one-forms the same, that is, $i_j = i_k$ for some $j \neq k$ then we have two rows that are the same, which gives a value of zero.

To close this section we will consider the one-form dx on the manifold $\mathbb{R}$, the two-form $dx_1 \wedge dx_2$ on the manifold $\mathbb{R}^2$, the three-form $dx_1 \wedge dx_2 \wedge dx_3$ on the manifold $\mathbb{R}^3$, and the n-form $dx_1 \wedge dx_2 \wedge \cdots \wedge dx_n$ on the manifold $\mathbb{R}^n$. These are very special differential forms that are very often called ***volume forms***, though in two dimensions the two-form $dx_1 \wedge dx_2$ is often also called an ***area form***. After this section the reason behind this terminology is fairly obvious. The one-form dx finds the one-dimensional volume, or length, of vectors on $\mathbb{R}$. Similarly, the two-form $dx_1 \wedge dx_2$, also written as $dx \wedge dy$, finds the two-dimensional volume, or area, of parallelepipeds on $\mathbb{R}^2$ (Fig. 3.9), the three-form $dx_1 \wedge dx_2 \wedge dx_3$, also written as $dx \wedge dy \wedge dz$, finds the three-dimensional volume of parallelepipeds on $\mathbb{R}^3$ (Fig. 3.10), and obviously the n-form $dx_1 \wedge dx_2 \wedge \cdots \wedge dx_n$ finds the n-dimensional volume of parallelepipeds on $\mathbb{R}^n$. Technically speaking, the parallelepipeds whose volumes are found are actually in some tangent space and not in the manifold, but for Euclidian space we can actually also imagine the parallelepipeds as being in the manifolds themselves so we will not belabor the point too much. These volume forms will play a vital role when we get to integration of forms.

Question 3.4 Consider the n-form $dx_{i_1} \wedge dx_{i_2} \wedge \cdots \wedge dx_{i_n}$. What happens if any one-forms $dx_{i_j} = dx_{i_k}$ for $i_j \neq i_k$? What happens if any vector $v_j = v_k$ for $j \neq k$?

Question 3.5 Show that if $i_j \neq i_k$ and we switch dx_{i_j} and dx_{i_k} that the wedgeproduct changes sign.

3.2 General Two-Forms and Three-Forms

Two-forms are built by wedgeproducting two one-forms together. Just as we denoted the cotangent space at the point p by $T_p^*\mathbb{R}^3$, which was the space of one-forms at p, we will denote the space of two-forms at p as $\bigwedge_p^2(\mathbb{R}^3)$. Keeping in line with this convention, you will occasionally see $T_p^*\mathbb{R}^3$ written as $\bigwedge_p^1(\mathbb{R}^3)$. As we have already mentioned, the basis of the space $\bigwedge_p^2(\mathbb{R}^3)$ is given by

$$\Big\{dx_1 \wedge dx_2,\ dx_2 \wedge dx_3,\ dx_3 \wedge dx_1\Big\}.$$

Why is the last element in the list written as $dx_3 \wedge dx_1$ instead of $dx_1 \wedge dx_3$? This is purely a matter of convention based on a desire to write the indexes cyclicly. Each of these basis elements $dx_1 \wedge dx_2,\ dx_2 \wedge dx_3,\ dx_3 \wedge dx_1$ finds the signed areas of the parallelepipeds formed by the projections of two vectors onto the appropriate planes of $T_p\mathbb{R}^3$, as depicted in Fig. 3.11.

Similarly, it is not difficult to see that the basis of the space $\bigwedge_p^2(\mathbb{R}^2)$ is given by

$$\Big\{dx_1 \wedge dx_2\Big\}$$

and the basis of the space $\bigwedge_p^2(\mathbb{R}^4)$ is given by

$$\Big\{dx_1 \wedge dx_2,\ dx_1 \wedge dx_3,\ dx_1 \wedge dx_4,\ dx_2 \wedge dx_3,\ dx_2 \wedge dx_4,\ dx_3 \wedge dx_4\Big\},$$

etc. The basis of $\bigwedge_p^2(\mathbb{R}^n)$ contains $\frac{(n)(n-1)}{2}$ elements.

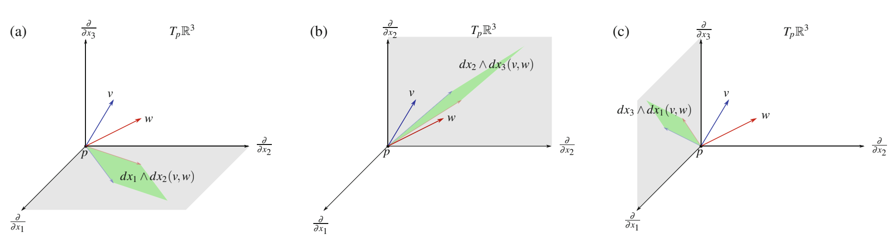


Fig. 3.11 The actions of the basis elements of $\bigwedge^2(\mathbb{R}^3)$ projecting v and w onto the appropriate planes in $T_{\mathbb{R}}^3$ and then finding the signed area of parallelepiped spanned by the projections, which is shown in green. (**a**) Action of $dx_1 \wedge dx_2$. (**b**) Action of $dx_2 \wedge dx_3$. (**c**) Action of $dx_3 \wedge dx_1$

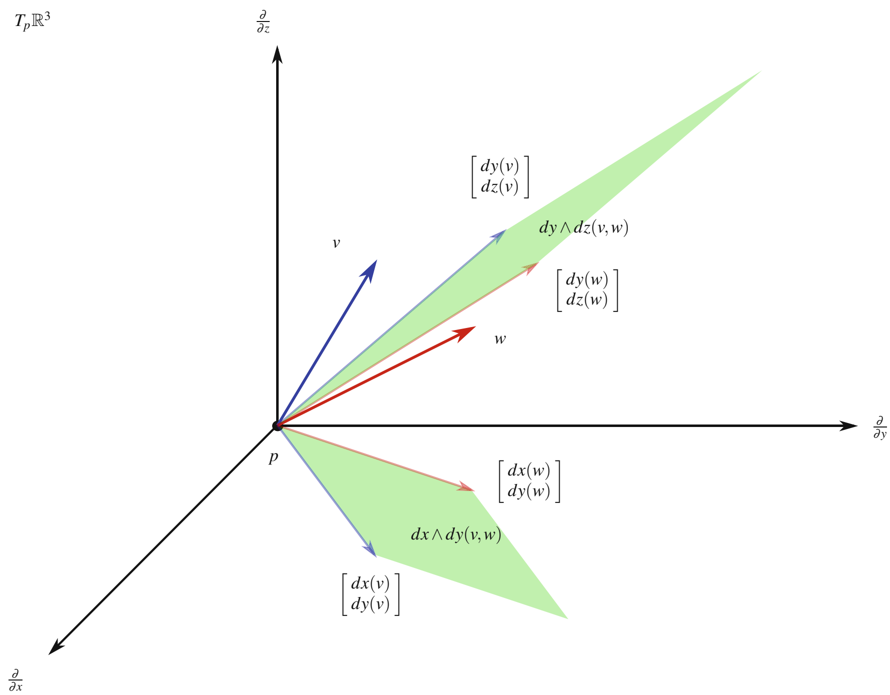


Fig. 3.12 The action of the two-form $dx \wedge dy + dy \wedge dz$ on the two vectors v and w. The two-form $dx \wedge dy + dy \wedge dz$ finds the sum of the areas of the two projected parallelepipeds

Question 3.6 Explain why $\{dx_1 \wedge dx_2\}$ is the basis of $\bigwedge_p^2(\mathbb{R}^2)$ and why $\{dx_1 \wedge dx_2,\ dx_1 \wedge dx_3,\ dx_1 \wedge dx_4,\ dx_2 \wedge dx_3,\ dx_2 \wedge dx_4,\ dx_3 \wedge dx_4\}$ is the basis of $\bigwedge_p^2(\mathbb{R}^4)$.

Question 3.7 Explain why the basis of $\bigwedge_p^2(\mathbb{R}^n)$ contains $\frac{(n)(n-1)}{2}$ elements.

Thus far we have simply looked at the basis elements of the two-forms. But what about the other elements in the spaces $\bigwedge_p^2(\mathbb{R}^2)$, $\bigwedge_p^2(\mathbb{R}^3)$, $\bigwedge_p^2(\mathbb{R}^4)$, or $\bigwedge_p^2(\mathbb{R}^n)$? What do they look like and what do they do? For the moment we will concentrate on the space $\bigwedge_p^2(\mathbb{R}^3)$ and so will resort to the notation x, y, z instead of x_1, x_2, x_3.

Consider the two-form $5dx \wedge dy$ acting on two vectors v_p and w_p for some p. What does the factor 5 do? $dx \wedge dy(v_p, w_p)$ gives the area of the parallelepiped spanned by the projection of v_p and w_p onto the $\partial_x\partial_y$-plane. The number five scales this area by multiplying the area by five. Thus the numerical factors in front of the basis elements $dx \wedge dy$, $dy \wedge dz$, $dz \wedge dx$ are scaling factors.

Figure 3.12 is an attempt to show what a two-form of the form $dx \wedge dy + dy \wedge dz$ does to two vectors v and w in $T_p\mathbb{R}^3$. The projections of v and w onto the $\partial_x\partial_y$-plane are found

$$\begin{bmatrix} dx(v) \\ dy(v) \end{bmatrix} \text{ and } \begin{bmatrix} dx(w) \\ dy(w) \end{bmatrix}$$

and the parallelepiped in the $\partial_x\partial_y$-plane formed from these two projected vectors is drawn. $dx \wedge dy(v, w)$ finds the signed area of this parallelepiped. Then the projections of v and w onto the $\partial_y\partial_z$-plane are found

$$\begin{bmatrix} dy(v) \\ dz(v) \end{bmatrix} \text{ and } \begin{bmatrix} dy(w) \\ dz(w) \end{bmatrix}$$

and the parallelepiped in the $\partial_y\partial_z$-plane formed from these two projected vectors is also drawn. $dx \wedge dy(v, w)$ finds the signed area of this parallelepiped. The one-form $dx \wedge dy + dy \wedge dz$ sums these two signed areas:

$$\Big(dx \wedge dy + dy \wedge dz\Big)(v, w) = dx \wedge dy(v, w) + dy \wedge dz(v, w).$$

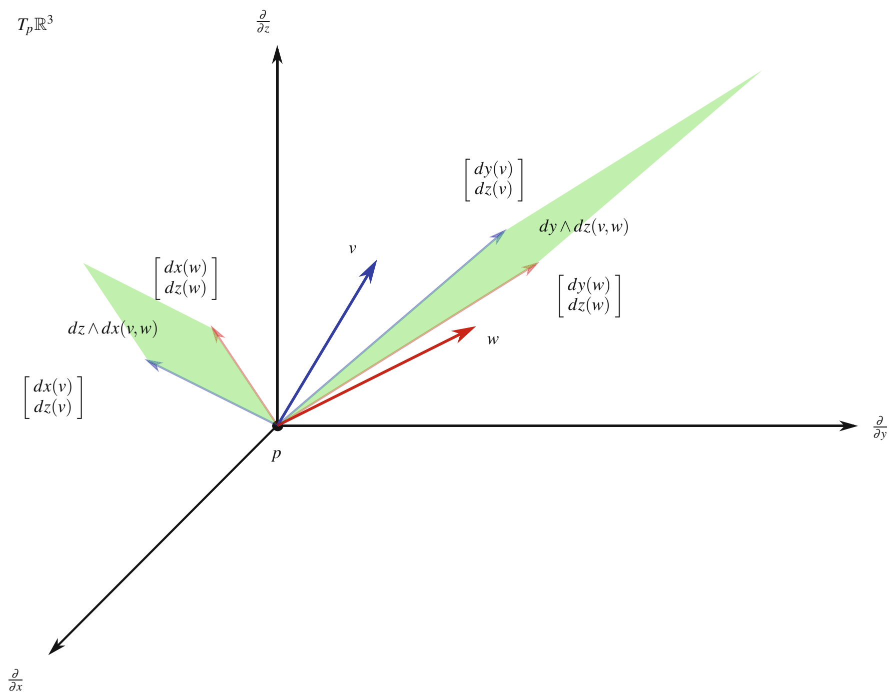


Fig. 3.13 The action of the two-form $dy \wedge dz + dz \wedge dx$ on the two vectors v and w. The two-form $dy \wedge dz + dz \wedge dx$ finds the sum of the areas of the two projected parallelepipeds

A two-form of the form $adx \wedge dy + bdy \wedge dz$, where $a, b \in \mathbb{R}$, multiplies the parallelepiped areas by the factors a and b before adding them:

$$\Big(adx \wedge dy + bdy \wedge dz\Big)(v, w) = adx \wedge dy(v, w) + bdy \wedge dz(v, w).$$

Figure 3.13 shows that the action of the two-form $dy \wedge dz + dz \wedge dx$ on two vectors v and w in $T_p\mathbb{R}^3$ is essentially similar. The projections of v and w onto the $\partial_y\partial_z$-plane are found

$$\begin{bmatrix} dy(v) \\ dz(v) \end{bmatrix} \text{ and } \begin{bmatrix} dy(w) \\ dz(w) \end{bmatrix}$$

and the parallelepiped in the $\partial_y\partial_z$-plane formed from these two projected vectors is drawn. $dy \wedge dz(v, w)$ finds the signed area of this parallelepiped. Then the projections of v and w onto the $\partial_x\partial_z$-plane are found

$$\begin{bmatrix} dx(v) \\ dz(v) \end{bmatrix} \text{ and } \begin{bmatrix} dx(w) \\ dz(w) \end{bmatrix}$$

and the parallelepiped in the $\partial_x\partial_z$-plane formed from these two projected vectors is also shown. $dz \wedge dx(v, w)$ finds the signed area of this parallelepiped. The one-form $dy \wedge dz + dz \wedge dx$ sums these two signed areas:

$$\Big(dy \wedge dz + dz \wedge dx\Big)(v, w) = dy \wedge dz(v, w) + dz \wedge dx(v, w).$$

A two-form of the form $ady \wedge dz + bdz \wedge dx$, where $a, b \in \mathbb{R}$, multiplies the parallelepiped areas by the factors a and b before adding them:

$$\Big(ady \wedge dz + bdz \wedge dx\Big)(v, w) = ady \wedge dz(v, w) + bdz \wedge dx(v, w).$$

Without going into all the details again, the two-forms $dx \wedge dy + dz \wedge dx$ and $adx \wedge dy + bdz \wedge dx$ behaves analogously when they act on two vectors $v, w \in T_p\mathbb{R}^3$, see Fig. 3.14.

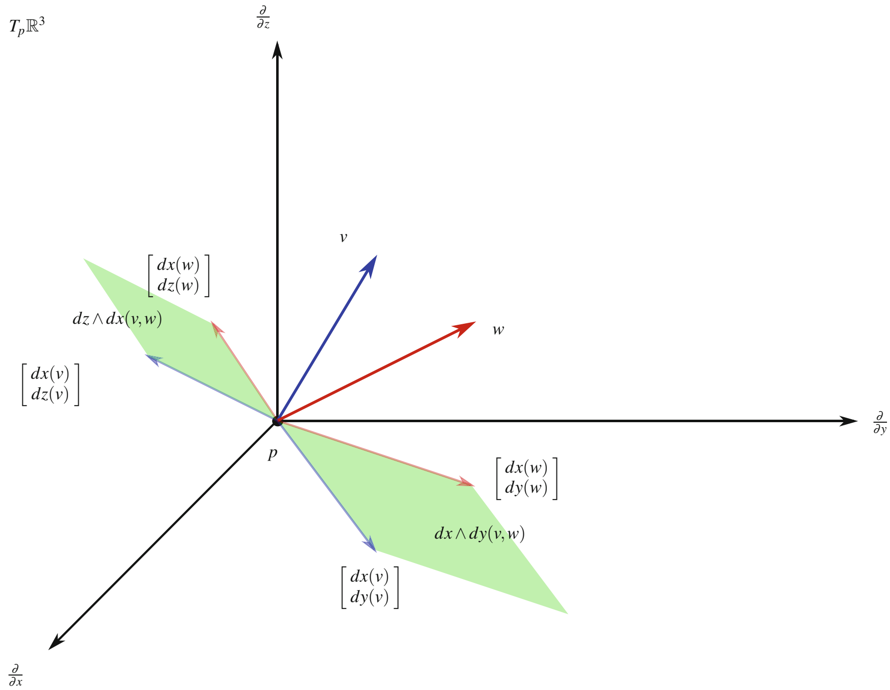


Fig. 3.14 The action of the two-form $dx \wedge dy + dz \wedge dx$ on the two vectors v and w. The two-form $dx \wedge dy + dz \wedge dx$ finds the sum of the areas of the two projected parallelepipeds

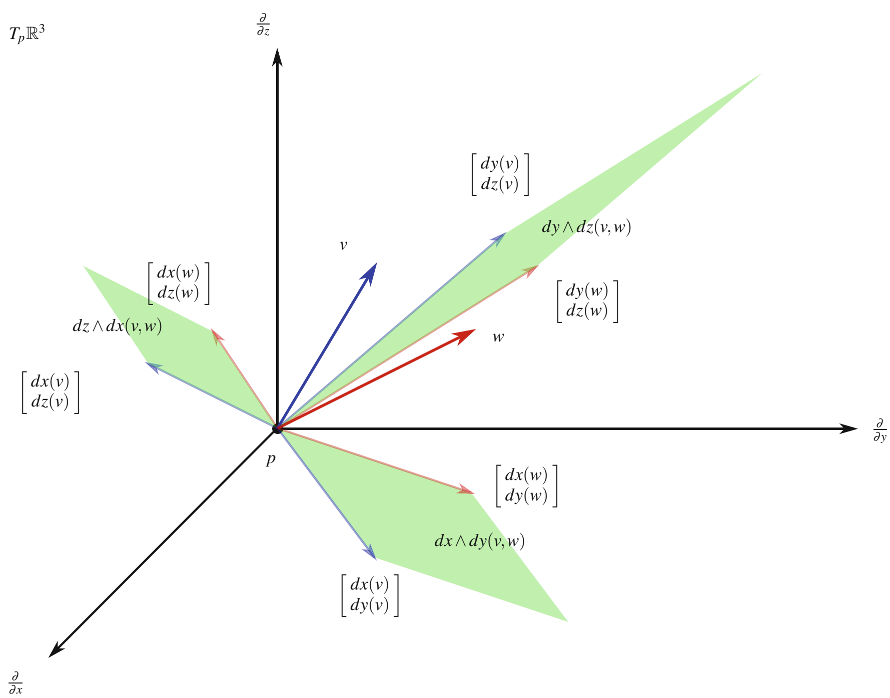


Fig. 3.15 The action of the two-form $dx \wedge dy + dy \wedge dz + dz \wedge dx$ on two vectors v and w, which are not shown

Finally, we consider two-forms of the form $dx \wedge dy + dy \wedge dz + dz \wedge dx$. This two-forms acts on two vectors $v, w \in T_p\mathbb{R}^3$ exactly as you would expect, as shown in Fig. 3.15. First, the projections onto the different planes in $T_p\mathbb{R}^3$ are found. Then the parallelepipeds formed from these projected vectors are drawn. $dx \wedge dy(v, w)$ finds the signed volume of the parallelepiped in the $\partial_x\partial_y$-plane, $dy \wedge dz(v, w)$ finds the signed volume of the parallelepiped in the $\partial_y\partial_z$-plane, and $dz \wedge dx(v, w)$ finds the signed volume of the parallelepiped in the $\partial_x \partial_z$-plane. These signed volumes are then summed up:

$$\Big(dx \wedge dy + dy \wedge dz + dz \wedge dx\Big)(v, w) = dx \wedge dy(v, w) + dy \wedge dz(v, w) + dz \wedge dx(v, w).$$

Two-forms of the form $adx \wedge dy + bdy \wedge dz + cdz \wedge dx$, where $a, b, c \in \mathbb{R}$ work analogously, only with the various signed volumes being scaled by the factors a, b, c before being summed together.

Two forms on n-dimensional manifolds, such as $\mathbb{R}^n$ where $n > 3$, behave in exactly analogous ways. Projections of two vectors v, w are found on the appropriate two dimensional subspaces of $T_p\mathbb{R}^n$, two-dimensional parallelepipeds are formed from these projected vectors, their volumes are first found and then scaled by the appropriate factor, and then the scaled volumes are summed. The only distinction is that $T_p\mathbb{R}^n$ has $\frac{n(n-1)}{2}$ distinct two dimensional subspaces.

We have already discussed the space $\bigwedge_p^3(\mathbb{R}^3)$ of three-forms on $\mathbb{R}^3$, but we will cover it again here for completeness. The basis of $\bigwedge_p^3(\mathbb{R}^3)$ is given by $\{dx \wedge dy \wedge dz\}$ since the three-form $dx \wedge dy \wedge dz$ projects three vectors u, v, w onto the $\partial_x \partial_y \partial_z$-subspace of $T_p\mathbb{R}^3$, which happens to be the whole space $T_p\mathbb{R}^3$. See Fig. 3.10. Thus, the three-form $dx \wedge dy \wedge dz$ simply finds the volume of the parallelepiped spanned by the vectors u, v, w. Elements of $\bigwedge_p^3(\mathbb{R}^3)$ of the form $adx \wedge dy \wedge dz$, where $a \in \mathbb{R}$, simply scales the volume by the factor a.

Three-forms only start to get interesting for manifolds of dimension four or higher. As an illustrative example let us consider $\bigwedge_p^3(\mathbb{R}^4)$, the three-forms on the manifold $\mathbb{R}^4$. A little bit of thought should convince you that the basis of $\bigwedge_p^3(\mathbb{R}^4)$ is given by

$$\Big\{dx_1 \wedge dx_2 \wedge dx_3,\ dx_1 \wedge dx_2 \wedge dx_4,\ dx_1 \wedge dx_3 \wedge dx_4,\ dx_2 \wedge dx_3 \wedge dx_4\Big\}$$

and all elements of $\bigwedge_p^3(\mathbb{R}^4)$ are of the form

$$adx_1 \wedge dx_2 \wedge dx_3 + bdx_1 \wedge dx_2 \wedge dx_4 + cdx_1 \wedge dx_3 \wedge dx_4 + ddx_2 \wedge dx_3 \wedge dx_4$$

for $a, b, c, d \in \mathbb{R}$.

Question 3.8 Explain why $\{dx_1 \wedge dx_2 \wedge dx_3,\ dx_1 \wedge dx_2 \wedge dx_4,\ dx_1 \wedge dx_3 \wedge dx_4,\ dx_2 \wedge dx_3 \wedge dx_4\}$ is the basis of $\bigwedge_p^3(\mathbb{R}^4)$. What are all the possible three-dimensional subspaces of $\mathbb{R}^4$?

The picture we will employ for "visualizing" the actions of these three-forms on three vectors $u, v, w \in T_p\mathbb{R}^4$ will, out of necessity, be a bit of a "cartoon," but it works well enough; see Fig. 3.16. In fact, similar cartoons can be used for all n-forms on m-dimensional manifolds. The projections of the four dimensional vectors

$$u = \begin{bmatrix} u_1 \\ u_2 \\ u_3 \\ u_4 \end{bmatrix},\quad v = \begin{bmatrix} v_1 \\ v_2 \\ v_3 \\ v_4 \end{bmatrix},\quad w = \begin{bmatrix} w_1 \\ w_2 \\ w_3 \\ w_4 \end{bmatrix}$$

onto the $\partial_{x_1} \partial_{x_2} \partial_{x_3}$-subspace are found to be

$$\begin{bmatrix} dx_1(u) \\ dx_2(u) \\ dx_3(u) \end{bmatrix} = \begin{bmatrix} u_1 \\ u_2 \\ u_3 \end{bmatrix},\quad \begin{bmatrix} dx_1(v) \\ dx_2(v) \\ dx_3(v) \end{bmatrix} = \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix},\quad \begin{bmatrix} dx_1(w) \\ dx_2(w) \\ dx_3(w) \end{bmatrix} = \begin{bmatrix} w_1 \\ w_2 \\ w_3 \end{bmatrix}.$$

The three-form $dx_1 \wedge dx_2 \wedge dx_3(u, v, w)$ finds the signed volume of the parallelepiped spanned by the projected vectors. The three-form $adx_1 \wedge dx_2 \wedge dx_3(u, v, w)$ scales the signed volume by the factor a. The other three-forms work similarly and can be imagined using the same sort of cartoon. Finally, an arbitrary element of $\bigwedge_p^3(\mathbb{R}^4)$ finds the appropriate signed volumes, scales the signed volumes by the appropriate factors, and then sums the results together according to

$$\begin{aligned}
&\Big(a dx_1 \wedge dx_2 \wedge dx_3 + b dx_1 \wedge dx_2 \wedge dx_4 \\
&\quad + c dx_1 \wedge dx_3 \wedge dx_4 + d dx_2 \wedge dx_3 \wedge dx_4\Big)(u, v, w) \\
=&a dx_1 \wedge dx_2 \wedge dx_3(u, v, w) + b dx_1 \wedge dx_2 \wedge dx_4(u, v, w) \\
&\quad + c dx_1 \wedge dx_3 \wedge dx_4(u, v, w) + d dx_2 \wedge dx_3 \wedge dx_4(u, v, w)
\end{aligned}$$

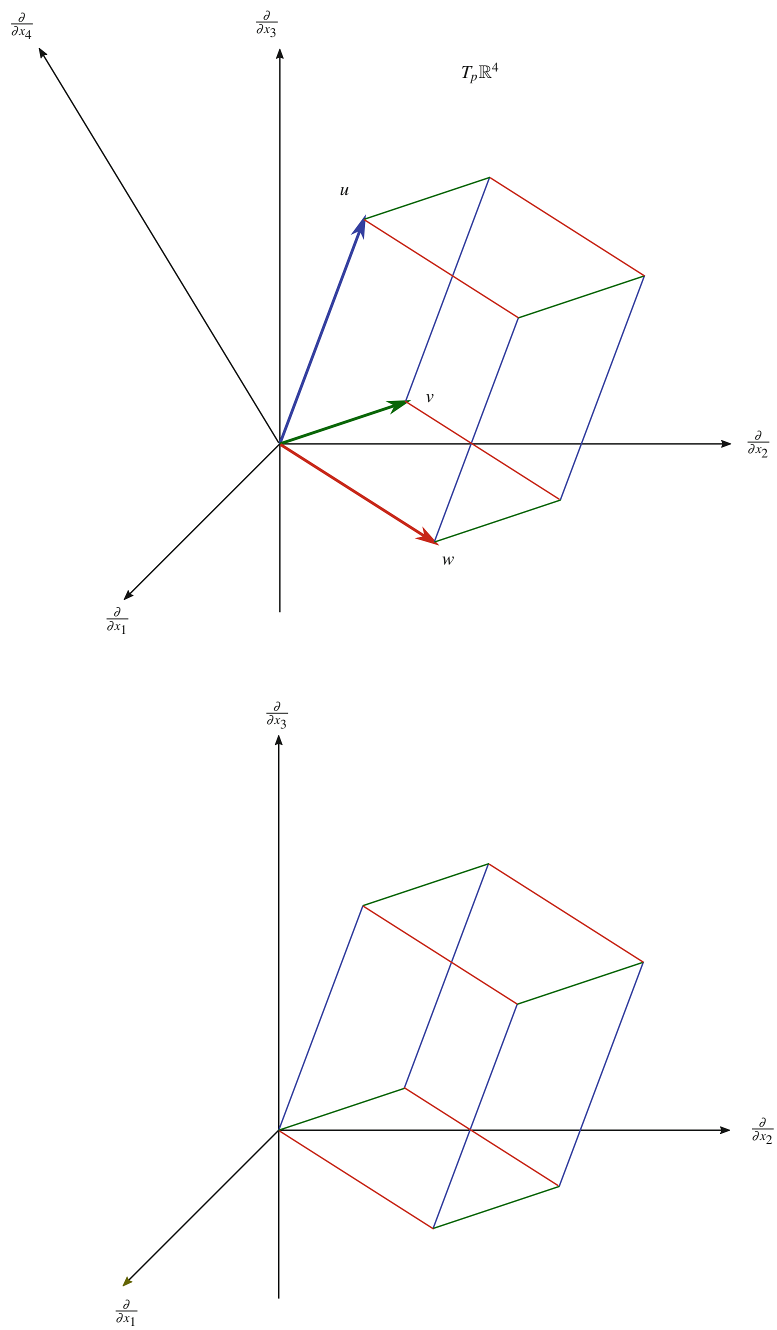


Fig. 3.16 An attempt is made to illustrate the tangent space $T_p\mathbb{R}^4$ and show the parallelepiped spanned by v_1, v_2, v_3 (top). The action of the three-form $dx_1 \wedge dx_2 \wedge dx_3$ on the three vectors u, v, w is shown as the projection of this parallelepiped to the three-dimensional $\partial_{x_1}\partial_{x_2}\partial_{x_3}$ subspace (below)

Question 3.9 Consider the manifold $\mathbb{R}^3$.

(a) Which of the following are one-forms on $\mathbb{R}^3$?

(i) $3dx_1$
(ii) $-4dx_2 + 7dx_3$
(iii) $5dx_1 + 3dx_2 - 6dx_1 + 4dx_3$
(iv) $2dx_1 \wedge dx_2 - 4dx_3$
(v) $dx_1 \wedge dx_2 - 3dx_2 \wedge dx_3$

(b) Which of the following are two-forms on $\mathbb{R}^3$?

(i) $-4dx_3 \wedge dx_2 \wedge dx_1$
(ii) $6dx_3 \wedge dx_2 + 8dx_3$
(iii) $-10dx_1 \wedge dx_3 + 5dx_2 \wedge dx_3 - dx_1 \wedge dx_2$
(iv) $-2dx_1 \wedge dx_2 + 3dx_3 \wedge dx_2 \wedge dx_1$
(v) $3dx_1 \wedge dx_3 - dx_2 \wedge dx_3$

(c) Which of the following are three-forms on $\mathbb{R}^3$?

(i) $5dx_3 \wedge dx_2 \wedge dx_1$
(ii) $-3dx_3 \wedge dx_2 + 2dx_3 \wedge dx_1 \wedge dx_2$
(iii) $5dx_2 \wedge dx_3 \wedge dx_1 + 5dx_1 \wedge dx_3 \wedge dx_2 - dx_1 \wedge dx_2 \wedge dx_3$
(iv) $-5dx_3 \wedge dx_2 + 3dx_1 \wedge dx_2 - dx_3 \wedge dx_1$
(v) $9dx_1 - 7dx_2 \wedge dx_3$

3.3 The Wedgeproduct of *n*-Forms

Now that we have an idea of what the wedgeproduct does geometrically we want to be able to manipulate wedgeproducts of forms quickly with algebraic formulas. In this section we learn some of the algebra related to differential forms. First of all we will see that differential forms essentially follows the rules we expect from algebra with only minor tweaks to take into account of the fact that $dx_i \wedge dx_j = -dx_j \wedge dx_i$. After that we will introduce a little bit of common notation, and then we will write out a general formula for the wedgeproduct.

3.3.1 Algebraic Properties

Suppose that $\omega, \omega_1, \omega_2$ are k-forms and η, η_1, η_2 are ℓ-forms. Then the following properties hold:

(1) $a\omega \wedge \eta = \omega \wedge a\eta = a\big(\omega \wedge \eta\big)$ where $a \in \mathbb{R}$,
(2) $\big(\omega_1 + \omega_2\big) \wedge \eta = \omega_1 \wedge \eta + \omega_2 \wedge \eta$,
(3) $\omega \wedge \big(\eta_1 + \eta_2\big) = \omega \wedge \eta_1 + \omega \wedge \eta_2$,
(4) $\omega \wedge \eta = (-1)^{k\ell}\eta \wedge \omega$.

We begin by showing these algebraic properties are reasonable by simply looking at a few easy examples on $\mathbb{R}^3$. We will begin with the first property and show that $(adx_1) \wedge dx_2 = dx_1 \wedge (adx_2) = a(dx_1 \wedge dx_2)$. First we have

$$
\begin{aligned}
(adx_1) \wedge dx_2(u, v) &= \begin{vmatrix} adx_1(u) & adx_1(v) \\ dx_2(u) & dx_2(v) \end{vmatrix} \\
&= adx_1(u)dx_2(v) - adx_1(v)dx_2(u) \\
&= a\big(dx_1(u)dx_2(v) - dx_1(v)dx_2(u)\big) \\
&= a \begin{vmatrix} dx_1(u) & dx_1(v) \\ dx_2(u) & dx_2(v) \end{vmatrix} \\
&= a\big(dx_1 \wedge dx_2\big)(u, v).
\end{aligned}
$$

Similarly,

$$
\begin{aligned}
dx_1 \wedge (adx_2)(u, v) &= \begin{vmatrix} dx_1(u) & dx_1(v) \\ adx_2(u) & adx_2(v) \end{vmatrix} \\
&= adx_1(u)dx_2(v) - adx_1(v)dx_2(u) \\
&= a\big(dx_1(u)dx_2(v) - dx_1(v)dx_2(u)\big) \\
&= a \begin{vmatrix} dx_1(u) & dx_1(v) \\ dx_2(u) & dx_2(v) \end{vmatrix} \\
&= a\big(dx_1 \wedge dx_2\big)(u, v).
\end{aligned}
$$

Putting all of this together and not writing the vectors gives us the identity we wanted. Next we look at an example that illustrates the second property; we show that $\big(dx_1 + dx_2\big) \wedge dx_3 = dx_1 \wedge dx_3 + dx_2 \wedge dx_3$.

$$
\begin{aligned}
&\Big(\big(dx_1 + dx_2\big) \wedge dx_3\Big)(u, v) \\
&= \begin{vmatrix} \big(dx_1 + dx_2\big)(u) & \big(dx_1 + dx_2\big)(v) \\ dx_3(u) & dx_3(v) \end{vmatrix} \\
&= \big(dx_1(u) + dx_2(u)\big)dx_3(v) - \big(dx_1(v) + dx_2(v)\big)dx_3(u) \\
&= dx_1(u)dx_3(v) + dx_2(u)dx_3(v) - dx_1(v)dx_3(u) - dx_2(v)dx_3(u) \\
&= dx_1(u)dx_3(v) - dx_1(v)dx_3(u) + dx_2(u)dx_3(v) - dx_2(v)dx_3(u) \\
&= \begin{vmatrix} dx_1(u) & dx_1(v) \\ dx_3(u) & dx_3(v) \end{vmatrix} + \begin{vmatrix} dx_2(u) & dx_2(v) \\ dx_3(u) & dx_3(v) \end{vmatrix} \\
&= \big(dx_1 \wedge dx_2\big)(u, v) + \big(dx_2 \wedge dx_3\big)(u, v).
\end{aligned}
$$

Notice that the third property $dx_1 \wedge \big(dx_2 + dx_3\big) = dx_1 \wedge dx_2 + dx_1 \wedge dx_3$ is essentially the same.

The last property is very straightforward. We already know that $dx_1 \wedge dx_2 = -dx_2 \wedge dx_1$. In fact, if $i \neq j$ we have $dx_i \wedge dx_j = -dx_j \wedge dx_i$. Suppose we had a three-form wedged with a two-form, $(dx_1 \wedge dx_2 \wedge dx_3) \wedge (dx_4 \wedge dx_5)$. We want to show that

$$
(dx_1 \wedge dx_2 \wedge dx_3) \wedge (dx_4 \wedge dx_5) = (-1)^{3 \cdot 2}(dx_4 \wedge dx_5) \wedge (dx_1 \wedge dx_2 \wedge dx_3).
$$

This involves nothing more than counting up how many switches are made to rearrange the terms.

Question 3.10 Rearrange $dx_1 \wedge dx_2 \wedge dx_3 \wedge dx_4 \wedge dx_5$ to $dx_4 \wedge dx_5 \wedge dx_1 \wedge dx_2 \wedge dx_3$ using six switches. For example, moving from $dx_1 \wedge dx_2 \wedge dx_3 \wedge dx_4 \wedge dx_5$ to $dx_1 \wedge dx_2 \wedge dx_4 \wedge dx_3 \wedge dx_5$ requires one switch, and hence

$$
dx_1 \wedge dx_2 \wedge dx_3 \wedge dx_4 \wedge dx_5 = -dx_1 \wedge dx_2 \wedge dx_4 \wedge dx_3 \wedge dx_5
$$

since $dx_3 \wedge dx_4 = -dx_4 \wedge dx_3$.

The general definition of the wedgeproduct of n one-forms was given in Sect. 3.1 as

$$
dx_{i_1} \wedge dx_{i_2} \wedge \cdots \wedge dx_{i_n}\big(v_1, v_2, \dots v_n\big) \equiv \begin{vmatrix} dx_{i_1}(v_1) & dx_{i_1}(v_2) & \cdots & dx_{i_1}(v_n) \\ dx_{i_2}(v_1) & dx_{i_2}(v_2) & \cdots & dx_{i_2}(v_n) \\ \vdots & \cdots & \ddots & \vdots \\ dx_{i_n}(v_1) & dx_{i_n}(v_2) & \cdots & dx_{i_n}(v_n) \end{vmatrix},
$$

where $v_1, v_2, \ldots, v_n$ represented vectors. Geometrically speaking, the n-form $dx_{i_1} \wedge dx_{i_2} \wedge \cdots \wedge dx_{i_n}$ found the volume of the parallelepiped spanned by the projections of $v_1, v_2, \ldots, v_n$ onto the $\frac{\partial}{\partial x_{i_1}} \frac{\partial}{\partial x_{i_2}} \cdots \frac{\partial}{\partial x_{i_n}}$-subspace. Now recalling our formula for the determinant that we found in Sect. 1.2,

$$\begin{vmatrix} a_{11} & a_{12} & & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & & a_{nn} \end{vmatrix} = \sum_{\sigma \in S_n} sgn(\sigma) \prod_{i=1}^{n} a_{\sigma(i)i},$$

we can combine this with our definition for the wedgeproduct of n one-forms to get the following important formula for the wedgeproduct, where S_n is the set of permutations on n elements,

Wedgeproduct of n one-forms

$$dx_{i_1} \wedge dx_{i_2} \wedge \cdots \wedge dx_{i_n}\big(v_1, v_2, \ldots v_n\big) = \sum_{\sigma \in S_n} \mathrm{sgn}(\sigma) \prod_{j=1}^{n} dx_{\sigma(i_j)}(v_j).$$

In fact, occasionally you see the wedgeproduct defined in terms of this formula. The problem with defining the wedgeproduct this way to start with is that it is not at all clear what the formula is actually doing or why this definition is important. At least we now thoroughly understand exactly what the wedgeproduct finds. It is not difficult to use this formula to actually prove the algebraic properties in this section, as the following questions show.

Question 3.11 Let $\omega = a\ dx_{i_1} \wedge \cdots dx_{i_k}$ and let $\eta = b\ dx_{i_{k+1}}$, where $a, b \in \mathbb{R}$. Using the formula for the wedgeproduct of n one-forms and a procedure similar to the example above, show that

$$\omega \wedge \eta = ab\ dx_{i_1} \wedge \cdots dx_{i_k} \wedge dx_{i_{k+1}}.$$

Question 3.12 Let $\omega_1 = a_1\ dx_{i_1} \wedge \cdots dx_{i_k}$, $\omega_2 = a_2\ dx_{j_1} \wedge \cdots dx_{j_k}$, and $\eta = b\ dx_\ell$. Use the formula for the wedgeproduct of n one-forms to show that

$$\big(\omega_1 + \omega_2\big) \wedge \eta = \omega_1 \wedge \eta + \omega_2 \wedge \eta.$$

Explain how this also implies $\omega \wedge \big(\eta_1 + \eta_2\big) = \omega \wedge \eta_1 + \omega \wedge \eta_2$.

Question 3.13 Let $\omega = dx_{i_1} \wedge \cdots dx_{i_k}$ and $\eta = dx_{j_1} \wedge \cdots dx_{j_\ell}$. Show that

$$\omega \wedge \eta = (-1)^{k\ell} \eta \wedge \omega.$$

Question 3.14 Suppose that $\omega, \omega_1, \omega_2$ are k-forms and η, η_1, η_2 are ℓ-forms. Using the above three questions, complete the proofs of the algebraic properties in this section,

(1) $a\omega \wedge \eta = \omega \wedge a\eta = a\big(\omega \wedge \eta\big)$ where $a \in \mathbb{R}$,
(2) $\big(\omega_1 + \omega_2\big) \wedge \eta = \omega_1 \wedge \eta + \omega_2 \wedge \eta$,
(3) $\omega \wedge \big(\eta_1 + \eta_2\big) = \omega \wedge \eta_1 + \omega \wedge \eta_2$,
(4) $\omega \wedge \eta = (-1)^{k\ell} \eta \wedge \omega$.

3.3.2 *Simplifying Notation*

In the last section we looked at two-forms on $\mathbb{R}^2$, denoted by $\bigwedge^2(\mathbb{R}^3)$, and found that $dx_1 \wedge dx_2$, $dx_2 \wedge dx_3$, and $dx_3 \wedge dx_1$ was a basis of $\bigwedge^2(\mathbb{R}^3)$. Hence any two-form $\alpha \in \bigwedge^2(\mathbb{R}^3)$ was of the form

$$\alpha = a_{12} dx_1 \wedge dx_2 + a_{23} dx_2 \wedge dx_3 + a_{31} dx_3 \wedge dx_1$$

where a_{12}, a_{23}, and a_{31} are just constants from $\mathbb{R}$. We could just as easily have written

$$\alpha = a dx_1 \wedge dx^2 + b dx_2 \wedge dx_3 + c dx_3 \wedge dx_1$$

as we did in the last section, or even written

$$\alpha = a_1 dx_1 \wedge dx_2 + a_2 dx_2 \wedge dx_3 + a_3 dx_3 \wedge dx_1.$$

After all, what does it matter what we name a constant? But it turns out that indexing our constants with the same numbers that appear in the basis element that follows it is notationally convenient. Recall, we wrote the indices of the two-form basis elements on $\mathbb{R}^3$ in cyclic order, which meant that we wrote $dx_3 \wedge dx_1$ instead of $dx_1 \wedge dx_3$. This is the convention in vector calculus. But for manifolds $\mathbb{R}^n$, where $n > 3$ this convention no longer works and we simply write our indices in increasing order. If $\alpha \in \bigwedge^2(\mathbb{R}^4)$ we would have

$$\begin{aligned}\alpha = {} & a_{12} dx_1 \wedge dx_2 + a_{13} dx_1 \wedge dx_3 + a_{14} dx_1 \wedge dx_4 \\ & a_{23} dx_2 \wedge dx_3 + a_{24} dx_2 \wedge dx_4 + a_{34} dx_3 \wedge dx_4.\end{aligned}$$

The basis elements of $\bigwedge^3(\mathbb{R}^4)$ are

$$dx_1 \wedge dx_2 \wedge dx_3, \quad dx_1 \wedge dx_2 \wedge dx_4, \quad dx_1 \wedge dx_3 \wedge dx_4, \quad dx_2 \wedge dx_3 \wedge dx_4.$$

Again, note that the indices are all in increasing order. An arbitrary three-form $\alpha \in \bigwedge^3(\mathbb{R}^4)$ would be written as

$$\begin{aligned}\alpha = {} & a_{123} dx_1 \wedge dx_2 \wedge dx_3 + a_{124} dx_1 \wedge dx_2 \wedge dx_4 \\ & + a_{134} dx_1 \wedge dx_3 \wedge dx_4 + a_{234} dx_2 \wedge dx_3 \wedge dx_4.\end{aligned}$$

For arbitrary k-forms in $\bigwedge^k(\mathbb{R}^n)$ we would not want to actually write out all of the elements of the basis of $\bigwedge^k(\mathbb{R}^n)$, so instead we will write

$$\alpha = \sum_I a_I dx^I.$$

Here the I stands for our elements in the set of k increasing indices $i_1 i_2 \ldots i_k$, where $1 \le i_1 < i_2 < \cdots < i_k \le n$. That is,

$$I \in \mathfrak{I}_{k,n} = \left\{ (i_1 i_2 \ldots i_k) \,\middle|\, 1 \le i_1 < i_2 < \cdots < i_k \le n \right\}.$$

So, for $k = 2$ and $n = 4$ we would have

$$I \in \mathfrak{I}_{2,4} = \{12, 13, 14, 23, 24, 34\}$$

and for $k = 3$ and $n = 4$ we have

$$I \in \mathfrak{I}_{3,4} = \{123, 124, 134, 234\}.$$

In this notation we would have $dx^{12} \equiv dx_1 \wedge dx_2$, $dx^{24} \equiv dx_2 \wedge dx_4$, and $dx^{134} \equiv dx_1 \wedge dx_3 \wedge dx_4$, and so on. In case you are wondering why the indices on the left are superscripts while the indices on the right are subscripts, this is so the notation is compatible with Einstein summation notation. We will explain Einstein summation notation in Sect. 9.4. For now don't worry about it.

Suppose that we have $\alpha \in \bigwedge^k(\mathbb{R}^n)$ and $\beta \in \bigwedge^l(\mathbb{R}^n)$ where $\alpha = \sum_I a_I dx^I$ and $\beta = \sum_J b_J dx^J$. Then $\alpha \wedge \beta$ is given by

Wedgeproduct of two arbitrary forms, Formula One

$$\boxed{\left(\sum_I a_I dx^I\right) \wedge \left(\sum_J b_J dx^J\right) = \sum a_I b_J dx^I \wedge dx^J.}$$

In the sum on the right hand side we have $dx^I \wedge dx^J = 0$ if I and J have any indices in common.

Question 3.15 Let $\alpha = \sum_I a_I dx^I \in \bigwedge^k(\mathbb{R}^n)$ and $\beta = \sum_J b_J dx^J \in \bigwedge^l(\mathbb{R}^n)$. By repeated applications of the algebraic properties of the wedgeproduct prove this formula for the wedgeproduct of two arbitrary forms.

If I and J are disjoint then we have $dx^I \wedge dx^J = \pm dx^K$ where $K = I \cup J$, but is reordered to be in increasing order and elements with repeated indices are dropped. Thus formula one automatically gives us formula two,

Wedgeproduct of two arbitrary forms, Formula Two	$$\left(\sum_I a_I dx^I\right) \wedge \left(\sum_J b_J dx^J\right) = \sum_K \left(\sum_{\substack{I \cup J \\ I,J \text{ disjoint}}} \pm a_I b_J \right) dx^K.$$

An example should hopefully make this clearer. Let α and β be the following forms on $\mathbb{R}^8$

$$\alpha = 5dx_1 \wedge dx_2 - 6dx_2 \wedge dx_4 + 7dx_1 \wedge dx_7 + 2dx_2 \wedge dx_8,$$
$$\beta = 3dx_3 \wedge dx_5 \wedge dx_8 - 4dx_5 \wedge dx_6 \wedge dx_8.$$

Then we have $\alpha = \sum_I a_I dx^I$ where the Is are the elements of the set $\{12, 24, 17, 28\} \subset \mathfrak{I}_{2,8}$ and our coefficients are $a_{12} = 5$, $a_{24} = -6$, $a_{17} = 7$, and $a_{28} = 2$. Similarly, we have $\beta = \sum_J b_J dx^J$ where the Js are the elements of the set $\{358, 568\} \subset \mathfrak{I}_{3,8}$ and the coefficients are $b_{358} = 3$ and $b_{568} = -4$. So, what are our elements K? They are given by the set $I \cup J$ but reordered into increasing order and with elements with repeated indices dropped. First we have

$$\begin{aligned} I \cup J = \{&12358, 24358, 17358, 28358 \\ &12568, 24568, 17568, 28568\}. \end{aligned}$$

The first thing that we notice is that two elements of this set have indices that are repeated. Both 28358 and 28568 repeat the 8. And since both $dx_2 \wedge dx_8 \wedge dx_3 \wedge dx_5 \wedge dx_8 = 0$ and $dx_2 \wedge dx_8 \wedge dx_5 \wedge dx_6 \wedge dx_8 = 0$ we can drop these elements.

To get K we need to put the other sets of indices in increasing order. The first element, 12358, is already in increasing order so nothing needs to be done. The coefficient in front of dx^{12358} is simply $a_{12}b_{358}$. To get 24358 into increasing order all we need to do is one switch, we need to switch the 4 and the 3 to get 23458. The one switch gives a negative sign which shows up in the coefficient in front of dx^{23458}, which is $-a_{24}b_{358} = -(-6)(3) = 18$. Similarly for the other terms. Thus we end up with

$$\begin{aligned} K = \{&12358, 23458, 13578, \cancel{23588} \\ &12568, 24568, 15678, \cancel{25688}\}. \end{aligned}$$

Notice that the way we have defined K means we are not keeping track of the sign changes that occur when we have to make the switches to get the indices in order. That is why we have the $\pm$ when we write $dx^I \wedge dx^J = \pm dx^K$.

Question 3.16 Finish writing out $\alpha \wedge \beta$.

Question 3.17 Let

$$\alpha = 4dx_1 \wedge dx_3 + 5dx_3 \wedge dx_5 - 7dx_3 \wedge dx_9$$

be a two-form on $\mathbb{R}^9$ and let

$$\beta = 7dx_1 \wedge dx_4 \wedge dx_6 \wedge dx_8 - 3dx_2 \wedge dx_3 \wedge dx_7 \wedge dx_9 + dx_5 \wedge dx_6 \wedge dx_8 \wedge dx_9$$

be a four-form on $\mathbb{R}^9$. Find $\alpha \wedge \beta$ and then find $\beta \wedge \alpha$.

From the preceding questions you can see that actually using this formula means we have to go through a lot of work rearranging the coefficients of I and J to get K and finding the correct sign. Where this formula really comes in useful is when we are doing general calculations, instead of working with specific examples, and are not so worried about the exact form the wedgeproduct takes.

3.3.3 The General Formula

This section introduces several new formulas for the wedgeproduct of two arbitrary forms. In fact, the following three general formulas are very often encountered as definitions of the wedgeproduct. We will discuss two of the three formulas in depth in this section. The third is provided here simply for completeness' sake. Suppose that α is a k-form and β is a ℓ-form. Then the wedgeproduct of α with β is given by

Wedgeproduct of two arbitrary forms, Formula Three

$$(\alpha \wedge \beta)\big(v_1, \ldots, v_{k+\ell}\big) = \frac{1}{k!\ell!} \sum_{\sigma \in S_{k+\ell}} \operatorname{sgn}(\sigma)\alpha\big(v_{\sigma(1)}, \ldots, v_{\sigma(k)}\big)\beta\big(v_{\sigma(k+1)}, \ldots, v_{\sigma(k+\ell)}\big),$$

which can also be written as

Wedgeproduct of two arbitrary forms, Formula Four

$$(\alpha \wedge \beta)\big(v_1, \ldots, v_{k+\ell}\big) = \sum_{\substack{\sigma \text{ is a} \\ (k+\ell)- \\ \text{shuffle}}} \operatorname{sgn}(\sigma)\alpha\big(v_{\sigma(1)}, \ldots, v_{\sigma(k)}\big)\beta\big(v_{\sigma(k+1)}, \ldots, v_{\sigma(k+\ell)}\big).$$

In books where tensors are introduced first and then differential forms are defined as a certain type of tensor, the wedgeproduct is often defined by the formula

Wedgeproduct of two arbitrary forms, Formula Five

$$\alpha \wedge \beta = \frac{(k+\ell)!}{k!\ell!}\mathscr{A}(\alpha \otimes \beta)$$

where $\otimes$ is the tensor product and $\mathscr{A}$ is the skew-symmetrization (or anti-symmetrization) operator. The explanation for this particular definition is given in the appendix on tensors, Sect. A.5.

We now turn our attention to understanding formulas three and four, and to seeing that they are indeed identical to our volume/determinant based definition of the wedgeproduct. The first step is to show that the right hand sides of formulas three and four are equal to each other. Once we have done that we will show that these formulas are in fact derivable from our definition of the wedgeproduct of n one-forms and thus can also be used as equations that define the wedgeproduct; that is, we will show that these formulas are indeed formulas for $(\alpha \wedge \beta)\big(v_1, \ldots, v_{k+\ell}\big)$.

To start with consider the formula given by the right hand side of formula three,

$$\frac{1}{k!\ell!} \sum_{\sigma \in S_{k+\ell}} \operatorname{sgn}(\sigma)\alpha\big(v_{\sigma(1)}, \ldots, v_{\sigma(k)}\big)\beta\big(v_{\sigma(k+1)}, \ldots, v_{\sigma(k+\ell)}\big).$$

Where does this term $\frac{1}{k!\ell!}$ in the front come from? Consider a permutation $\sigma \in S_{k+\ell}$. There are $k!$ different permutations τ in $S_{k+\ell}$ that permute the first k terms but leave the terms $k+1, \ldots, k+\ell$ fixed. This means that for any permutation σ we have $\sigma\tau(k+1) = \sigma(k+1), \ldots, \sigma\tau(k+\ell) = \sigma(k+\ell)$. That means we have

$$\begin{aligned}
&\operatorname{sgn}(\sigma\tau)\alpha\big(v_{\sigma\tau(1)}, \ldots, v_{\sigma\tau(k)}\big)\beta\big(v_{\sigma\tau(k+1)}, \ldots, v_{\sigma\tau(k+\ell)}\big) \\
&= \operatorname{sgn}(\sigma\tau)\operatorname{sgn}(\tau)\alpha\big(v_{\sigma(1)}, \ldots, v_{\sigma(k)}\big)\beta\big(v_{\sigma(k+1)}, \ldots, v_{\sigma(k+\ell)}\big) \\
&= \operatorname{sgn}(\sigma)\underbrace{\operatorname{sgn}(\tau)\operatorname{sgn}(\tau)}_{=1}\alpha\big(v_{\sigma(1)}, \ldots, v_{\sigma(k)}\big)\beta\big(v_{\sigma(k+1)}, \ldots, v_{\sigma(k+\ell)}\big) \\
&= \operatorname{sgn}(\sigma)\alpha\big(v_{\sigma(1)}, \ldots, v_{\sigma(k)}\big)\beta\big(v_{\sigma(k+1)}, \ldots, v_{\sigma(k+\ell)}\big)
\end{aligned}$$

where the first equality follows from the fact that $\tau(1), \ldots \tau(k)$ is just a permutation of $1, \ldots, k$, which results in $\alpha\big(v_{\sigma\tau(1)}, \ldots, v_{\sigma\tau(k)}\big) = \operatorname{sgn}(\tau)\alpha\big(v_{\sigma(1)}, \ldots, v_{\sigma(k)}\big)$, and that $\sigma\tau(k+1) = \sigma(k+1), \ldots, \sigma\tau(k+\ell) = \sigma(k+l)$, which gives $\beta\big(v_{\sigma\tau(k+1)}, \dots, v_{\sigma\tau(k+\ell)}\big) = \beta\big(v_{\sigma(k+1)}, \dots, v_{\sigma(k+\ell)}\big)$. This means that given a particular σ then there are $k!$ other terms in the sum

$$\sum_{\sigma \in S_{k+\ell}} \operatorname{sgn}(\sigma)\alpha\big(v_{\sigma(1)}, \dots, v_{\sigma(k)}\big)\beta\big(v_{\sigma(k+1)}, \dots, v_{\sigma(k+\ell)}\big)$$

that are exactly the same. We divide the sum by $k!$ to eliminate the $k!$ repeated terms.

Similarly, given a permutation $\sigma \in S_{k+\ell}$ there are $\ell!$ different permutations τ in $S_{k+\ell}$ that permute the last ℓ terms but hold the terms $1, \dots, k$ fixed. An identical arguments shows we also need to divide by $\ell!$ to eliminate the $\ell!$ repeated terms. Thus the factor $\frac{1}{k!\ell!}$ eliminates all the repeated terms in the sum over $\sigma \in S_{k+\ell}$. This means that the right hand side of our first definition of the wedgeproduct eliminates all of the repeated terms.

Next we turn our attention to seeing that

$$\begin{aligned}
&\frac{1}{k!\ell!}\sum_{\sigma \in S_{k+\ell}} \operatorname{sgn}(\sigma)\alpha\big(v_{\sigma(1)}, \dots, v_{\sigma(k)}\big)\beta\big(v_{\sigma(k+1)}, \dots, v_{\sigma(k+\ell)}\big) \\
&= \sum_{\substack{\sigma \text{ is a} \\ (k,\ell)-\text{shuffle}}} \operatorname{sgn}(\sigma)\alpha\big(v_{\sigma(1)}, \dots, v_{\sigma(k)}\big)\beta\big(v_{\sigma(k+1)}, \dots, v_{\sigma(k+\ell)}\big).
\end{aligned}$$

A permutation $\sigma \in S_{k+\ell}$ is called a (k, ℓ)-shuffle if

$$\sigma(1) < \cdots < \sigma(k) \quad \text{and} \quad \sigma(k+1) < \cdots < \sigma(k+\ell).$$

In other words, if both the first k terms of the permutation are in increasing order and the last ℓ terms of the permutation are in increasing order.

Question 3.18 Find the $(2, 3)$-shuffles of 1 2 3 4 5.

When we are summing over $\sigma \in S_{k+\ell}$ on the left hand side, for each of these σ there are $k!\ell! - 1$ other permutations which give the same term. This is essentially the same argument that we just went through. Exactly one of these $k!\ell!$ different permutations that give the same term will actually be a (k, l)-shuffle. Thus, by choosing the particular permutation which is the (k, ℓ)-shuffle what we are doing is choosing one representative permutation from the $k!\ell!$ identical terms. So when we sum over all the (k, ℓ)-shuffles we are only summing over one term from each of the sets of identical terms that appear in the sum over all $\sigma \in S_{k+l}$, thereby eliminating the need to divide by $k!\ell!$. Hence we have shown that the two formulas on the right hand sides of formula three and formula four for the wedgeproduct of two differential forms are indeed equal.

Question 3.19 Fill in the details of this argument.

Now we will show that the right hand side of formula four is indeed equal to the left hand side of formula four. Once this is done formula three will automatically follow by the above. Before actually doing the general case we will show it is true for a simple example. Consider $\alpha = dx_2 \wedge dx_3 \in \bigwedge^2(\mathbb{R}^5)$ and $\beta = dx_5 \in \bigwedge^1(\mathbb{R}^5)$. We will first compute $\alpha \wedge \beta(v_1, v_2, v_3)$ from the volume/determinant-based definition of the wedgeproduct. We will then apply the right hand side of formula four to α and β and thereby see that they are equal, which is exactly what we want to show.

First we write out $\alpha \wedge \beta(v_1, v_2, v_3)$ using our volume/determinat based definition,

$$\begin{aligned}
&\alpha \wedge \beta(v_1, v_2, v_3) \\
&= \big(dx_2 \wedge dx_3 \wedge dx_5\big)(v_1, v_2, v_3) \\
&= \begin{vmatrix} dx_2(v_1) & dx_2(v_2) & dx_2(v_3) \\ dx_3(v_1) & dx_3(v_2) & dx_3(v_3) \\ dx_5(v_1) & dx_5(v_2) & dx_5(v_3) \end{vmatrix} \\
&= dx_2(v_1)dx_3(v_2)dx_5(v_3) + dx_2(v_2)dx_3(v_3)dx_5(v_1) \\
&\quad + dx_2(v_3)dx_3(v_1)dx_5(v_2) - dx_2(v_3)dx_3(v_2)dx_5(v_1) \\
&\quad - dx_2(v_2)dx_3(v_1)dx_5(v_3) - dx_2(v_1)dx_3(v_3)dx_5(v_2) \\
&= \big(dx_2dx_3dx_5 + dx_5dx_2dx_3 + dx_3dx_5dx_2 \\
&\quad - dx_5dx_3dx_2 - dx_3dx_2dx_5 - dx_2dx_5dx_3\big)(v_1, v_2, v_3).
\end{aligned}$$

Now we will use the right hand side of formula four and find $\sum_\sigma \text{sgn}(\sigma)\alpha(v_{\sigma(1)}, v_{\sigma(2)})\beta(v_{\sigma(3)})$ where σ is a (2, 1)-shuffle. The (2, 1)-shuffles of 123 are 231, 132, and 123. Thus we have the three (2, 1)-shuffle permutations

$$\sigma_1 = \begin{pmatrix} 1 & 2 & 3 \\ \downarrow & \downarrow & \downarrow \\ 2 & 3 & 1 \end{pmatrix}, \quad \sigma_2 = \begin{pmatrix} 1 & 2 & 3 \\ \downarrow & \downarrow & \downarrow \\ 1 & 3 & 2 \end{pmatrix}, \quad \sigma_3 = \begin{pmatrix} 1 & 2 & 3 \\ \downarrow & \downarrow & \downarrow \\ 1 & 2 & 3 \end{pmatrix}.$$

It is easy to show that $\text{sgn}(\sigma_1)$ and $\text{sgn}(\sigma_3)$ are positive and $\text{sgn}(\sigma_2)$ is negative, hence

$$\begin{aligned}
&\alpha \wedge \beta(v_1, v_2, v_3) \\
&= \sum \text{sgn}(\sigma)(dx_2 \wedge dx_3)(v_{\sigma(1)}, v_{\sigma(2)})dx_5(v_{\sigma(3)}) \\
&= (dx_2 \wedge dx_3)(v_2, v_3)dx_5(v_1) - (dx_2 \wedge dx_3)(v_1, v_3)dx_5(v_2) \\
&\quad + (dx_2 \wedge dx_3)(v_1, v_2)dx_5(v_3) \\
&= \begin{vmatrix} dx_2(v_2) & dx_2(v_3) \\ dx_3(v_2) & dx_3(v_3) \end{vmatrix} dx_5(v_1) - \begin{vmatrix} dx_2(v_1) & dx_2(v_3) \\ dx_3(v_1) & dx_3(v_3) \end{vmatrix} dx_5(v_2) \\
&\quad + \begin{vmatrix} dx_2(v_1) & dx_2(v_2) \\ dx_3(v_1) & dx_3(v_2) \end{vmatrix} dx_5(v_3) \\
&= dx_2(v_2)dx_3(v_3)dx_5(v_1) - dx_2(v_3)dx_3(v_2)dx_5(v_1) \\
&\quad - dx_2(v_1)dx_3(v_3)dx_5(v_2) + dx_2(v_3)dx_3(v_1)dx_5(v_2) \\
&\quad + dx_2(v_1)dx_3(v_2)dx_5(v_3) + dx_2(v_2)dx_3(v_1)dx_5(v_3) \\
&= \big(dx_5dx_2dx_3 - dx_5dx_3dx_2 - dx_2dx_5dx_3 \\
&\quad + dx_3dx_5dx_2 + dx_2dx_3dx_5 - dx_3dx_2dx_5\big)(v_1, v_2, v_3).
\end{aligned}$$

Upon rearrangement we see that indeed, $\alpha \wedge \beta(v_1, v_2, v_3)$ found using the volume/determinant-based definition of wedgeproduct is exactly the same as what we found when we used the right hand side of formula four, $\sum_\sigma \text{sgn}(\sigma)\alpha(v_{\sigma(1)}, v_{\sigma(2)})\beta(v_{\sigma(3)})$ where σ is a (2, 1)-shuffle. Thus, indeed, $\alpha \wedge \beta(v_1, v_2, v_3) = \sum_\sigma \text{sgn}(\sigma)\alpha(v_{\sigma(1)}, v_{\sigma(2)}) \beta(v_{\sigma(3)})$ where σ is a (2, 1)-shuffle

Now we turn our attention to the general case. The procedure we use will mirror that of the example. We want to show this equality for a k-form $\alpha \in \bigwedge^k(\mathbb{R}^n)$ and ℓ-form $\beta \in \bigwedge^\ell(\mathbb{R}^n)$. In general we have $\alpha = \sum_I a_I dx^I$ and $\beta = \sum_J b_J dx^J$, where

$$\begin{aligned}
I &\in \mathfrak{I}_{k,n} = \big\{i_1 i_2 \cdots i_k \;\big|\; 1 \le i_1 < i_2 < \cdots < i_k \le n\big\}, \\
J &\in \mathfrak{J}_{\ell,n} = \big\{j_1 j_2 \cdots j_\ell \;\big|\; 1 \le j_1 < j_2 < \cdots < j_\ell \le n\big\}.
\end{aligned}$$

The first thing to notice is that our definition of the wedgeproduct based on the volume/determinant only applies to elements of the form $dx_{i_1} \wedge dx_{i_2} \wedge \cdots \wedge dx_{i_m}$, which are the basis elements of $\bigwedge^m(\mathbb{R}^n)$. If we had a general k-form such as $\alpha = \sum_I a_I dx^I$ and wanted to find $\alpha(v_1, \dots, v_k)$ using this definition we first have to find the value of $dx^I(v_1, \dots, v_k) = dx_{i_1} \wedge \cdots \wedge dx_{i_k}(v_1, \dots, v_k)$ for each I by taking the determinant of the projected vectors, multiply what we found by the coefficient a_I, and then added all the terms up. Similarly for β.

Taking the wedgeproduct of α and β and using formula one proved in the last section we have

$$\begin{aligned}
&\alpha \wedge \beta \\
&= \left(\sum_I a_I dx^I\right) \wedge \left(\sum_J b_J dx^J\right) \\
&= \sum_{I,J} a_I b_J dx^I \wedge dx^J \\
&= \sum_{I,J} a_I b_J dx_{i_1} \wedge \cdots \wedge dx_{i_k} \wedge dx_{j_1} \wedge dx_{j_\ell}.
\end{aligned}$$

So using our volume/determinant definition of wedgeproduct we have

$$
\begin{aligned}
&\alpha \wedge \beta(v_1, \cdots v_{k+\ell}) \\
&= \sum_{I,J} a_I b_J dx_{i_1} \wedge \cdots \wedge dx_{i_k} \wedge dx_{j_1} \wedge dx_{j_\ell}(v_1, \cdots, v_{k+\ell}) \\
&= \sum_{\mathbf{I}} a_{\mathbf{I}} dx_{i_1} \wedge \cdots \wedge dx_{i_k} \wedge dx_{i_{k+1}} \wedge \cdots \wedge dx_{i_{k+\ell}}(v_1, \cdots, v_{k+\ell}) \\
&= \sum_{\mathbf{I}} a_{\mathbf{I}} \sum_{\sigma \in S_{k+\ell}} \operatorname{sgn}(\sigma) \prod_{m=1}^{k+\ell} dx_{i_{\sigma(m)}}(v_m),
\end{aligned}
$$

where we make a notation change to $\mathbf{I} = I \cup J$ in the second equality to make our indice labels line up nicely. Hence $j_1 = i_{k+1}, \ldots j_\ell = i_{k+\ell}$ and the $a_{\mathbf{I}}$ term is the appropriate $a_I b_J$ term; that is, $a_{i_1 \cdots i_{k+\ell}} = a_{i_1 \cdots i_k} b_{j_1 \cdots j_\ell}$.

Now, using our (k, ℓ)-shuffle formula we have

$$
\begin{aligned}
&\sum_{\substack{\sigma \text{ a} \\ (k,\ell)- \\ \text{shuffle}}} \operatorname{sgn}(\sigma) \alpha(v_{\sigma(1)}, \cdots v_{\sigma(k)}) \beta(v_{\sigma(k+1)}, \cdots v_{\sigma(k+\ell)}) \\
&= \sum_{\substack{\sigma \text{ a} \\ (k,\ell)- \\ \text{shuffle}}} \operatorname{sgn}(\sigma) \left(\sum_I a_I dx^I \right) (v_{\sigma(1)}, \cdots, v_{\sigma(k)}) \left(\sum_J b_J dx^J \right) (v_{\sigma(k+1)}, \cdots, v_{\sigma(k+\ell)}) \\
&= \sum_{\substack{\sigma \text{ a} \\ (k,\ell)- \\ \text{shuffle}}} \operatorname{sgn}(\sigma) \sum_{I,J} a_I b_J dx^I (v_{\sigma(1)}, \cdots, v_{\sigma(k)}) dx^J (v_{\sigma(k+1)}, \cdots, v_{\sigma(k+\ell)}) \\
&= \sum_{\mathbf{I}} a_{\mathbf{I}} \sum_{\substack{\sigma \text{ a} \\ (k,\ell)- \\ \text{shuffle}}} \operatorname{sgn}(\sigma) dx_{i_1} \wedge \cdots \wedge dx_{i_k}(v_{\sigma(1)}, \cdots, v_{\sigma(k)}) \underbrace{\cdot}_{\substack{\text{multiplication} \\ \text{not } \wedge}} dx_{i_{k+1}} \wedge \cdots \wedge dx_{i_{k+1}}(v_{\sigma(k+1)}, \cdots, v_{\sigma(k+\ell)})
\end{aligned}
$$

where in the last equality we made the same notation changes as before and the order of summation is changed. Our goal is to see that this is identical to what the volume/determinant definition gave us in the preceding formula.

First, using the determinant definition of wedgeproduct we have

$$
dx_{i_1} \wedge \cdots \wedge dx_{i_k}(v_{\sigma(1)}, \cdots, v_{\sigma(k)}) = \sum_{\tilde{\sigma} \in S_k} \operatorname{sgn}(\tilde{\sigma}) \prod_{m=1}^{k} dx_{i_{\tilde{\sigma}(m)}}(v_{\sigma(m)})
$$

and

$$
dx_{i_{k+1}} \wedge \cdots \wedge dx_{i_{k+1}}(v_{\sigma(k+1)}, \cdots, v_{\sigma(k+\ell)}) = \sum_{\tilde{\tilde{\sigma}} \in S_\ell} \operatorname{sgn}(\tilde{\tilde{\sigma}}) \prod_{m=k+1}^{k+\ell} dx_{i_{\tilde{\tilde{\sigma}}(m)}}(v_{\sigma(m)}).
$$

Combining this we have

$$
\begin{aligned}
&\sum_{\mathbf{I}} a_{\mathbf{I}} \sum_{\substack{\sigma \text{ a} \\ (k,\ell)- \\ \text{shuffle}}} \operatorname{sgn}(\sigma) dx_{i_1} \wedge \cdots \wedge dx_{i_k}(v_{\sigma(1)}, \cdots, v_{\sigma(k)}) \underbrace{\cdot}_{\substack{\text{mult.} \\ \text{not } \wedge}} dx_{i_{k+1}} \wedge \cdots \wedge dx_{i_{k+1}}(v_{\sigma(k+1)}, \cdots, v_{\sigma(k+\ell)}) \\
&= \sum_{\mathbf{I}} a_{\mathbf{I}} \sum_{\substack{\sigma \text{ a} \\ (k,\ell)- \\ \text{shuffle}}} \operatorname{sgn}(\sigma) \left(\sum_{\tilde{\sigma} \in S_k} \operatorname{sgn}(\tilde{\sigma}) \prod_{m=1}^{k} dx_{i_{\tilde{\sigma}(m)}}(v_{\sigma(m)}) \right) \cdot \left(\sum_{\tilde{\tilde{\sigma}} \in S_\ell} \operatorname{sgn}(\tilde{\tilde{\sigma}}) \prod_{m=k+1}^{k+\ell} dx_{i_{\tilde{\tilde{\sigma}}(m)}}(v_{\sigma(m)}) \right)
\end{aligned}
$$

$$= \sum_{\mathbf{I}} a_{\mathbf{I}} \sum_{\substack{\sigma \text{ a} \\ (k,\ell)- \\ \text{shuffle}}} \sum_{\tilde{\sigma} \in S_k} \sum_{\tilde{\tilde{\sigma}} \in S_\ell} \text{sgn}(\sigma)\text{sgn}(\tilde{\sigma})\text{sgn}(\tilde{\tilde{\sigma}}) \left(\prod_{m=1}^{k} dx_{i_{\tilde{\sigma}(m)}}(v_{\sigma(m)}) \right) \left(\prod_{m=k+1}^{k+\ell} dx_{i_{\tilde{\tilde{\sigma}}(m)}}(v_{\sigma(m)}) \right)$$

$$= \sum_{\mathbf{I}} a_{\mathbf{I}} \sum_{\sigma \in S_{k+\ell}} \text{sgn}(\sigma) \prod_{m=1}^{k+\ell} dx_{i_{\sigma(m)}}(v_m),$$

which is exactly the same as $\alpha \wedge \beta(v_1, \cdots v_{k+\ell})$ given by volume/determinant definition. Hence, formula four is indeed a formula for $\alpha \wedge \beta(v_1, \cdots v_{k+\ell})$. Formula three then follows. A little thought may be required to convince yourself of the final equality.

Question 3.20 Explain the final equality above.

3.4 The Interior Product

We now introduce something called the **interior product**, or **inner product**, of a vector and a k-form. Given the vector v and a k-form α the interior product of v with α is denoted by $\iota_v \alpha$. The interior product of a k-form α with a vector v is a $(k-1)$-form defined by

$$\boxed{\text{Interior Product:} \qquad \iota_v \alpha(v_1, \ldots, v_{k-1}) = \alpha(v, v_1, \ldots, v_{k-1}).}$$

where $v_1, \ldots, v_{k-1}$ are any $k-1$ vectors. In other words, $\iota_v \alpha$ just puts the vector v into α's first slot. So, when α is a k-form then $\iota_v \alpha$ is a $(k-1)$-form.

We now point out one simple identity. If both α and β are k-forms then

$$\boxed{\iota_v \big(\alpha + \beta\big) = \iota_v \alpha + \iota_v \beta.}$$

This comes directly from how the sums of forms are evaluated,

$$\begin{aligned}
&\iota_v \big(\alpha + \beta\big)(v_1, \ldots, v_{k-1}) \\
&= \big(\alpha + \beta\big)(v, v_1, \ldots, v_{k-1}) \\
&= \alpha(v, v_1, \ldots, v_{k-1}) + \beta(v, v_1, \ldots, v_{k-1}) \\
&= \iota_v \alpha(v_1, \ldots, v_{k-1}) + \iota_v \beta(v_1, \ldots, v_{k-1}),
\end{aligned}$$

which gives us the identity. Now suppose α is a k-form and v, w are vectors. We also have

$$\boxed{\iota_{(v+w)} \alpha = \iota_v \alpha + \iota_w \alpha.}$$

Thus the interior product $\iota_v \alpha$ is linear in both v and α.

Question 3.21 If α is a k-form and v, w are vectors, show that $\iota_{(v+w)} \alpha = \iota_v \alpha + \iota_w \alpha$.

Let us see how this works for some simple examples. Let us begin by looking at two-forms on the manifold $\mathbb{R}^3$. First we will consider the most basic two-forms, $dx \wedge dy$, $dy \wedge dz$, and $dx \wedge dz$. For the moment we will use increasing notation instead of cyclic notation in order to stay consistent with the notation of the last section. Here this simply means we will write $dx \wedge dz$ instead of $dz \wedge dx$. Letting

$$v = \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix}$$

we get

$$\begin{aligned}
\iota_v\big(dx \wedge dy\big) &= dx \wedge dy(v, \cdot) \\
&= \begin{vmatrix} dx(v) & dx(\cdot) \\ dy(v) & dy(\cdot) \end{vmatrix} \\
&= dx(v)dy(\cdot) - dy(v)dx(\cdot) \\
&= v_1 dy - v_2 dx,
\end{aligned}$$

which is clearly a one-form. Similarly, for the second term we have

$$\begin{aligned}
\iota_v\big(dy \wedge dz\big) &= dy \wedge dz(v, \cdot) \\
&= \begin{vmatrix} dy(v) & dy(\cdot) \\ dz(v) & dz(\cdot) \end{vmatrix} \\
&= dy(v)dz(\cdot) - dz(v)dy(\cdot) \\
&= v_2 dz - v_3 dy.
\end{aligned}$$

And for the third term we have

$$\begin{aligned}
\iota_v\big(dx \wedge dz\big) &= dx \wedge dz(v, \cdot) \\
&= \begin{vmatrix} dx(v) & dx(\cdot) \\ dz(v) & dz(\cdot) \end{vmatrix} \\
&= dx(v)dz(\cdot) - dz(v)dx(\cdot) \\
&= v_1 dz - v_3 dx.
\end{aligned}$$

Question 3.22 Show that $\iota_v\big(f dx \wedge dy\big) = f \iota_v\big(dx \wedge dy\big)$.

Now suppose we had the more general two-form,

$$\begin{aligned}
\alpha &= f(x, y, z)dx \wedge dy + g(x, y, z)dy \wedge dz + h(x, y, z)dx \wedge dz \\
&= f dx \wedge dy + g dy \wedge dz + h dx \wedge dz.
\end{aligned}$$

Using the above we can write

$$\begin{aligned}
\iota_v \alpha &= \iota_v\big(f dx \wedge dy + g dy \wedge dz + h dx \wedge dz\big) \\
&= \iota_v\big(f dx \wedge dy\big) + \iota_v\big(g dy \wedge dz\big) + \iota_v\big(h dx \wedge dz\big) \\
&= f \iota_v\big(dx \wedge dy\big) + g \iota_v\big(dy \wedge dz\big) + h \iota_v\big(dx \wedge dz\big) \\
&= f\big(v_1 dy - v_2 dx\big) + g\big(v_2 dz - v_3 dy\big) + h\big(v_1 dz - v_3 dx\big) \\
&= -\big(f v_2 + h v_3\big)dx + \big(f v_1 - g v_3\big)dy + \big(g v_2 + h v_1\big)dz.
\end{aligned}$$

To get the second line we used the fact that $\iota_v\big(\alpha + \beta\big) = \iota_v \alpha + \iota_v \beta$ and the third line came from the above question. Clearly, simply writing down what $\iota_v \alpha$ is can be a fairly complicated endeavor. So we give the general formula

$$\boxed{\iota_v\big(dx_1 \wedge \cdots \wedge dx_k\big) = \sum_{i=1}^{k} (-1)^{i-1} dx_i(v)\big(dx_1 \wedge \cdots \widehat{dx_i} \cdots \wedge dx_k\big),}$$

where $\widehat{dx_i}$ means that the element dx_i is omitted.

Question 3.23 Use the general formula for $\iota_v \alpha$ to find $\iota_v\big(dx \wedge dy\big)$, $\iota_v\big(dy \wedge dz\big)$, and $\iota_v\big(dx \wedge dz\big)$.

Question 3.24 (Requires some linear algebra) Using the determinant definition of the wedgeproduct,

$$dx_1 \wedge dx_2 \wedge \cdots \wedge dx_k\big(v_1, v_2, \ldots v_k\big) = \sum_{\sigma \in S_k} \operatorname{sgn}(\sigma) \prod_{i=1}^{k} dx_{\sigma(i)}(v_i)$$

deduce the formula

$$\iota_v\big(dx_1 \wedge \cdots \wedge dx_k\big) = \sum_{i=1}^{k} (-1)^{i-1} dx_i(v)\big(dx_1 \wedge \cdots \widehat{dx_i} \cdots \wedge dx_k\big)$$

by expanding the determinant in terms of the elements of the first column and their cofactors.

We now turn toward showing two important identities. If α is a k-form and β is any form then

$$\boxed{\iota_v\big(\alpha \wedge \beta\big) = \big(\iota_v \alpha\big) \wedge \beta + (-1)^k \alpha \wedge \big(\iota_v \beta\big).}$$

We first show this for the basis elements and then use the linearity of the interior product to show it is true for arbitrary forms. Suppose $\alpha = dx_1 \wedge \cdots \wedge dx_k$ and $\beta = dy_1 \wedge \cdots \wedge dy_q$. Then we have

$$\begin{aligned}
\iota_v(\alpha \wedge \beta) &= \iota_v\big(dx_1 \wedge \cdots \wedge dx_k \wedge dy_1 \wedge \cdots \wedge dy_q\big) \\
&= \sum_{i=1}^{k} (-1)^{i-1} dx_i(v)\big(dx_1 \wedge \cdots \widehat{dx_i} \cdots \wedge dx_k \wedge dy_1 \wedge \cdots \wedge dy_q\big) \\
&\quad + \sum_{j=1}^{q} (-1)^{k+j-1} dy_j(v)\big(dx_1 \wedge \cdots \wedge dx_p \wedge dy_1 \wedge \cdots \widehat{dy_j} \cdots \wedge dy_q\big) \\
&= \big(\iota_v \alpha\big) \wedge \beta + (-1)^k \alpha \wedge \big(\iota_v \beta\big).
\end{aligned}$$

Question 3.25 Suppose we have the general k-form $\alpha = \sum \alpha_{i_1 \cdots i_k} dx_{i_1} \wedge \cdots \wedge dx_{i_k}$ and ℓ-from $\beta = \sum \beta_{j_1 \cdots j_\ell} dy_{j_1} \wedge \cdots \wedge dy_{j_\ell}$. Use the linearity of the interior product to show the identity $\iota_v\big(\alpha \wedge \beta\big) = \big(\iota_v \alpha\big) \wedge \beta + (-1)^k \alpha \wedge \big(\iota_v \beta\big)$ applies to general forms.

Next we prove

$$\boxed{\big(\iota_u \iota_v + \iota_v \iota_u\big)\alpha = 0,}$$

which is often simply written as $\iota_v \iota_w + \iota_w \iota_v = 0$. Again we begin by supposing $\alpha = dx_1 \wedge \cdots \wedge dx_k$. We start by finding $\iota_u \iota_v \alpha$,

$$\begin{aligned}
\iota_u \iota_v \alpha &= \iota_u \iota_v\big(dx_1 \wedge \cdots \wedge dx_k\big) \\
&= \sum_{i=1}^{k} (-1)^{i-1} dx_i(v) \iota_u\big(dx_1 \wedge \cdots \widehat{dx_i} \cdots \wedge dx_k\big) \\
&= \sum_{i=1}^{k-1} (-1)^{i-1} dx_i(v) \sum_{j=i+1}^{k} dx_j(u)\big(dx_1 \wedge \cdots \widehat{dx_i} \cdots \widehat{dx_j} \cdots \wedge dx_k\big) \\
&\quad + \sum_{i=2}^{k} (-1)^{i-1} dx_i(v) \sum_{j=1}^{i-1} dx_j(u)\big(dx_1 \wedge \cdots \widehat{dx_j} \cdots \widehat{dx_i} \cdots \wedge dx_k\big).
\end{aligned}$$

Since i and j are just dummy indices we can switch them in the second double summation and we get

$$\begin{aligned}\iota_u \iota_v \alpha &= \sum_{i=1}^{k-1} \sum_{j=i+1}^{k} (-1)^{i+j-3} dx_i(v) dx_j(u) \big(dx_1 \wedge \cdots \widehat{dx_i} \cdots \widehat{dx_j} \cdots \wedge dx_k \big) \\ &\quad + \sum_{j=2}^{k} \sum_{i=1}^{j-1} (-1)^{i+j-2} dx_i(u) dx_j(v) \big(dx_1 \wedge \cdots \widehat{dx_i} \cdots \widehat{dx_j} \cdots \wedge dx_k \big).\end{aligned}$$

To get the second term $\iota_v \iota_u \alpha$ we simply interchange the u and v.

Question 3.26 Using $\iota_u \iota_v \alpha$ and $\iota_v \iota_u \alpha$ found above, show that $(\iota_u \iota_v + \iota_v \iota_u)\alpha = 0$.

Question 3.27 Using the above identity, show that $\iota_v^2 = 0$.

3.5 Summary, References, and Problems

3.5.1 Summary

A way to "multiply" together one-forms in a way that gives a very precise geometric meaning is introduced. This "multiplication" product is called a wedgeproduct. In words, the wedgeproduct of two one-forms $dx_i \wedge dx_j$ eats two vectors v_p and w_p and finds the two-dimensional volume of the parallelepiped spanned by the projections of those two vectors onto the $\partial_{x_i} \partial_{x_j}$-plane in $T_p M$. First the projections of v_p and w_p onto the $\partial_{x_i} \partial_{x_j}$-plane are found to be

$$\begin{bmatrix} dx_i(v_p) \\ dx_j(v_p) \end{bmatrix} = \begin{bmatrix} v_i \\ v_j \end{bmatrix} \qquad \text{and} \qquad \begin{bmatrix} dx_i(w_p) \\ dx_j(w_p) \end{bmatrix} = \begin{bmatrix} w_i \\ w_j \end{bmatrix}.$$

Recalling that the determinate of a matrix finds the volume of the parallelepiped spanned by the columns of a matrix we take the determinant of the 2×2 matrix with these vectors as its columns,

$$dx_i \wedge dx_j(v_p, w_p) = \begin{vmatrix} dx_i(v_p) & dx_i(w_p) \\ dx_j(v_p) & dx_j(w_p) \end{vmatrix}.$$

A two-form is defined to be a linear combination of elements of the form $dx_i \wedge dx_j$ and the space of two-forms on M at the point p is denoted by $\bigwedge_p^2(\mathbb{R}^2)$. The geometric meaning and formula for the wedgeproduct of n one-forms is defined analogously,

$$\boxed{\begin{array}{c}\text{Wedgeproduct}\\ \text{of } n\\ \text{one-forms}\end{array} \quad dx_{i_1} \wedge dx_{i_2} \wedge \cdots \wedge dx_{i_n}\big(v_1, v_2, \ldots v_n\big) \equiv \begin{vmatrix} dx_{i_1}(v_1) & dx_{i_1}(v_2) & \cdots & dx_{i_1}(v_n) \\ dx_{i_2}(v_1) & dx_{i_2}(v_2) & \cdots & dx_{i_2}(v_n) \\ \vdots & \vdots & \ddots & \vdots \\ dx_{i_n}(v_1) & dx_{i_n}(v_2) & \cdots & dx_{i_n}(v_n) \end{vmatrix}.}$$

An n-form is defined to be a linear combination of elements of the form $dx_{i_1} \wedge dx_{i_2} \wedge \cdots \wedge dx_{i_n}$ and the space of n-forms on M at point p is denoted by $\bigwedge_p^n(M)$. Using the definition of the determinant the wedgeproduct of n one-forms can also be written as

$$\boxed{\begin{array}{c}\text{Wedgeproduct}\\ \text{of } n\\ \text{one-forms}\end{array} \quad dx_{i_1} \wedge dx_{i_2} \wedge \cdots \wedge dx_{i_n}\big(v_1, v_2, \ldots v_n\big) = \sum_{\sigma \in S_n} \operatorname{sgn}(\sigma) \prod_{j=1}^{n} dx_{\sigma(i_j)}(v_j).}$$

A number of different formulas for the wedgeproduct of an arbitrary k-form and an ℓ-form were derived:

Wedgeproduct of two arbitrary forms, Formula One
$$\Big(\sum_I a_I dx^I\Big) \wedge \Big(\sum_J b_J dx^J\Big) = \sum a_I b_J dx^I \wedge dx^J,$$

Wedgeproduct of two arbitrary forms, Formula Two
$$\Big(\sum_I a_I dx^I\Big) \wedge \Big(\sum_J b_J dx^J\Big) = \sum_K \left(\sum_{\substack{I,J \\ I \cup J \text{ disjoint}}} \pm a_I b_J\right) dx^K,$$

Wedgeproduct of two arbitrary forms, Formula Three
$$(\alpha \wedge \beta)\big(v_1, \ldots, v_{k+\ell}\big) = \frac{1}{k!\ell!} \sum_{\sigma \in S_{k+\ell}} \operatorname{sgn}(\sigma) \alpha\big(v_{\sigma(1)}, \ldots, v_{\sigma(k)}\big) \beta\big(v_{\sigma(k+1)}, \ldots, v_{\sigma(k+\ell)}\big),$$

Wedgeproduct of two arbitrary forms, Formula Four
$$(\alpha \wedge \beta)\big(v_1, \ldots, v_{k+\ell}\big) = \sum_{\substack{\sigma \text{ is a} \\ (k+\ell)- \\ \text{shuffle}}} \operatorname{sgn}(\sigma) \alpha\big(v_{\sigma(1)}, \ldots, v_{\sigma(k)}\big) \beta\big(v_{\sigma(k+1)}, \ldots, v_{\sigma(k+\ell)}\big),$$

Wedgeproduct of two arbitrary forms, Formula Five
$$\alpha \wedge \beta = \frac{(k+\ell)!}{k!\ell!} \mathscr{A}(\alpha \otimes \beta).$$

Properties of the wedgeproduct were also derived. If $\omega, \omega_1, \omega_2$ are k-forms and η, η_1, η_2 are ℓ-forms then:

(1) $a\omega \wedge \eta = \omega \wedge a\eta = a\big(\omega \wedge \eta\big)$ where $a \in \mathbb{R}$,
(2) $\big(\omega_1 + \omega_2\big) \wedge \eta = \omega_1 \wedge \eta + \omega_2 \wedge \eta$,
(3) $\omega \wedge \big(\eta_1 + \eta_2\big) = \omega \wedge \eta_1 + \omega \wedge \eta_2$,
(4) $\omega \wedge \eta = (-1)^{k\ell} \eta \wedge \omega$.

The interior product of a k-form α and a vector v was defined to be

Interior Product:
$$\iota_v \alpha(v_1, \ldots, v_{k-1}) = \alpha(v, v_1, \ldots, v_{k-1}).$$

And a number of interior product identities were found,

$$\boxed{\iota_v\big(\alpha + \beta\big) = \iota_v \alpha + \iota_v \beta,}$$

$$\boxed{\iota_{(v+w)} \alpha = \iota_v \alpha + \iota_w \alpha,}$$

$$\boxed{\iota_v\big(dx_1 \wedge \cdots \wedge dx_k\big) = \sum_{i=1}^{k} (-1)^{i-1} dx_i(v) \big(dx_1 \wedge \cdots \widehat{dx_i} \cdots \wedge dx_k\big),}$$

$$\boxed{\iota_v\big(\alpha \wedge \beta\big) = \big(\iota_v \alpha\big) \wedge \beta + (-1)^k \alpha \wedge \big(\iota_v \beta\big),}$$

$$\boxed{\big(\iota_u \iota_v + \iota_v \iota_u\big)\alpha = 0.}$$

3.5.2 References and Further Reading

The first two sections of this chapter are a slow and careful introduction to the wedgeproduct from a geometric point of view. As much as possible we have attempted to focus on the meanings behind the fairly complicated formula that defines the wedgeproduct of two differential forms. The general approach taken in the first few sections of this chapter generally follows Bachman [4], though Edwards [18] also uses this approach as well. But very often the opposite approaches taken, the geometric meaning of the wedgeproduct is deduced, or at least alluded to, based on the formula; see for example Darling [12], Arnold [3], and even Spivak [41]. In Spivak what we have referred to as the "scaling factors" for general k-forms are called "signed scalar densities," which is another good way of thinking about the scaling factors.

In section three we have attempted to clearly connect the geometrically motivated formula for the wedgeproduct with the various other formulas that are often used to define the wedgeproduct. In reality, most books take a much more formal approach to defining the wedgeproduct than we do. See for example Tu [46], Munkres [35], Martin [33], or just about any other book on manifold theory or differential geometry. In a sense this is understandable, in these books differential forms are simply one of many topics that need to be covered, and so a much briefer approach is needed. It is our hope that this section will provide the link to help readers of those books gain a deeper understanding of what the wedgeproduct actually is.

3.5.3 Problems

Question 3.28 Which of the following make sense? If the expression makes sense then evaluate it.

$$a)\ dx_1 \wedge dx_2 \left(\begin{bmatrix} 3 \\ -2 \\ 1 \end{bmatrix}, \begin{bmatrix} 5 \\ 3 \\ -4 \end{bmatrix} \right) \qquad d)\ dx_1 \wedge dx_2 \wedge dx_4 \left(\begin{bmatrix} 6 \\ 1 \\ -3 \end{bmatrix}, \begin{bmatrix} 5 \\ -2 \\ -1 \end{bmatrix}, \begin{bmatrix} 5 \\ -2 \\ -1 \end{bmatrix} \right)$$

$$b)\ dx_1 \wedge dx_3 \left(\begin{bmatrix} -4 \\ -2 \end{bmatrix}, \begin{bmatrix} -3 \\ 4 \end{bmatrix} \right) \qquad e)\ dx_2 \wedge dx_3 \wedge dx_4 \left(\begin{bmatrix} -1 \\ 4 \\ 5 \\ -2 \end{bmatrix}, \begin{bmatrix} 3 \\ -1 \\ 1 \\ -3 \end{bmatrix}, \begin{bmatrix} -2 \\ 3 \\ 4 \\ 1 \end{bmatrix} \right)$$

$$c)\ dx_1 \wedge dx_4 \left(\begin{bmatrix} 5 \\ -2 \\ 3 \\ -1 \end{bmatrix}, \begin{bmatrix} 1 \\ 2 \\ -3 \\ -4 \end{bmatrix} \right) \qquad f)\ (3\, dx_2 \wedge dx_4 - 2\, dx_1 \wedge dx_3) \left(\begin{bmatrix} -1 \\ 4 \\ 5 \\ -2 \end{bmatrix}, \begin{bmatrix} 3 \\ -1 \\ 1 \\ -3 \end{bmatrix} \right)$$

Question 3.29 Compute the following numbers. Then explain in words what the number represents.

$$a)\ dx_1 \wedge dx_2 \left(\begin{bmatrix} 2 \\ 3 \\ -5 \\ 2 \end{bmatrix}, \begin{bmatrix} 2 \\ 3 \\ 4 \\ -5 \end{bmatrix} \right) \qquad d)\ dx_2 \wedge dx_3 \left(\begin{bmatrix} 2 \\ 3 \\ -5 \\ 2 \end{bmatrix}, \begin{bmatrix} 2 \\ 3 \\ 4 \\ -5 \end{bmatrix} \right)$$

$$b)\ dx_1 \wedge dx_3 \left(\begin{bmatrix} 2 \\ 3 \\ -5 \\ 2 \end{bmatrix}, \begin{bmatrix} 2 \\ 3 \\ 4 \\ -5 \end{bmatrix} \right) \qquad e)\ dx_2 \wedge dx_4 \left(\begin{bmatrix} 2 \\ 3 \\ -5 \\ 2 \end{bmatrix}, \begin{bmatrix} 2 \\ 3 \\ 4 \\ -5 \end{bmatrix} \right)$$

$$c)\ dx_1 \wedge dx_4 \left(\begin{bmatrix} 2 \\ 3 \\ -5 \\ 2 \end{bmatrix}, \begin{bmatrix} 2 \\ 3 \\ 4 \\ -5 \end{bmatrix} \right) \qquad f)\ dx_3 \wedge dx_4 \left(\begin{bmatrix} 2 \\ 3 \\ -5 \\ 2 \end{bmatrix}, \begin{bmatrix} 2 \\ 3 \\ 4 \\ -5 \end{bmatrix} \right)$$

Question 3.30 Using the results you obtained in problem 3.29 compute the following numbers. Then explain in words what the number represents.

$$a)\ 3\,dx_2 \wedge dx_3 \left(\begin{bmatrix} 2 \\ 3 \\ -5 \\ 2 \end{bmatrix}, \begin{bmatrix} 2 \\ 3 \\ 4 \\ -5 \end{bmatrix} \right) \qquad d)\ \left(-2\,dx_2 \wedge dx_3 + \frac{1}{4}\,dx_3 \wedge dx_4 \right) \left(\begin{bmatrix} 2 \\ 3 \\ -5 \\ 2 \end{bmatrix}, \begin{bmatrix} 2 \\ 3 \\ 4 \\ -5 \end{bmatrix} \right)$$

$$b)\ -2\,dx_1 \wedge dx_4 \left(\begin{bmatrix} 2 \\ 3 \\ -5 \\ 2 \end{bmatrix}, \begin{bmatrix} 2 \\ 3 \\ 4 \\ -5 \end{bmatrix} \right) \qquad e)\ (4\,dx_1 \wedge dx_2 - 3\,dx_3 \wedge dx_4) \left(\begin{bmatrix} 2 \\ 3 \\ -5 \\ 2 \end{bmatrix}, \begin{bmatrix} 2 \\ 3 \\ 4 \\ -5 \end{bmatrix} \right)$$

$$c)\ \frac{1}{2}\,dx_2 \wedge dx_4 \left(\begin{bmatrix} 2 \\ 3 \\ -5 \\ 2 \end{bmatrix}, \begin{bmatrix} 2 \\ 3 \\ 4 \\ -5 \end{bmatrix} \right) \qquad f)\ \left(\frac{4}{3}\,dx_1 \wedge dx_4 + 3\,dx_2 \wedge dx_4 \right) \left(\begin{bmatrix} 2 \\ 3 \\ -5 \\ 2 \end{bmatrix}, \begin{bmatrix} 2 \\ 3 \\ 4 \\ -5 \end{bmatrix} \right)$$

Question 3.31 Compute the following functions for the given forms on manifold $\mathbb{R}^n$ and constant vector fields. Then explain in words what the function represents.

$$a)\ xyz\,dz \wedge dx \left(\begin{bmatrix} 4 \\ -2 \\ -3 \end{bmatrix}, \begin{bmatrix} -3 \\ 2 \\ 2 \end{bmatrix} \right) \qquad d)\ \left(x\,dx \wedge dy + \sin(z)\,dy \wedge dz \right) \left(\begin{bmatrix} 1 \\ -2 \\ -2 \end{bmatrix}, \begin{bmatrix} 3 \\ 1 \\ 2 \end{bmatrix} \right)$$

$$b)\ y^z \sin(x)\,dy \wedge dz \left(\begin{bmatrix} 4 \\ -2 \\ 1 \end{bmatrix}, \begin{bmatrix} 3 \\ -2 \\ -1 \end{bmatrix} \right) \qquad e)\ \left(e^{xy}\,dx_1 \wedge dx_3 + \sqrt{|z|}\,dx_1 \wedge dx_4 \right) \left(\begin{bmatrix} 3 \\ 2 \\ -2 \\ 5 \end{bmatrix}, \begin{bmatrix} -4 \\ -3 \\ 2 \\ -1 \end{bmatrix} \right)$$

$$c)\ \left(x_1 + x_2^{x_4} \right) dx_2 \wedge dx_3 \left(\begin{bmatrix} -4 \\ 3 \\ -2 \\ 6 \end{bmatrix}, \begin{bmatrix} -1 \\ 5 \\ 4 \\ -5 \end{bmatrix} \right) \qquad f)\ e_1^x \sqrt{|x_3|}\,dx_2 \wedge dx_3 \wedge dx_4 \left(\begin{bmatrix} 3 \\ 4 \\ -4 \\ -2 \end{bmatrix}, \begin{bmatrix} 3 \\ 2 \\ -2 \\ -1 \end{bmatrix}, \begin{bmatrix} 1 \\ -3 \\ 2 \\ 4 \end{bmatrix} \right)$$

Question 3.32 Compute the following functions for the given forms on manifold $\mathbb{R}^n$ and the non-constant vector fields. Then explain in words what the function represents.

$$a)\ xyz\,dz \wedge dx \left(\begin{bmatrix} 3x \\ -y \\ -3 \end{bmatrix}, \begin{bmatrix} 2 \\ z \\ x \end{bmatrix} \right) \qquad d)\ \left(x\,dx \wedge dy + \sin(z)\,dy \wedge dz \right) \left(\begin{bmatrix} x^y \\ -z \\ -2 \end{bmatrix}, \begin{bmatrix} 1 \\ x^2 \\ z^3 \end{bmatrix} \right)$$

$$b)\ y^z \sin(x)\,dy \wedge dz \left(\begin{bmatrix} 2x \\ -y^z \\ z \end{bmatrix}, \begin{bmatrix} 3 \\ -x \\ -2y \end{bmatrix} \right) \qquad e)\ \left(e^{xy}\,dx_1 \wedge dx_3 + \sqrt{|z|}\,dx_1 \wedge dx_4 \right) \left(\begin{bmatrix} x_4 \\ x_2 + 3 \\ -2 \\ x_3 \end{bmatrix}, \begin{bmatrix} -x_2 \\ -3 \\ x_3^{x_1} \\ -1 \end{bmatrix} \right)$$

$$c)\ \left(x_1 + x_2^{x_4} \right) dx_2 \wedge dx_3 \left(\begin{bmatrix} -2x_2 \\ x_3 \\ x_4 - 2 \\ 6 \end{bmatrix}, \begin{bmatrix} 2 \\ x_2^4 \\ -x_4 \\ -5 \end{bmatrix} \right) \qquad f)\ e_1^x \sqrt{|x_3|}\,dx_2 \wedge dx_3 \wedge dx_4 \left(\begin{bmatrix} 1 \\ 2^{x_1} \\ 3 \\ -2x_3 \end{bmatrix}, \begin{bmatrix} x_3 \\ 2 \\ e^{x_4} \\ -1 \end{bmatrix}, \begin{bmatrix} 4 \\ x_3x_3 \\ x_3x_4 \\ x_4x_1 \end{bmatrix} \right)$$

Question 3.33 Given $\alpha = \cos(xz)\, dx \wedge dy$ a two-form on $\mathbb{R}^3$ and the constant vectors fields

$$v = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} \qquad \text{and} \qquad w = \begin{bmatrix} 2 \\ 2 \\ 3 \end{bmatrix}$$

on $\mathbb{R}^3$ find $\alpha(v, w)$. Then find $\alpha_p(w_p, v_p)$ for both points $(1, 2, \pi)$ and $\left(\frac{1}{2}, 2, \pi\right)$.

Question 3.34 Given $\beta = \sin(xz)\, dy \wedge dz$ a two-form on $\mathbb{R}^3$ and the constant vector fields v and w on $\mathbb{R}^3$ from problem 3.33, find $\beta(v, w)$. Then find $\beta_p(w_p, v_p)$ for both points $(1, 2, \pi)$ and $\left(\frac{1}{2}, 2, \pi\right)$.

Question 3.35 Given $\gamma = xyz\, dz \wedge dx$ a two-form on $\mathbb{R}^3$ and the constant vector fields v and w from problem 3.33, find $\gamma(v, w)$. Then find $\gamma_p(w_p, v_p)$ for both points $(1, 2, \pi)$ and $\left(\frac{1}{2}, 2, \pi\right)$.

Question 3.36 Given $\phi = \cos(xz)\, dx \wedge dy + \sin(xz)\, dy \wedge dz + xyz\, dz \wedge dx$ a two-form on $\mathbb{R}^3$ and the constant vector fields v and w from problem 3.33, find $\phi(v, w)$. Then find $\phi_p(w_p, v_p)$ for both points $(1, 2, \pi)$ and $\left(\frac{1}{2}, 2, \pi\right)$.

Question 3.37 Find the basis of $\bigwedge_p^2(\mathbb{R}^5)$. That is, list the elementary two-forms on $\mathbb{R}^5$. Then find the basis of $\bigwedge_p^3(\mathbb{R}^5)$, $\bigwedge_p^4(\mathbb{R}^5)$, and $\bigwedge_p^5(\mathbb{R}^5)$.

Question 3.38 Find the basis of $\bigwedge_p^1(\mathbb{R}^6)$, $\bigwedge_p^2(\mathbb{R}^6)$, $\bigwedge_p^3(\mathbb{R}^6)$, $\bigwedge_p^4(\mathbb{R}^6)$, $\bigwedge_p^5(\mathbb{R}^6)$, and $\bigwedge_p^6(\mathbb{R}^6)$.

Question 3.39 Using the algebraic properties of differential forms simplify the following expressions. Put indices in increasing order.

a) $\left(3\, dx + 2\, dy\right) \wedge dz$ *d)* $dz \wedge \left(4\, dx + 3\, dy\right)$ *g)* $(dx_1 \wedge dx_3) \wedge \left(3\, dx_2 - 4\, dx_4\right)$

b) $\left(z\, dx - x\, dy\right) \wedge dz$ *e)* $dx \wedge \left(6\, dy - z\, dz\right)$ *h)* $\left(x_3\, dx_2 \wedge dx_5 + x_1\, dx_4 \wedge dx_6\right) \wedge (-5\, dx_1 \wedge dx_3)$

c) $\left(x^y\, dz + 4z\, dx\right) \wedge (4\, dy)$ *f)* $-dy \wedge \left(e^z\, dz - x\, dx\right)$ *i)* $(2\, dx_3 \wedge dx_4) \wedge \left(x_4\, dx_1 \wedge dx_2 + e^{x_6}\, dx_1 \wedge dx_6\right)$

Question 3.40 Using the algebraic properties of differential forms simplify the following expressions. Put indices in increasing order.

a) $\left(3\, dx_1 \wedge dx_3 + 2\, dx_2 \wedge dx_4\right) \wedge \left(-\, dx_1 \wedge dx_2 + 3\, dx_2 \wedge dx_4 - 6\, dx_2 \wedge dx_3\right)$

b) $\left(-4\, dx_2 \wedge dx_4 \wedge dx_5\right) \wedge \left(3\, dx_1 \wedge dx_3 \wedge dx_7 + 5\, dx_1 \wedge dx_6 \wedge dx_7\right)$

c) $\left(x_2 x_5\, dx_3 \wedge dx_4 \wedge dx_8 + \sin(x_3)\, dx_3 \wedge dx_6 \wedge dx_8\right) \wedge \left(e^{x_7}\, dx_2 \wedge dx_4 \wedge dx_7\right)$

d) $\left(\sin(x_2)\, dx_2 \wedge dx_3 - e^{x_3} dx_5 \wedge dx_7\right) \wedge \left(x_2^{x_4}\, dx_1 \wedge dx_4 \wedge dx_6 + (x_3 + x_4)\, dx_4 \wedge dx_8 \wedge dx_9\right)$

Question 3.41 Let $v_1 = \begin{bmatrix} -2 \\ 3 \\ -1 \end{bmatrix}$, $v_2 = \begin{bmatrix} -5 \\ 6 \\ -4 \end{bmatrix}$, and $v_3 = \begin{bmatrix} 7 \\ -5 \\ -2 \end{bmatrix}$. Find

a) $\iota_{v_1}\left(3\, dx \wedge dy\right)$ *d)* $\iota_{v_1}\left(-6\, dx \wedge dy + 2\, dx \wedge dz\right)$ *g)* $\iota_{v_1}\left(x^y\, dx \wedge dz\right)$

b) $\iota_{v_2}\left(5\, dx \wedge dz\right)$ *e)* $\iota_{v_2}\left(3\, dx \wedge dz - 4\, dy \wedge dz\right)$ *h)* $\iota_{v_2}\left(y \sin(x) e^z\, dy \wedge dz\right)$

c) $\iota_{v_3}\left(-4\, dy \wedge dz\right)$ *f)* $\iota_{v_3}\left(2\, dx \wedge dy + 7\, dy \wedge dz\right)$ *i)* $\iota_{v_3}\left((x + y + z)\, dx \wedge dy\right)$

Question 3.42 Let $v_1 = e_1 - 2e_2 - 5e_3 + 5e_4 - 3e_5 + 6e_6$ and $v_2 = 4e_1 - 7e_2 - 3e_3 + 2e_4 + e_5 + 7e_6$. Find

$$a)\ \iota_{v_1}\Big(dx_1 \wedge dx_3 \wedge dx_5 \wedge dx_6\Big) \qquad c)\ \iota_{v_1}\Big(dx_1 \wedge dx_2 \wedge dx_4 \wedge dx_5 \wedge dx_6\Big)$$

$$b)\ \iota_{v_2}\Big(dx_2 \wedge dx_4 \wedge dx_5 \wedge dx_6\Big) \qquad d)\ \iota_{v_2}\Big(dx_1 \wedge dx_2 \wedge dx_3 \wedge dx_4 \wedge dx_6\Big)$$

Chapter 4
Exterior Differentiation

One of the central concepts of calculus, and in fact of all mathematics, is that of differentiation. One way to view differentiation is to view it as a mathematical object that measures how another mathematical object varies. For example, in one dimensional calculus the derivative of a function $f : \mathbb{R} \to \mathbb{R}$ is another function. In vector calculus you learned that the derivative of a transformation is the Jacobian matrix. You learned that both the divergence and curl are in some sense "derivatives" that measure how vector fields are varying. Now we want to consider the derivatives of forms.

From these examples you can see that the idea of differentiation can actually get rather complicated. Often the more complicated or abstract a space or object is, the more ways there are for the object to vary. This actually leads to different types of differentiations, each of which is useful in different circumstances. In this section we will introduce the most common type of differentiation for differential forms, the exterior derivative. There is also another way to define the derivative of a differential form, the Lie derivative of a form, which is introduced in Appendix A. Exterior differentiation is an extremely powerful concept, as we will come to see. Section one provides an overview of the four different approaches to exterior differentiation that books usually take. Since our goal in this chapter is to completely understand exterior differentiation each of these four approaches are then explored in detail in the following four sections. The chapter then concludes with a section devoted to some examples.

4.1 An Overview of the Exterior Derivative

In essence the basic objects that were studied in calculus were functions. You learned to take derivatives of functions and to integrate functions. In calculus on manifolds the basic objects are differential forms. We want to be able to take derivatives of differential forms and to integrate differential forms.

Derivatives of real-valued functions $f : \mathbb{R} \to \mathbb{R}$ and directional derivatives of multivariable real-valued functions $f : \mathbb{R}^n \to \mathbb{R}$ were both defined in terms of the limit of a difference quotient. This gave a very clean and precise geometrical meaning to the concept of the derivative of a function. Unfortunately, with the exterior derivative of differential form we do not have such a clean, precise, and easily understood geometrical meaning. Therefore, it is difficult to see right away why the definition of the exterior derivative of a differential form is the "right" definition to use.

In a sense, with exterior differentiation, we rely on the idea in the idiom "the proof is in the pudding." Basically, the end results of the definition of exterior derivative are just so amazingly "nice" that the definition has to be "right." What we mean is that by defining the exterior derivative of a differential form the way it is defined, we are able to perfectly generalize major ideas from vector calculus. For example, it turns out that the concepts of gradient, curl, and divergence from vector calculus all become special cases of exterior differentiation and the fundamental theorem of line integrals, Stokes theorem, and the divergence theorem all become special cases of what is called the generalized Stokes theorem. It allows us to summarize Maxwell's equations in electromagnetism into a single line, it gives us the Poncaré lemma, which is a powerful tool for the study of manifolds, and gives us the canonical symplectic form, which allows us to provide an abstract theoretical formulation of classical mechanics. We will study all of these consequences of the exterior derivative through the rest of this book. The amount of mathematics that works out perfectly because of this definition tells us that the definition has to be "right" even if it is difficult to see, geometrically, what the definition means.

In case all of this makes you uncomfortable, and feels a little circular to you, that is okay. In a sense it is. But you would be surprised by how often this sort of thing happens in mathematics. The definitions that prove the most useful are the ones

J. P. Fortney, *A Visual Introduction to Differential Forms and Calculus on Manifolds*,
https://doi.org/10.1007/978-3-319-96992-3_4

that survive to be taught in courses. Other, less useful, definitions eventually fall by the wayside. With all of this in mind we are now finally ready to take a look at exterior differentiation.

Most books introduce exterior differentiation in one of four different ways, and which approach a book's author takes depends largely on the flavor of the book.

1. The local ("in coordinates") formula is given. For example, if we have the one-form $\alpha = \sum f_i dx_i$ then the formula for the exterior derivative of α is simply given as

$$d\alpha = \sum df_i \wedge dx_i.$$

Notice that the coordinates that α is written in, here x_i, show up in the above formula. This is probably the most common approach taken in most introductory math and physics textbooks. From the formula the various properties are then derived. Frequently in these books the global formula is not given. The differential of a function df was defined in Definition 2.4.2 in Chap. 2 to be $df(v_p) = v_p[f] = \lim_{t\to 0} \frac{f(p+tv_p)-f(p)}{f}$ and thus df is related to our good old-fashioned directional derivative of f. Thus we can think of the exterior derivative as being, in some sense, a generalization of directional derivatives.
2. A list of the algebraic properties (or axioms) that the exterior derivative should have is given. These properties are then used to show that the exterior derivative is unique and to derive the (local and/or global) formula. Books that use this approach tend to be quite formal and axiomatic in nature, quite the antithesis of this book.
3. The global (also called invariant) formula is given. Most books that take this approach are quite advanced and most readers already have some knowledge of exterior differentiation. At this point in your education you have probably only ever seen the local ("in coordinates") form of any formula, so this concept is probably unfamiliar to you. The general idea is that you want to provide a formula but you do not want your formula to depend on which coordinates (Cartesian, polar, spherical, cylindrical, etc.) you are using at the time. In other words, you want your formula to be independent of the coordinates, or invariant. Without yet getting into what the formula means, if a one-form is written in-coordinates as $\alpha = \sum f_i dx_i$ (the x_i are the coordinates that are being used) then the in-coordinates formula for the exterior derivative of the one-form α is given by

$$d\alpha = \sum df_i \wedge dx_i.$$

The global formula for $d\alpha$, which is a two-form that acts on vectors v and w, is given by

$$d\alpha(v, w) = v\big[\alpha(w)\big] - w\big[\alpha(v)\big] - \alpha([v, w]).$$

Recall that $v[f]$ is one notation for the directional derivative of f in the direction of v. We also have that $[v, w]$ is the lie-bracket of two vector fields, which is defined by $[v, w] = vw - wv$. This will be explained later. Thus, the square brackets in the first two terms mean something different from the square brackets in the last term.

Notice that this formula is written entirely in terms of the one-form α and the vectors v and w. Nowhere in this formula do the particular coordinates that you are using (the x_i from the previous formula) show up. This type of formula is sometimes called a coordinate-free formula, a global formula, or an invariant formula. If α is an arbitrary k-form and $v_0, \ldots, v_k$ are $k+1$ vector fields then the global formula is given by

$$d\alpha(v_0, \ldots, v_k) = \sum_i (-1)^i v_i\big[\alpha(v_0, \ldots, \widehat{v_i}, \ldots, v_k)\big] + \sum_{i<j} (-1)^{i+j} \alpha\big([v_i, v_j], v_0, \ldots, \widehat{v_i}, \ldots, \widehat{v_j}, \ldots, v_k\big).$$

A hat over a vector means that the vector is omitted. Again, notice this formula is written entirely in terms of the k-form α and the vectors fields. It is also quite a complicated looking formula, not nearly as neat as the in-coordinates formula turns out to be.
4. A geometric definition in terms of the limit of the integral of the form over the boundary of a parallelepiped is given. This is a fairly infrequent approach that tends to show up in engineering or applied physics texts that want to emphasis the physical meaning of the exterior derivative. While this approach is really the most geometric of the approaches it either requires the book be structured so that integration is covered before differentiation or is necessarily somewhat imprecise.

The whole purpose of this book is to develop a deeper intuitive geometric understanding of what differential forms are, and since exterior differentiation is such a central concept we will attempt to look at it from all of these perspectives, trying to see how they relate to each other, as well as to actually spend some time looking at what is going on geometrically.

4.2 The Local Formula

We are taking things slowly and have simplified our lives considerably by only dealing with manifolds $\mathbb{R}^n$, and even then primarily with the cases of $n = 2$ or 3. In fact, we have even gone a step further in our simplification and only considered $\mathbb{R}^n$ with the standard Cartesian coordinates. These simplifications have allowed us to sidestep a large number of technical issues and to genuinely concentrate on developing an intuitive understanding of what differential forms are. We shall continue in this vein a bit longer, but will caution that by doing so we are also missing some of the richness and subtlety of the subject.

First we will recall what we did in chapter two. Given a function $f : \mathbb{R}^n \to \mathbb{R}$ on the manifold $\mathbb{R}^n$ we want to write the directional derivative of f in the direction of

$$v_p = \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix}_p$$

at the point $p = (x_{1_0}, x_{2_0}, \ldots, x_{n_0})$. Writing $p + tv_p$ as $p + tv_p = (x_{1_0} + tv_1, x_{2_0} + tv_2, \ldots, x_{n_0} + tv_n)$, and then writing the intermediate functions $x_1(t) = x_{1_0} + tv_1$ through $x_n(t) = x_{n_0} + tv_n$ we have

$$\begin{aligned}
df(v_p) &\equiv v_p[f] \\
&= \lim_{t\to 0} \frac{f\big(x_{1_0} + tv_1, x_{2_0} + tv_2, \cdots, x_{n_0} + tv_n\big) - f\big(x_{1_0}, x_{2_0}, \ldots, x_{n_0}\big)}{t} \\
&= \frac{d}{dt}\Big(f\big(p + tv_p\big)\Big)\Big|_{t=0} \\
&= \frac{d}{dt} f\big(\underbrace{x_{1_0} + tv_1}_{x_1(t)}, \underbrace{x_{2_0} + tv_2}_{x_2(t)}, \ldots, \underbrace{x_{n_0} + tv_n}_{x_n(t)}\big)\Big|_{t=0} \\
&= \frac{\partial f(x_1, \ldots, x_n)}{\partial x_1}\bigg|_p \cdot \frac{dx_1(t)}{dt}\bigg|_{t=0} + \frac{\partial f(x_1, \ldots, x_n)}{\partial x_2}\bigg|_p \cdot \frac{dx_2(t)}{dt}\bigg|_{t=0} + \cdots + \frac{\partial f(x_1, \ldots, x_n)}{\partial x_n}\bigg|_p \cdot \frac{dx_n(t)}{dt}\bigg|_{t=0} \\
&= \frac{\partial f}{\partial x_1}\bigg|_p \cdot v_1 + \frac{\partial f}{\partial x_2}\bigg|_p \cdot v_2 + \cdots + \frac{\partial f}{\partial x_n}\bigg|_p \cdot v_n \\
&= \sum_{i=1}^n v_i \cdot \frac{\partial f}{\partial x_i}\bigg|_p \\
&= \sum_{i=1}^n \frac{\partial f}{\partial x_i}\bigg|_p \, dx_i(v_p) \\
&= \left(\sum_{i=1}^n \frac{\partial f}{\partial x_i}\bigg|_p \, dx_i\right)(v_p)
\end{aligned}$$

and so, leaving off the base point, we have

$$df = \sum_{i=1}^n \frac{\partial f}{\partial x_i}\, dx_i.$$

Note, often you see $df(v_p)$ written as $df \cdot v_p$.

As an example, for $\mathbb{R}^2$, if $v_p = \begin{bmatrix} v_1 \\ v_2 \end{bmatrix}_p$ we found, recalling the use of the chain rule, that

$$\begin{aligned} df(v_p) &\equiv v_p[f] \\ &= \frac{\partial f}{\partial x} v_1 + \frac{\partial f}{\partial y} v_2 \\ &= \frac{\partial f}{\partial x} dx(v_p) + \frac{\partial f}{\partial y} dy(v_p) \\ &= \left(\frac{\partial f}{\partial x} dx + \frac{\partial f}{\partial y} dy \right)(v_p) \end{aligned}$$

so that we end up with

$$df = \frac{\partial f}{\partial x} dx + \frac{\partial f}{\partial y} dy.$$

Similarly, for $\mathbb{R}^3$ we have that

$$df = \frac{\partial f}{\partial x} dx + \frac{\partial f}{\partial y} dy + \frac{\partial f}{\partial z} dz.$$

Now we will introduce a little more terminology. If α is a one-form on the manifold M then $\alpha : T_pM \to \mathbb{R}$. That is, α "eats" a vector at each point p of the manifold M and "spits out" a real number for each point p of the manifold M. The space of one-forms is written as T_p^*M or $\bigwedge^1(M)$. If α is a two-form on the manifold M then $\alpha : T_pM \times T_pM \to \mathbb{R}$; it eats two vectors at each point p of the manifold M and spits out a real number for each point p of the manifold M. The space of two-forms is denoted by $\bigwedge^2(M)$. Similarly, if α is an n-form

$$\alpha : \underbrace{T_pM \times T_pM \times \cdots \times T_pM}_{n \text{ times}} \longrightarrow \mathbb{R}$$

eats n vectors at each point p of the manifold M. The space of n-forms is denoted by $\bigwedge^n(M)$.

In other words, we can think of a one-form as eating a point and a vector, a two-form as eating a point and two vectors, and an n-form at eating a point and n vectors. So, what would a zero-form eat? What would it spit out? It stands to reason that a zero-form would eat a point and zero vectors and spit out a real number for each point. And what sorts of things do that? Functions of course. Functions $f : M \to \mathbb{R}$ are things that we are very familiar with, and functions are nothing more than zero-forms. Functions are generally classified by how many times they can be differentiated on a domain, in this case the manifold M. Without getting into the details, we will generally assume our functions can be differentiated as many times as we like, an infinite number of times in fact. Thus we can denote our space of functions as $\mathscr{C}^\infty(M)$ or, if we are thinking of our functions as zero-forms, as $\bigwedge^0(M)$.

In the expression

$$df = \sum \frac{\partial f}{\partial x_i} dx_i$$

d is an **operator** that takes a function, a zero-form, and produces as output a one-form denoted df. The one-form df evaluated at a point p in some sense encodes the information about the tangent plane to f at the point p. In other words, df is the closest linear approximation to f at p. As a linear functional we should recognize that df is also a one-form. But notice, when the operator d is applied to a function and then paired with a vector that is exactly the old-fashioned directional derivative.

We want to extend this idea, we want to see how the operator d can take a one-form and produce as output a two-form, or take in an n-form and produce as output an $(n+1)$-form. In other words, we want an operator

$$d : \bigwedge^n(M) \longrightarrow \bigwedge^{n+1}(M).$$

This operator d will be called exterior differentiation. Since d when applied to a zero-form f is essentially a rewriting of the directional derivative then we can think of exterior differentiation as being a generalization of the directional derivatives from vector calculus. We will begin by simply giving the local (or "in-coordinates") formula for the exterior derivative.

Definition 4.2.1 Suppose f is a zero-form. Then the **exterior derivative of** f is defined by

$$df = \sum \frac{\partial f}{\partial x_i} dx_i.$$

Suppose $\alpha = \sum f_i dx_i$ is a one-form. Then the **exterior derivative of** α is defined by

$$d\alpha = \sum df_i \wedge dx_i.$$

Suppose $\omega = \sum f_{i_1 \cdots i_n} dx_{i_1} \wedge \cdots \wedge dx_{i_n}$ is an n-form. Then the **exterior derivative of** ω is defined by

$$d\omega = \sum df_{i_1 \cdots i_n} \wedge dx_{i_1} \wedge \cdots \wedge dx_{i_n}.$$

These three formulas are all very intuitive and nicely cover all the possible cases, but it is of course possible to combine them all into a single formula,

Exterior derivative of an n-form

$$\boxed{d\Big(\sum \alpha_{i_1 \cdots i_n}\, dx_{i_1} \wedge \cdots \wedge dx_{i_n}\Big) = \sum \sum_{j=1}^{n} \frac{\partial\, \alpha_{i_1 \cdots i_n}}{\partial x_{i_j}}\, dx_{i_j} \wedge dx_{i_1} \wedge \cdots \wedge dx_{i_n}.}$$

Each f_i in the definition of α and ω are functions, or zero-forms, and hence the df_i in the definition of the exterior derivatives $d\alpha$ and $d\omega$ are given by the definition of the exterior derivatives of a zero-form. Let us see how this works in a simple example. We will consider a one-form on the manifold $\mathbb{R}^2$. Suppose $\alpha = f_1 dx + f_2 dy$ is a one-form on the manifold $\mathbb{R}^2$ for some functions $f_1, f_2 : \mathbb{R}^2 \to \mathbb{R}$. Then

$$\begin{aligned}
d\alpha &= df_1 \wedge dx + df_2 \wedge dy \\
&= \left(\frac{\partial f_1}{\partial x} dx + \frac{\partial f_1}{\partial y} dy\right) \wedge dx + \left(\frac{\partial f_2}{\partial x} dx + \frac{\partial f_2}{\partial y} dy\right) \wedge dy \\
&= \frac{df_1}{dx} \underbrace{dx \wedge dx}_{=0} + \frac{df_1}{dy} \underbrace{dy \wedge dx}_{=-dx\wedge dy} + \frac{df_2}{dx} dx \wedge dy + \frac{df_2}{dy} \underbrace{dy \wedge dy}_{=0} \\
&= \left(\frac{df_2}{dx} - \frac{df_1}{dy}\right) dx \wedge dy.
\end{aligned}$$

Since α was a one-form then $d\alpha$ is a two-form. We go one step further to see how this two-form acts on the vectors

$$v = v_1 \frac{\partial}{\partial x} + v_2 \frac{\partial}{\partial y} = \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} \quad \text{and} \quad w = w_1 \frac{\partial}{\partial x} + w_2 \frac{\partial}{\partial y} = \begin{bmatrix} w_1 \\ w_2 \end{bmatrix}.$$

We have

$$\begin{aligned}
d\alpha(v, w) &= \left(\frac{df_2}{dx} - \frac{df_1}{dy}\right) dx \wedge dy(v, w) \\
&= \left(\frac{df_2}{dx} - \frac{df_1}{dy}\right) \begin{vmatrix} dx(v) & dx(w) \\ dy(v) & dy(w) \end{vmatrix} \\
&= \left(\frac{df_2}{dx} - \frac{df_1}{dy}\right) \begin{vmatrix} v_1 & w_1 \\ v_2 & w_2 \end{vmatrix}
\end{aligned}$$

$$
\begin{aligned}
&= \left(\frac{df_2}{dx} - \frac{df_1}{dy}\right)(v_1 w_2 - w_1 v_2) \\
&= v_1 w_2 \frac{df_2}{dx} - v_1 w_2 \frac{df_1}{dy} - w_1 v_2 \frac{df_2}{dx} + w_1 v_2 \frac{df_1}{dy}.
\end{aligned}
$$

We will return to this example later when we try to motivate the global formula for exterior differentiation.

Let us consider one more example. Let $\alpha = xy^3z^2\ dx + 5x^2y\ dz$ be a one-form on the manifold $\mathbb{R}^3$. We will find the exterior derivative of α,

$$
\begin{aligned}
d\alpha &= d\left(xy^3z^2\ dx + 5x^2y\ dz\right) \\
&= d\big(xy^3z^2\big) \wedge dx + \big(5x^2y\big) \wedge dz \\
&= \left(\frac{\partial\ xy^3z^2}{\partial x} dx + \frac{\partial\ xy^3z^2}{\partial y} dy + \frac{\partial\ xy^3z^2}{\partial z} dz\right) \wedge dx \\
&\quad + \left(\frac{\partial\ 5x^2y}{\partial x} dx + \frac{\partial\ 5x^2y}{\partial y} dy + \frac{\partial\ 5x^2y}{\partial z} dz\right) \wedge dz \\
&= \left(y^3z^2\ dx + 3xy^2z^2\ dy + 2xy^3z\ dz\right) \wedge dx \\
&\quad + \left(10xy\ dx + 5x^2\ dy + 0\ dz\right) \wedge dz \\
&= y^3z^2 \underbrace{dx \wedge dx}_{=0} + 3xy^2z^2 \underbrace{dy \wedge dx}_{-dx\wedge dy} + 2xy^3z\ dz \wedge dx \\
&\quad + 10xy \underbrace{dx \wedge dz}_{-dz\wedge dx} + 5x^2\ dy \wedge dz + 0 \underbrace{dz \wedge dz}_{=0} \\
&= -3xy^2z^2\ dx \wedge dy + 5x^2\ dy \wedge dz + \big(2xy^3z - 10xy\big)\ dz \wedge dx.
\end{aligned}
$$

We see that $d\alpha$ is a two-form.

Question 4.1 Find $dd\alpha$ for this example.

Question 4.2 Let $\alpha = 2x^2yz^3\ dx + 3yz^2\ dy + 7x^3y^2z^2\ dz$ be a one-form on the manifold $\mathbb{R}^3$. Find $d\alpha$ and $dd\alpha$.

When the exterior derivative is defined by just giving the formulas, as in this section, one generally proceeds to show that the exterior derivative operator d has a number of algebraic properties. One then needs to prove that the formulas remain unchanged under a change of coordinates. This is not at all difficult to do, but since we have not discussed changes of coordinates yet we will not do that here. Now suppose α, β are n-forms and ω is an m-form. Exterior differentiation satisfies the following three algebraic properties:

1. $d(\alpha + \beta) = d\alpha + d\beta$,
2. $d(\alpha \wedge \omega) = d\alpha \wedge \omega + (-1)^n \alpha \wedge d\omega$,
3. for each n-form α, $d(d\alpha) = 0$.

Showing these properties is not difficult, but the general case becomes somewhat notationally tedious.

Question 4.3 Using the notation from the last chapter, let $\alpha = \sum \alpha_I dx^I$, $\beta = \sum \beta_J dx^J$ both be n-forms and $\omega = \sum \omega_K dx^K$ be an m-form. Show the three algebraic properties above.

4.3 The Axioms of Exterior Differentiation

The second approach to introducing exterior differentiation one often encounters is to list the algebraic properties one wants exterior differentiation to have and then to show that such an operation exists and is unique.

Theorem 4.1 *There exists a unique operator*

$$d : \bigwedge^{n}(M) \longrightarrow \bigwedge^{n+1}(M).$$

called the ***exterior derivative*** *that satisfies the following four properties. Suppose α, β are n-forms and ω is an m-form on M, the operator d satisfies*

1. $d(\alpha + \beta) = d\alpha + d\beta$,
2. $d(\alpha \wedge \omega) = d\alpha \wedge \omega + (-1)^n \alpha \wedge d\omega$,
3. *for each n-form α, $d(d\alpha) = 0$,*
4. *in local coordinates, for each function f, $df = \sum \frac{\partial f}{\partial x_i} dx_i$.*

The first three of these properties are of course the properties that were listed at the end of the last section. Books that take this approach in introducing exterior differentiation essentially give these four properties and then use them to show that an operator d that satisfies them both exists and is unique. In the process of doing this they find the local formula for d. In essence they are simply going in the opposite direction as the last section. The last property is included because we want the operator d, when applied to a zero-form, to be exactly the same as the directional derivative of the zero-form function. By doing this we are essentially forcing our idea of exterior differentiation to be a generalization of the directional derivative. By making this specification, along with listing the properties we want d to have, it is fairly straight forward to derive the actual formula.

First of all we show existence of the operator d. To show existence we use the properties to find a formula for d. Once we know the formula for d we know it exists. And since we find only one formula for it then it must be unique as well. Let us now make explicit a notational convention that we have implicitly been using. If f is a zero-form (function) and α is an n-form then we generally write $f \wedge \alpha$ simply as $f\alpha$. This convention used in the third equality below. Suppose that we have the n-from

$$\alpha = \sum \alpha_{i_1 \cdots i_n} \, dx_{i_1} \wedge \cdots \wedge dx_{i_n}.$$

We apply d to α to get

$$\begin{aligned}
d\alpha &= d\Big(\sum \alpha_{i_1 \cdots i_n} \, dx_{i_1} \wedge \cdots \wedge dx_{i_n} \Big) \\
&= \sum d\Big(\alpha_{i_1 \cdots i_n} \, dx_{i_1} \wedge \cdots \wedge dx_{i_n} \Big) \\
&= \sum \Big(d\alpha_{i_1 \cdots i_n} \wedge dx_{i_1} \wedge \cdots \wedge dx_{i_n} + (-1)^0 \alpha_{i_1 \cdots i_n} \, d\big(dx_{i_1} \wedge \cdots \wedge dx_{i_n}\big) \Big) \\
&= \sum \Big(d\alpha_{i_1 \cdots i_n} \wedge dx_{i_1} \wedge \cdots \wedge dx_{i_n} + (-1)^0 \alpha_{i_1 \cdots i_n} \, \big(ddx_{i_1} \wedge \cdots \wedge dx_{i_n} + (-1)^1 dx_{i_1} \wedge d\big(dx_{i_2} \wedge \cdots \wedge dx_{i_n}\big)\big) \Big) \\
&\;\;\vdots \\
&= \sum \Bigg(d\alpha_{i_1 \cdots i_n} \wedge dx_{i_1} \wedge \cdots \wedge dx_{i_n} \\
&\qquad + (-1)^0 \alpha_{i_1 \cdots i_n} \Big(\underbrace{ddx_{i_1}}_{=0} \wedge \cdots \wedge dx_{i_n} + (-1)^1 dx_{i_1} \wedge \underbrace{ddx_{i_2}}_{=0} \wedge \cdots \wedge dx_{i_n} + \cdots + (-1)^{n-1} dx_{i_1} \wedge \cdots \wedge dx_{i_{n-1}} \wedge \underbrace{ddx_{i_n}}_{=0} \Big) \Bigg) \\
&= \sum d\alpha_{i_1 \cdots i_n} \wedge dx_{i_1} \wedge \cdots \wedge dx_{i_n} \\
&= \sum \left(\sum_{j=1}^{n} \frac{\partial \, \alpha_{i_1 \cdots i_n}}{\partial x_{i_j}} \, dx_{i_j} \right) \wedge dx_{i_1} \wedge \cdots \wedge dx_{i_n} \\
&= \sum \sum_{j=1}^{n} \frac{\partial \, \alpha_{i_1 \cdots i_n}}{\partial x_{i_j}} \, dx_{i_j} \wedge dx_{i_1} \wedge \cdots \wedge dx_{i_n}.
\end{aligned}$$

Thus our formula for the exterior differentiation operator d is thus given by

$$d\Big(\sum \alpha_{i_1\cdots i_n}\, dx_{i_1}\wedge\cdots\wedge dx_{i_n}\Big) = \sum\sum_{j=1}^{n} \frac{\partial\, \alpha_{i_1\cdots i_n}}{\partial x_{i_j}}\, dx_{i_j}\wedge dx_{i_1}\wedge\cdots\wedge dx_{i_n}.$$

This is exactly the formula for exterior differentiation d that was given in the last section. Notice that in the above string of equalities we used all four properties. Even though we used properties $i - iv$ to find this formula, to be completely rigorous we ought to turn around and show that this formula satisfies the properties. This is essentially what Question 4.3 asked you to do.

Question 4.4 Show where each of the properties $1 - 4$ were used in the above computation.

There is one final comment that needs to be made. What we have done is shown that the exterior differential operator exists and is unique, locally. What we mean by locally is within a single *coordinate patch* where the coordinates we used to find the formula apply. We discuss coordinate patches in Chap. 10 when we discuss general manifolds. However, for a Euclidian manifold $\mathbb{R}^n$ there is only one coordinate patch, so for the Euclidian manifold $\mathbb{R}^n$ we now know that the exterior differential operator exists and is unique on the whole manifold. For a general manifold we need to show existence and uniqueness is global and does not depend on the coordinate patch being used. We will address this issue in Sect. 10.4 once we have the necessary concepts and tools at our disposal.

4.4 The Global Formula

In the last two sections we did not try to address or think about what the geometric meaning behind exterior differentiation was; the last two sections were all about the formula and properties of the exterior derivative. In this section we will start to think about the underlying geometry in an attempt to both justify and find the global formula for the exterior derivative. As far as we are aware you will not see a presentation along these lines in any other introductions to exterior differentiation. However we believe it provides both a geometrical justification for exterior differentiation as well as allowing for a deeper intuitive understanding of both the exterior derivative and the global formula for it. A different geometric approach is presented in the next section.

4.4.1 Exterior Differentiation with Constant Vector Fields

Recall that given a one-form ϕ on the manifold $\mathbb{R}^n$, at each point $p \in \mathbb{R}^n$ we have a mapping $\phi : T_p(\mathbb{R}^n) \to \mathbb{R}$. If we were given a vector field v on manifold $\mathbb{R}^n$, then for each $p \in \mathbb{R}^n$ we have $\phi_p(v_p) \in \mathbb{R}$, that is, a real number. Hence we could view $\phi(v)$ as a function on the manifold $\mathbb{R}^n$. That is, its inputs are points p on the manifold and its outputs are real numbers, that is defined as so:

$$\begin{aligned} \phi(v) : \mathbb{R}^n &\longrightarrow \mathbb{R} \\ p &\longmapsto \phi_p(v_p). \end{aligned}$$

Also, we recall the notation $\langle\cdot,\cdot\rangle$ to mean the canonical pairing between a one-form and a vector, that is, $\phi(v) = \langle \phi, v\rangle$. If ϕ were an n-form then we would write the canonical pairing as $\phi(v_1,\ldots,v_n) = \langle \phi, (v_1,\ldots,v_n)\rangle$.

Now let us go back to the idea that differentiation is some sort of a measure of how a mathematical object, in this case a differential form, varies. For the moment we will continue to consider the special case of a one-form $\alpha = f_1 dx + f_2 dy$ on the manifold $\mathbb{R}^2$ since we can draw the associated pictures easily. Suppose we are given a *constant vector field* v on the manifold $\mathbb{R}^2$, where

$$v_p = v_1\frac{\partial}{\partial x} + v_2\frac{\partial}{\partial y} = \begin{bmatrix} v_1 \\ v_2 \end{bmatrix}_p$$

at every point $p \in \mathbb{R}^2$ and $v_1, v_2 \in \mathbb{R}$. The fact that we assume the vector field to be constant simplifies the computations considerably thereby making the underlying ideas clearer. Of course, vector fields need not be constant, and in fact generally they are not. The case of the non-constant vector field is handled next.

One of the ways that we could try to measure how the differential one-form $\alpha : T(\mathbb{R}^2) \to \mathbb{R}$ varies is to instead consider how the function $\langle \alpha, v \rangle$ varies. True, doing this does not just measure how just α varies, instead it measure how $\langle \alpha, v \rangle$ taken together vary. But be that as it may, at least it is an approach that we know how to handle, meaning that we already know how to measure how functions vary in different directions. This is just our usual directional derivative.

But to do this we need to vary $\langle \alpha, v \rangle$ in some direction, say in the direction w. Again, to keep things simple we will choose w to be a constant vector field on the manifold $\mathbb{R}^2$,

$$w_p = w_1 \frac{\partial}{\partial x} + w_2 \frac{\partial}{\partial y} = \begin{bmatrix} w_1 \\ w_2 \end{bmatrix}_p$$

at every point $p \in \mathbb{R}^2$. Thus, seeing how the function $\langle \alpha, v \rangle$ varies in the direction w means finding the directional derivative of the function $\langle \alpha, v \rangle$ in the direction of w, that is, finding $d\langle \alpha, v \rangle\big(w\big) = w[\langle \alpha, v \rangle]$. In order to do this we first find the function $\langle \alpha, v \rangle$,

$$\begin{aligned}
\langle \alpha, v \rangle &= (f_1 dx + f_2 dy)\left(v_1 \frac{\partial}{\partial x} + v_2 \frac{\partial}{\partial y}\right) \\
&= [f_1, f_2] \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} \\
&= v_1 f_1 + v_2 f_2.
\end{aligned}$$

Notice, this is just some sort of a "weighted sum" of the functions f_1 and f_2 with weights given by v_1 and v_2, which come from our constant vector v. We already know how to take the differential of a function, so we have

$$\begin{aligned}
d\langle \alpha, v \rangle &= \frac{\partial \langle \alpha, v \rangle}{\partial x} dx + \frac{\partial \langle \alpha, v \rangle}{\partial y} dy \\
&= \frac{\partial (v_1 f_1 + v_2 f_2)}{\partial x} dx + \frac{\partial (v_1 f_1 + v_2 f_2)}{\partial y} dy \\
&= \left(v_1 \frac{\partial f_1}{\partial x} + v_2 \frac{\partial f_2}{\partial x}\right) dx + \left(v_1 \frac{\partial f_1}{\partial y} + v_2 \frac{\partial f_2}{\partial y}\right) dy.
\end{aligned}$$

Using this we can finally find the directional derivative of $\langle \alpha, v \rangle$ in the direction of w as follows,

$$\begin{aligned}
d\langle \alpha, v \rangle\big(w\big) &= \big\langle d\langle \alpha, v \rangle, w \big\rangle \\
&= \left[v_1 \frac{\partial f_1}{\partial x} + v_2 \frac{\partial f_2}{\partial x},\ v_1 \frac{\partial f_1}{\partial y} + v_2 \frac{\partial f_2}{\partial y}\right] \begin{bmatrix} w_1 \\ w_2 \end{bmatrix} \\
&= \left(v_1 \frac{\partial f_1}{\partial x} + v_2 \frac{\partial f_2}{\partial x}\right) w_1 + \left(v_1 \frac{\partial f_1}{\partial y} + v_2 \frac{\partial f_2}{\partial y}\right) w_2 \\
&= v_1 w_1 \frac{\partial f_1}{\partial x} + v_2 w_1 \frac{\partial f_2}{\partial x} + v_1 w_2 \frac{\partial f_1}{\partial y} + v_2 w_2 \frac{\partial f_2}{\partial y}.
\end{aligned}$$

We try to show this pictorially in Fig. 4.1.

In summary, what we have here,

$$d\langle \alpha, v \rangle\big(w\big) = v_1 w_1 \frac{\partial f_1}{\partial x} + v_2 w_1 \frac{\partial f_2}{\partial x} + v_1 w_2 \frac{\partial f_1}{\partial y} + v_2 w_2 \frac{\partial f_2}{\partial y}$$

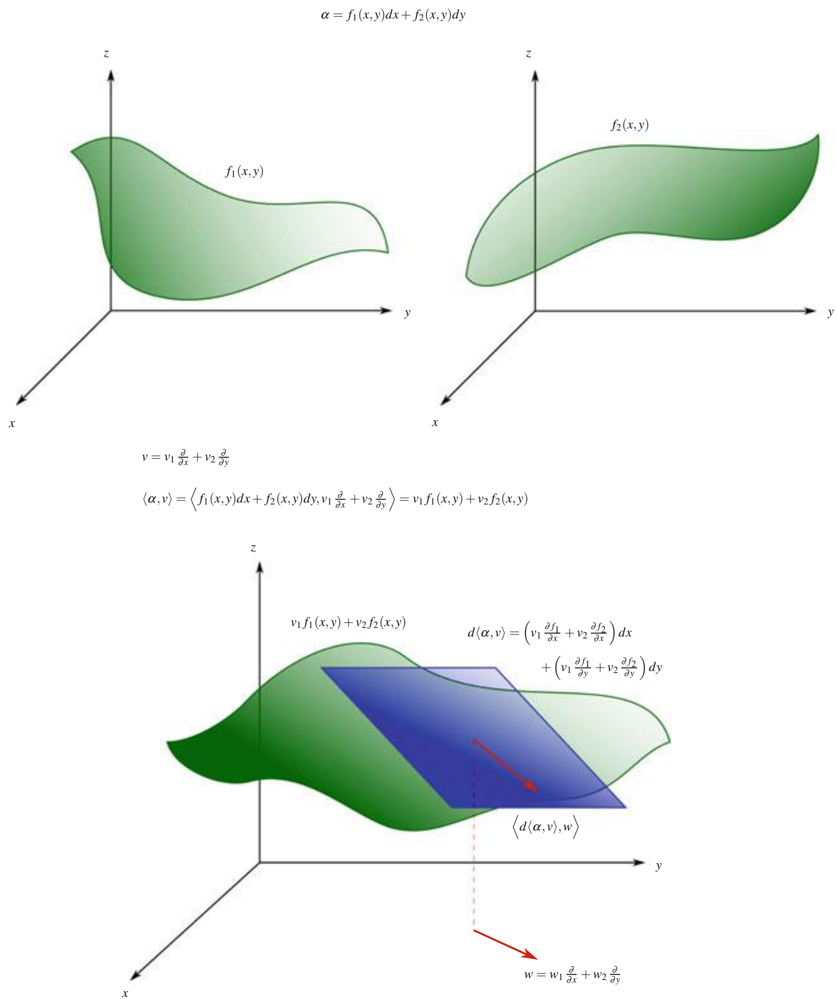


Fig. 4.1 The one-form $\alpha_{(x,y)} = f_1(x, y)dx + f_2(x, y)dy$ on the manifold $\mathbb{R}^2$ is made up of two functions $f_1, f_2 : \mathbb{R}^2 \to \mathbb{R}$, shown above. Once we are given a vector field v on manifold $\mathbb{R}^3$, here the constant vector field $v = v_1\partial_x + v_2\partial_y$, then this can be used to find the real-valued function $\langle \alpha, v\rangle = v_1 f_1 + v_2 f_2$, which can be viewed as a linear combination of the two functions f_1, f_2. The directional derivative of this function can then be found in the direction of another given vector $w = w_1\partial_x + w_2\partial_y$. Here the differential $d\langle \alpha, v\rangle$ in essence encodes the information about the tangent plane to the function $\langle \alpha, v\rangle$, shown in green

is a function from the manifold $\mathbb{R}^2$ to $\mathbb{R}$. When we input some point $p = (x_0, y_0)$ what we get is

$$\begin{aligned} d\langle \alpha_{(x_0,y_0)}, v\rangle\big(w\big) =& v_1 w_1 \left.\frac{\partial f_1(x,y)}{\partial x}\right|_{(x_0,y_0)} + v_2 w_1 \left.\frac{\partial f_2(x,y)}{\partial x}\right|_{(x_0,y_0)} \\ &+ v_1 w_2 \left.\frac{\partial f_1(x,y)}{\partial y}\right|_{(x_0,y_0)} + v_2 w_2 \left.\frac{\partial f_2(x,y)}{\partial y}\right|_{(x_0,y_0)}, \end{aligned}$$

which is a number that measures how much the function $\langle \alpha, v\rangle = v_1 f_1 + v_2 f_2$ varies as we move along the vector w at the given point $p = (x_0, y_0)$.

This clearly is not simply a measure of how the one-form α varies. Instead, it requires two additional vectors, v and w. The first vector v is used along with the one-form to make the real-valued function $\langle \alpha, v\rangle$ and then the second vector is used to find the directional derivative of this function. While this seems like a somewhat odd way to measure how α varies it is perhaps about as straight-forward as we can expect to get given how complicated a mathematical object a one-form actually is.

But before we try to somehow define the differential of a one-form α with this formula we decide to be cautious and see what happens when we switch the order of the vectors v and w. That is, we want to know what $d\langle \alpha, w\rangle\big(v\big)$ is. After all, if we simply change the order of the vectors v and w we would hope that the results are "the same" in some sense. Proceeding to do just this we have $\langle\alpha, w\rangle = w_1 f_1 + w_2 f_2$, which gives

$$d\langle\alpha, w\rangle = \left(w_1\frac{\partial f_1}{\partial x} + w_2\frac{\partial f_2}{\partial x}\right)dx + \left(w_1\frac{\partial f_1}{\partial y} + w_2\frac{\partial f_2}{\partial y}\right)dy,$$

resulting in

$$\begin{aligned} d\langle\alpha, w\rangle\big(v\big) &= \big\langle d\langle\alpha, w\rangle, v\big\rangle \\ &= v_1 w_1\frac{\partial f_1}{\partial x} + v_1 w_2\frac{\partial f_2}{\partial x} + v_2 w_1\frac{\partial f_1}{\partial y} + v_2 w_2\frac{\partial f_2}{\partial y}. \end{aligned}$$

Let us put these two formulas side by side to see how they match up.

$$\begin{aligned} \Big\langle d\langle\alpha, v\rangle, w\Big\rangle &= v_1 w_1\frac{\partial f_1}{\partial x} + v_2 w_1\frac{\partial f_2}{\partial x} + v_1 w_2\frac{\partial f_1}{\partial y} + v_2 w_2\frac{\partial f_2}{\partial y} \\ \Big\langle d\langle\alpha, w\rangle, v\Big\rangle &= v_1 w_1\frac{\partial f_1}{\partial x} + v_1 w_2\frac{\partial f_2}{\partial x} + v_2 w_1\frac{\partial f_1}{\partial y} + v_2 w_2\frac{\partial f_2}{\partial y}. \end{aligned}$$

So, these two equalities are not the same as we may have hoped. Nor do they simply differ from each other by a sign. But clearly there is a relation between $d\langle\alpha, v\rangle\big(w\big)$ and $d\langle\alpha, w\rangle\big(v\big)$. The relation between them is a little more complicated. The first and last terms of each equality are identical but the middle two terms are not. We can not use either of these formulas as a definition for the derivative of a one-form. However, all is not lost. A bit of clever thinking may yet salvage the situation.

In the past we have discovered that the "volume" of a parallelepiped has a sign attached to it that depends on the orientation, or order, of the vectors. This sign "ambiguity" has shown up in the definition of determinant, and then in the definition of wedgeproduct. So perhaps we should not be too surprised, or concerned, if it shows up in whatever definition of derivative of a form we end up settling on. So we will use a little "trick" that pops up in mathematics from time to time. What would happen if we take the difference between these two terms as the definition of exterior differentiation? That is, suppose we define the exterior derivative of α, $d\alpha$, by the following formula

$$d\alpha(v, w) = \Big\langle d\langle\alpha, w\rangle, v\Big\rangle - \Big\langle d\langle\alpha, v\rangle, w\Big\rangle.$$

With this formula we then notice that by changing the order of v and w the sign is changed,

$$\begin{aligned} d\alpha(w, v) &= \big\langle d\langle\alpha, v\rangle, w\big\rangle - \big\langle d\langle\alpha, w\rangle, v\big\rangle \\ &= -\Big(\big\langle d\langle\alpha, w\rangle, v\big\rangle - \big\langle d\langle\alpha, v\rangle, w\big\rangle\Big) \\ &= -d\alpha(v, w). \end{aligned}$$

By doing this we have that $d\alpha(v, w)$ and $d\alpha(w, v)$ give us the same answer up to sign, something that we are already perfectly used to and comfortable with. In fact, given the way signs seem to keep behaving in this subject it would be surprising if switching the order of two vectors didn't change the sign of something. So, we will settle on

$$d\alpha(v, w) = \big\langle d\langle\alpha, w\rangle, v\big\rangle - \big\langle d\langle\alpha, v\rangle, w\big\rangle.$$

as our working definition for the derivative of α. (In reality this working definition only works if v and w happen to be constant vector fields, but more on that later.) Geometrically, $d\alpha(v, w)$ measures the change in the function $\langle\alpha, w\rangle$ in the v direction minus the change in the function $\langle\alpha, v\rangle$ in the w direction. We attempt to show this in Fig. 4.2.

Spend a little time thinking about what we have done. Since the one-form $\alpha = f_1 dx + f_2 dy$ is a rather complicated mathematical object, no obvious direct way of measuring how it changes is apparent. Thus we decided to try to measure how α changes indirectly using techniques that we were already familiar with, namely directional derivatives. In order to do this we needed two (constant) vector fields v and w. Thus, once we were given a particular v and w we used these to indirectly measure how α changes.

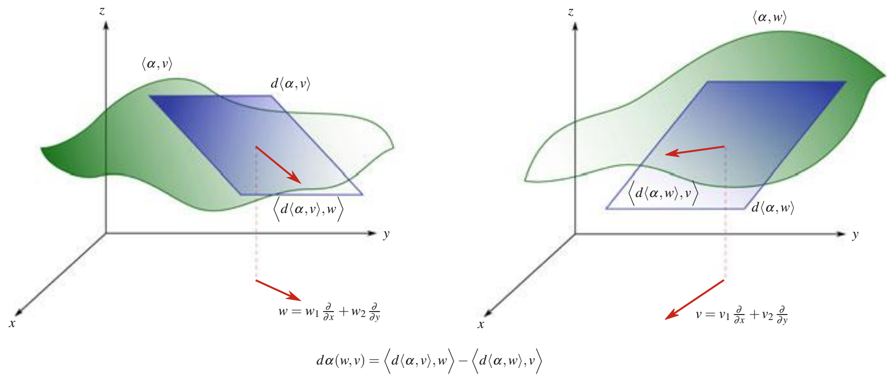


Fig. 4.2 The change in $\langle\alpha, v\rangle$ in the direction w is shown on the left and the change in $\langle\alpha, w\rangle$ in the direction v is shown on the left. The difference in these two values is then used to define $d\alpha(w, v)$. This is an indirect way to measure how α changes

Let us use this to get an actual expression for $d\alpha(v, w)$,

$$
\begin{aligned}
d\alpha(v, w) =& \Big\langle d\langle\alpha, w\rangle, v\Big\rangle - \Big\langle d\langle\alpha, v\rangle, w\Big\rangle \\
=& \left(v_1w_1\frac{\partial f_1}{\partial x} + v_1w_2\frac{\partial f_2}{\partial x} + v_2w_1\frac{\partial f_1}{\partial y} + v_2w_2\frac{\partial f_2}{\partial y}\right) \\
&- \left(v_1w_1\frac{\partial f_1}{\partial x} + v_2w_1\frac{\partial f_2}{\partial x} + v_1w_2\frac{\partial f_1}{\partial y} + v_2w_2\frac{\partial f_2}{\partial y}\right) \\
=& v_1w_2\frac{\partial f_2}{\partial x} - v_1w_2\frac{\partial f_1}{\partial y} - v_2w_1\frac{\partial f_2}{\partial x} + v_2w_1\frac{\partial f_1}{\partial y} \\
=& \left(\frac{\partial f_2}{\partial x} - \frac{\partial f_1}{\partial y}\right)(v_1w_2 - w_1v_2) \\
=& \left(\frac{\partial f_2}{\partial x} - \frac{\partial f_1}{\partial y}\right)\begin{vmatrix} v_1 & w_1 \\ v_2 & w_2 \end{vmatrix} \\
=& \left(\frac{\partial f_2}{\partial x} - \frac{\partial f_1}{\partial y}\right)\begin{vmatrix} dx(v) & dx(w) \\ dy(v) & dy(w) \end{vmatrix} \\
=& \left(\frac{\partial f_2}{\partial x} - \frac{\partial f_1}{\partial y}\right) dx \wedge dy(v, w) \\
=& \left(\frac{\partial f_1}{\partial x}\underbrace{dx \wedge dx}_{=0} + \frac{\partial f_1}{\partial y}\underbrace{dy \wedge dx}_{=-dx\wedge dy} + \frac{\partial f_2}{\partial x}dx \wedge dy + \frac{\partial f_2}{\partial y}\underbrace{dy \wedge dy}_{=0}\right)(v, w) \\
=& \left(\underbrace{\left(\frac{\partial f_1}{\partial x}dx + \frac{\partial f_1}{\partial y}dy\right)}_{df_1} \wedge dx + \underbrace{\left(\frac{\partial f_2}{\partial x}dx + \frac{\partial f_2}{\partial y}dy\right)}_{df_2} \wedge dy\right)(v, w) \\
=& \Big(df_1 \wedge dx + df_2 \wedge dy\Big)(v, w).
\end{aligned}
$$

Thus our "geometric" definition of $d\alpha$ as the change in the function $\langle\alpha, w\rangle$ in the v direction minus the change in the function $\langle\alpha, v\rangle$ in the w direction leads us all the way back to our original formula for $d\alpha$, $\sum df_i \wedge dx_i$, when $\alpha = \sum f_i dx_i$. Furthermore, we have a wonderful bonus, $d\alpha$ turns out to be a two-form. Okay, in reality this is actually more than just a "bonus." The fact that this definition results in $d\alpha$ being a two-form is actually a prime reason we settle on this admittedly slightly awkward geometrical meaning for the exterior derivative of a one-form.

Actually, in order to keep the computations and pictures simple we looked at a one-form α on the manifold $\mathbb{R}^2$, but we could have done the whole analysis for any manifold and arrived at the same formula; the computations would have simply been more cumbersome.

We recognize that we tend to bounce between different notations, and we will continue to do so. We are actually doing this on purpose, it is important that you become comfortable with all of these different notations. Different books, papers, and authors all tend to use different notations. Recall that $df(v) = v[f]$ and $d\langle\alpha, w\rangle\big(v\big) = v[\langle\alpha, w\rangle]$. Also, we have $\langle\alpha, w\rangle = \alpha(w)$, so we could write $\big\langle d\langle\alpha, w\rangle, v\big\rangle$ as $v[\alpha(w)]$. Similarly, we have $\big\langle d\langle\alpha, v\rangle, w\big\rangle = w[\alpha(v)]$, so we can also write the definition of the exterior derivative of α, where v and w are constant vector fields, as

Global formula for exterior derivative of a one-form, constant vector fields	$d\alpha(v, w) = \big\langle d\langle\alpha, w\rangle, v\big\rangle - \big\langle d\langle\alpha, v\rangle, w\big\rangle$

or as

Global formula for exterior derivative of a one-form, constant vector fields	$d\alpha(v, w) = v[\alpha(w)] - w[\alpha(v)].$

Notice, these are exactly the first two terms in our global formula given in the overview section,

$$d\alpha(v, w) = v[\alpha(w)] - w[\alpha(v)] - \alpha([v, w]).$$

Thus we see that our “geometric” definition of $d\alpha$ is tied to the global formula for $d\alpha$. However, we are missing the final term $\alpha([v, w])$ from the global formula. This term is zero when both v and w are constant vector fields. Avoiding the mess of this term was the reason we chose constant vector fields v and w. Again, we point out that in this global formula nowhere do the actual coordinates that we are using show up. Sometimes this is called a coordinate-free formula.

Now we will repeat this quickly for a general one-form on a general n-dimensional manifold. The computations apply to what is called a coordinate patch of a general manifold. That is a concept we will introduce in a latter chapter, so for now just imagine that we are on the manifold $\mathbb{R}^n$. Consider the one-form and vectors

$$\begin{aligned}
\alpha &= \sum_i \alpha_i \, dx_i, \\
v &= \sum_i v_i \, \partial_{x_i}, \\
w &= \sum_i w_i \, \partial_{x_i}.
\end{aligned}$$

Then we have

$$\begin{aligned}
\alpha(v) &= \langle\alpha, v\rangle \\
&= \left(\sum_i \alpha_i \, dx_i\right)\left(\sum_i v_i \, \partial_{x_i}\right) \\
&= \sum_i v_i \alpha_i .
\end{aligned}$$

As before, $\langle\alpha, v\rangle = \sum_i v_i\alpha_i$ can be viewed as a function on the manifold. Basically this function it is the sum of the functions α_i weighted by v_i. Using d as defined in the directional derivative of a function case, we have

$$\begin{aligned}
d\langle\alpha, v\rangle &= \sum_j \frac{\partial\left(\sum_i v_i\alpha_i\right)}{\partial x_j} dx_j \\
&= \sum_j \sum_i v_i \frac{\partial\alpha_i}{\partial x_j} dx_j .
\end{aligned}$$

So the directional derivative of the function $\langle \alpha, v\rangle = \sum_i v_i\alpha_i$ in the direction w is given by

$$\begin{aligned} d\langle \alpha, v\rangle(w) &= \left(\sum_j \sum_i v_i \frac{\partial \alpha_i}{\partial x_j} dx_j\right)\left(\sum_k w_k \partial_{x_k}\right) \\ &= \sum_j \sum_i v_i w_j \frac{\partial \alpha_i}{\partial x_j}. \end{aligned}$$

Recall, basically $d\langle \alpha, v\rangle$ encodes the tangent space of the function $\langle \alpha, v\rangle$ and $d\langle \alpha, v\rangle(w)$ tell us the change that occurs in as we move by w. A completely identical computation gives us

$$d\langle \alpha, w\rangle(v) = \sum_j \sum_i v_j w_i \frac{\partial \alpha_i}{\partial x_j}.$$

Comparing $d\langle \alpha, v\rangle(w)$ and $d\langle \alpha, w\rangle(v)$ we can see that when $i \neq j$ the terms differ, so like before we will define

$$\begin{aligned} d\alpha(v, w) &= \big\langle d\langle \alpha, w\rangle, v\big\rangle - \big\langle d\langle \alpha, v\rangle, w\big\rangle \\ &= \sum_j \sum_i v_j w_i \frac{\partial \alpha_i}{\partial x_j} - \sum_j \sum_i v_i w_j \frac{\partial \alpha_i}{\partial x_j} \\ &= \sum_j \sum_i \big(v_j w_i - v_i w_j\big) \frac{\partial \alpha_i}{\partial x_j} \\ &= \sum_j \sum_i \frac{\partial \alpha_i}{\partial x_j} \begin{vmatrix} v_j & w_j \\ v_i & w_i \end{vmatrix} \\ &= \sum_j \sum_i \frac{\partial \alpha_i}{\partial x_j} \begin{vmatrix} dx_j(v) & dx_j(w) \\ dx_i(v) & dx_i(w) \end{vmatrix} \\ &= \sum_j \sum_i \frac{\partial \alpha_i}{\partial x_j} dx_j \wedge dx_i(v, w) \\ &= \sum_i \underbrace{\left(\sum_j \frac{\partial \alpha_i}{\partial x_j} dx_j\right)}_{=d\alpha_i} \wedge dx_i(v, w). \end{aligned}$$

Hence we have the in-coordinates formula

$$d\alpha = \sum_i d\alpha_i \wedge dx_i,$$

which is of course exactly what we had before.

As in the case of the one-forms we could attempt to make an geometrical argument for the global formula of a k-form ω. The first step would be completely analogous. In order to use the concept of the directional derivative of a function we would have to turn a k-from into a function by putting in not one, but k vectors $v_0, \ldots, v_{k-1}$ and then we would measure how that function changes in the direction of yet another vector v_k. However, the next step, where we actually use the math "trick" to define $d\omega$ would be difficult to rationalize. We would end up with

Global formula for exterior derivative of a k-form, constant vector fields	$d\omega(v_0, \ldots, v_k) = \sum_i (-1)^i \Big\langle d\big\langle \omega, (v_0, \ldots, \widehat{v_i}, \ldots, v_k)\big\rangle, v_i \Big\rangle,$

which could also be written as

Global formula for exterior derivative of a k-form, constant vector fields	$d\omega(v_0,\dots,v_k) = \sum_i (-1)^i v_i\big[\omega(v_0,\dots,\widehat{v_i},\dots,v_k)\big].$

Using this global formula to arrive at the in-coordinates formula would also be a very computationally intense process. Therefore we will forgo it. However, notice that the above formulas are different from the global formula in the overview. Like before, this is because we are using constant vector fields.

4.4.2 Exterior Differentiation with Non-Constant Vector Fields

What we have done for constant vector fields v and w is to make some geometrical augments that $d\alpha(v,w) = v[\alpha(w)] - w[\alpha(v)]$ would be a good definition for the global formula of the exterior derivative of α. We then used this formula to find the in-coordinates formula for the exterior derivative of the one-form $\alpha = \sum_i \alpha_i dx_i$ and found it to be $d\alpha = \sum_i d\alpha_i \wedge dx_i$. We will now assume that this is the form we want the in-coordinates version of the general global formula to take.

Since we are not trying to present an actual proof of the global formula for exterior differentiation but instead are trying to give you a feel for how and why the formula is what it is; we will not attempt a logically rigorous argument. What we will be doing is actually logically backwards. We know what answer we want to get, that is, we know what the in-coordinates formula for the exterior derivative should be, so we will actually use this knowledge of where we want to end up to help us find and understand the general global formula for exterior differentiation. The general global formula for exterior differentiation will actually be proved in Sect. A.7 in a completely different way using a number of concepts and identities that will be introduced in the appendix. The proof in that section is very computational and the formula seems to miraculously appear out of a stew of different ingredients. This section is meant to be more down-to-earth, giving motivation and addressing the underlying geometry to a greater extent.

We begin with the one-form α and two *non-constant vector fields* v and w on $\mathbb{R}^2$,

$$\alpha = f_1 dx + f_2 dy,$$

$$v = g_1\partial_x + g_2\partial_y = \begin{bmatrix} g_1 \\ g_2 \end{bmatrix},$$

$$w = h_1\partial_x + h_2\partial_y = \begin{bmatrix} h_1 \\ h_2 \end{bmatrix},$$

where $f_1, f_2, g_1, g_1, h_1, h_2$ are real-valued functions on the manifold $\mathbb{R}^2$.

As we said, we will start with the answer we know we want to arrive at. We want the in-coordinates formula for the exterior derivative of α to be $d\alpha = \sum_i d\alpha_i \wedge dx_i$. So first we compute that,

$$\begin{aligned} d\alpha &= df_1 \wedge dx + df_2 \wedge dy \\ &= \left(\frac{\partial f_1}{\partial x}dx + \frac{\partial f_1}{\partial y}dy\right) \wedge dx + \left(\frac{\partial f_2}{\partial x}dx + \frac{\partial f_2}{\partial y}dy\right) \wedge dy \\ &= \frac{\partial f_1}{\partial x}\underbrace{dx\wedge dx}_{=0} + \frac{\partial f_1}{\partial y}\underbrace{dy\wedge dx}_{=-dx\wedge dy} + \frac{\partial f_2}{\partial x}dx\wedge dy + \frac{\partial f_2}{\partial y}\underbrace{dy\wedge dy}_{=0} \\ &= \left(\frac{\partial f_2}{\partial x} - \frac{\partial f_1}{\partial y}\right) dx\wedge dy. \end{aligned}$$

Plugging in our vector fields v and w we get

$$d\alpha(v,w) = \left(\frac{\partial f_2}{\partial x} - \frac{\partial f_1}{\partial x}\right) dx\wedge dy(v,w)$$

$$
\begin{aligned}
&= \left(\frac{\partial f_2}{\partial x} - \frac{\partial f_1}{\partial x}\right)\begin{vmatrix} dx(v) & dx(w) \\ dy(v) & dy(w) \end{vmatrix} \\
&= \left(\frac{\partial f_2}{\partial x} - \frac{\partial f_1}{\partial x}\right)\begin{vmatrix} g_1 & h_1 \\ g_2 & h_2 \end{vmatrix} \\
&= \left(\frac{\partial f_2}{\partial x} - \frac{\partial f_1}{\partial x}\right)(g_1h_2 - g_2h_1) \\
&= g_1h_2\frac{\partial f_2}{\partial x} - g_1h_2\frac{\partial f_1}{\partial y} - g_2h_1\frac{\partial f_2}{\partial x} + g_2h_1\frac{\partial f_1}{\partial y}.
\end{aligned}
$$

By now we should be comfortable with this calculation. The only difference is that the constants v_1, v_1, w_1, w_2 have been replaced with the functions g_1, g_2, h_1, h_2. This is the answer we want to get to. Later on we will refer back to this equation.

Our strategy is the same as in the constant vector field case for the same geometric reasons. In fact, based on the constant vector field case we will first hypothesize that the global formula is $d\alpha(v, w) = v[\alpha(w)] - w[\alpha(v)]$. We will do this (admittedly messy) calculation and discover there are a lot more terms than the local formula for $d\alpha(v, w)$ has, which we found just above. We will then group those terms together and call them $-\alpha([v, w])$ thereby giving the full global formula $d\alpha(v, w) = v[\alpha(w)] - w[\alpha(v)] - \alpha([v, w])$ for the one-form case. Of course, at that point you will have no idea what $\alpha([v, w])$ is. We will then spend some time explaining what this final term actually represents. The brackets $[\cdot, \cdot]$, with the dots replaced by vectors, is call the Lie bracket, so $[v, w]$ is the Lie bracket of v and w, and turns out to be another vector field, which is good since it is being eaten by the one-form α.

Now we are ready to begin. The same arguments apply. We want to somehow measure how α varies. Since we are not sure how to do this we first change α to something that we know how to study, namely a function, which we can do by pairing it with the vector field v, which no longer needs to be a constant vector field,

$$
\langle \alpha, v\rangle = [f_1, f_2]\begin{bmatrix} g_1 \\ g_2 \end{bmatrix} = f_1g_1 + f_2g_2.
$$

Compare this to the case with the constant vector field, where we had obtained $v_1f_1 + v_2f_2$ with v_1 and v_2 being constants. The situation is immediately more complicated since it involves products of functions and not just a weighted sum of functions.

In order to find the directional derivative of the function $f_1g_1 + f_2g_2$ we first find the differential of the function $\langle\alpha, v\rangle$,

$$
\begin{aligned}
d\langle\alpha, v\rangle =& \frac{\partial(f_1g_1 + f_2g_2)}{\partial x}dx + \frac{\partial(f_1g_1 + f_2g_2)}{\partial y}dy \\
=& \left(\frac{\partial(f_1g_1)}{\partial x} + \frac{\partial(f_2g_2)}{\partial x}\right)dx + \left(\frac{\partial(f_1g_1)}{\partial y} + \frac{\partial(f_2g_2)}{\partial y}\right)dy \\
=& \left(\frac{\partial f_1}{\partial x}\cdot g_1 + f_1\cdot\frac{\partial g_1}{\partial x} + \frac{\partial f_2}{\partial x}\cdot g_2 + f_2\cdot\frac{\partial g_2}{\partial x}\right)dx \\
&+ \left(\frac{\partial f_1}{\partial y}\cdot g_1 + f_1\cdot\frac{\partial g_1}{\partial y} + \frac{\partial f_2}{\partial y}\cdot g_2 + f_2\cdot\frac{\partial g_2}{\partial y}\right)dy.
\end{aligned}
$$

Notice that the required use of the product rule here results in twice the number of terms as we had previously when we were dealing with constant vector fields. And these terms are themselves products of functions. This is what makes dealing with non-constant vector fields so much more complicated. We now find the directional derivative of $\langle\alpha, v\rangle$ in the direction of w by

$$
\begin{aligned}
\Big\langle d\langle\alpha, v\rangle, w\Big\rangle =& \left(\frac{\partial f_1}{\partial x}\cdot g_1 + f_1\cdot\frac{\partial g_1}{\partial x} + \frac{\partial f_2}{\partial x}\cdot g_2 + f_2\cdot\frac{\partial g_2}{\partial x}\right)dx\left(\begin{bmatrix} h_1 \\ h_2 \end{bmatrix}\right) \\
&+ \left(\frac{\partial f_1}{\partial y}\cdot g_1 + f_1\cdot\frac{\partial g_1}{\partial y} + \frac{\partial f_2}{\partial y}\cdot g_2 + f_2\cdot\frac{\partial g_2}{\partial y}\right)dy\left(\begin{bmatrix} h_1 \\ h_2 \end{bmatrix}\right)
\end{aligned}
$$

$$
\begin{aligned}
&=\left(\frac{\partial f_1}{\partial x}\cdot g_1+f_1\cdot\frac{\partial g_1}{\partial x}+\frac{\partial f_2}{\partial x}\cdot g_2+f_2\cdot\frac{\partial g_2}{\partial x}\right)h_1\\
&\quad+\left(\frac{\partial f_1}{\partial y}\cdot g_1+f_1\cdot\frac{\partial g_1}{\partial y}+\frac{\partial f_2}{\partial y}\cdot g_2+f_2\cdot\frac{\partial g_2}{\partial y}\right)h_2\\
&=g_1h_1\frac{\partial f_1}{\partial x}+f_1h_1\frac{\partial g_1}{\partial x}+g_2h_1\frac{\partial f_2}{\partial x}+f_2h_1\frac{\partial g_2}{\partial x}\\
&\quad+g_1h_2\frac{\partial f_1}{\partial y}+f_1h_2\frac{\partial g_1}{\partial y}+g_2h_2\frac{\partial f_2}{\partial y}+f_2h_2\frac{\partial g_2}{\partial y}.
\end{aligned}
$$

Next we repeat this procedure exchanging the vector fields v and w,

$$
\langle\alpha, w\rangle=[f_1, f_2]\begin{bmatrix}h_1\\h_2\end{bmatrix}=f_1h_1+f_2h_2.
$$

In order to find the directional derivative we need

$$
\begin{aligned}
d\langle\alpha, w\rangle&=\frac{\partial(f_1h_1+f_2h_2)}{\partial x}dx+\frac{\partial(f_1h_1+f_2h_2)}{\partial y}dy\\
&=\left(\frac{\partial(f_1h_1)}{\partial x}+\frac{\partial(f_2h_2)}{\partial x}\right)dx+\left(\frac{\partial(f_1h_1)}{\partial y}+\frac{\partial(f_2h_2)}{\partial y}\right)dy\\
&=\left(\frac{\partial f_1}{\partial x}\cdot h_1+f_1\cdot\frac{\partial h_1}{\partial x}+\frac{\partial f_2}{\partial x}\cdot h_2+f_2\cdot\frac{\partial h_2}{\partial x}\right)dx\\
&\quad+\left(\frac{\partial f_1}{\partial y}\cdot h_1+f_1\cdot\frac{\partial h_1}{\partial y}+\frac{\partial f_2}{\partial y}\cdot h_2+f_2\cdot\frac{\partial h_2}{\partial y}\right)dy.
\end{aligned}
$$

Thus the directional derivative of $\langle\alpha, w\rangle$ in the direction of v is given by

$$
\begin{aligned}
\Big\langle d\langle\alpha, w\rangle, v\Big\rangle&=\left(\frac{\partial f_1}{\partial x}\cdot h_1+f_1\cdot\frac{\partial h_1}{\partial x}+\frac{\partial f_2}{\partial x}\cdot h_2+f_2\cdot\frac{\partial h_2}{\partial x}\right)dx\left(\begin{bmatrix}g_1\\g_2\end{bmatrix}\right)\\
&\quad+\left(\frac{\partial f_1}{\partial y}\cdot h_1+f_1\cdot\frac{\partial h_1}{\partial y}+\frac{\partial f_2}{\partial y}\cdot h_2+f_2\cdot\frac{\partial h_2}{\partial y}\right)dy\left(\begin{bmatrix}g_1\\g_2\end{bmatrix}\right)\\
&=\left(\frac{\partial f_1}{\partial x}\cdot h_1+f_1\cdot\frac{\partial h_1}{\partial x}+\frac{\partial f_2}{\partial x}\cdot h_2+f_2\cdot\frac{\partial h_2}{\partial x}\right)g_1\\
&\quad+\left(\frac{\partial f_1}{\partial y}\cdot h_1+f_1\cdot\frac{\partial h_1}{\partial y}+\frac{\partial f_2}{\partial y}\cdot h_2+f_2\cdot\frac{\partial h_2}{\partial y}\right)g_2\\
&=h_1g_1\frac{\partial f_1}{\partial x}+f_1g_1\frac{\partial h_1}{\partial x}+h_2g_1\frac{\partial f_2}{\partial x}+f_2g_1\frac{\partial h_2}{\partial x}\\
&\quad+h_1g_2\frac{\partial f_1}{\partial y}+f_1g_2\frac{\partial h_1}{\partial y}+h_2g_2\frac{\partial f_2}{\partial y}+f_2g_2\frac{\partial h_2}{\partial y}.
\end{aligned}
$$

Recall that in our strategy we were first going to hypothesis that $d\alpha(v, w)=v[\alpha(w)]-w[\alpha(v)]$. We chose this in the constant vector case simply because it was something that seemed to work, and it also resulted in the fact that switching the vectors v and w caused the sign to change. Proceeding with our hypothesis we have

$$
\begin{aligned}
d\alpha(v, w)&\underset{\text{hyp.}}{=}v[\alpha(w)]-w[\alpha(v)]\\
&=\Big\langle d\langle\alpha, w\rangle, v\Big\rangle-\Big\langle d\langle\alpha, v\rangle, w\Big\rangle
\end{aligned}
$$

$$
\begin{aligned}
&= \cancel{h_1 g_1 \frac{\partial f_1}{\partial x}} + f_1 g_1 \frac{\partial h_1}{\partial x} + h_2 g_1 \frac{\partial f_2}{\partial x} + f_2 g_1 \frac{\partial h_2}{\partial x} \\
&\quad + h_1 g_2 \frac{\partial f_1}{\partial y} + f_1 g_2 \frac{\partial h_1}{\partial y} + \cancel{h_2 g_2 \frac{\partial f_2}{\partial y}} + f_2 g_2 \frac{\partial h_2}{\partial y} \\
&\quad - \cancel{g_1 h_1 \frac{\partial f_1}{\partial x}} - f_1 h_1 \frac{\partial g_1}{\partial x} - g_2 h_1 \frac{\partial f_2}{\partial x} - f_2 h_1 \frac{\partial g_2}{\partial x} \\
&\quad - g_1 h_2 \frac{\partial f_1}{\partial y} - f_1 h_2 \frac{\partial g_1}{\partial y} - \cancel{g_2 h_2 \frac{\partial f_2}{\partial y}} - f_2 h_2 \frac{\partial g_2}{\partial y}.
\end{aligned}
$$

Notice that four terms cancel out right away. Since we already know what we want our local in-coordinates formula to look like we now turn to that. Based on the computation above, what we are trying to arrive at is

$$
d\alpha(v, w) = g_1 h_2 \frac{\partial f_2}{\partial x} - g_1 h_2 \frac{\partial f_1}{\partial y} - g_2 h_1 \frac{\partial f_2}{\partial x} + g_2 h_1 \frac{\partial f_1}{\partial y}.
$$

These terms are shown in green above. Thus we have eight extra terms

$$
\begin{aligned}
& f_1 g_1 \frac{\partial h_1}{\partial x} + f_2 g_1 \frac{\partial h_2}{\partial x} + f_1 g_2 \frac{\partial h_1}{\partial y} + f_2 g_2 \frac{\partial h_2}{\partial y} \\
& - f_1 h_1 \frac{\partial g_1}{\partial x} - f_2 h_1 \frac{\partial g_2}{\partial x} - f_1 h_2 \frac{\partial g_1}{\partial y} - f_2 h_2 \frac{\partial g_2}{\partial y}.
\end{aligned}
$$

As we said we would do in our strategy, we will group these terms together and call them $-\alpha([v, w])$ resulting in

$$
\begin{aligned}
\alpha([v, w]) = \; & f_1 h_1 \frac{\partial g_1}{\partial x} + f_2 h_1 \frac{\partial g_2}{\partial x} + f_1 h_2 \frac{\partial g_1}{\partial y} + f_2 h_2 \frac{\partial g_2}{\partial y} \\
& - f_1 g_1 \frac{\partial h_1}{\partial x} - f_2 g_1 \frac{\partial h_2}{\partial x} - f_1 g_2 \frac{\partial h_1}{\partial y} - f_2 g_2 \frac{\partial h_2}{\partial y}.
\end{aligned}
$$

which in turn gives us the global formula

$$
d\alpha(v, w) = v[\alpha(w)] - w[\alpha(v)] - \alpha([v, w]).
$$

Of course, at this point the last term in this global formula seems like simply a matter of notational convenience; we have no idea what it actually represents. The fact of the matter is that the third term genuinely does represent something that is geometrically interesting and important, which we now turn to explaining.

The definition of the Lie bracket of two vector fields v and w, which is denoted by $[v, w]$, is defined in terms of how those vector fields act on a function $F : \mathbb{R}^2 \to \mathbb{R}$. Given a vector field w defined on the manifold $\mathbb{R}^2$ we can find the directional derivative of F in the direction of w. At a particular point $p \in \mathbb{R}^2$ we have that $w_p[F]$ is a value telling how quickly F is changing in the w direction at that point p. But if we do not specify a point p the $w[F]$ is another function on the manifold $\mathbb{R}^2$. That is,

$$
\begin{aligned}
w[F] : \mathbb{R}^2 &\longrightarrow \mathbb{R} \\
p &\longmapsto w_p[F].
\end{aligned}
$$

We can again take the directional derivative of this function $w[F]$ in the direction of v. Again, at the point $p \in \mathbb{R}^2$ we have that $v_p\big[w[F]\big]$ is a value that tells how quickly $w[F]$ changes in the v direction at that point p. However, if we do not specify the point then $v\big[w[F]\big]$ is again another function on the manifold $\mathbb{R}^2$,

$$
\begin{aligned}
v\big[w[F]\big] : \mathbb{R}^2 &\longrightarrow \mathbb{R} \\
p &\longmapsto v_p\big[w[F]\big].
\end{aligned}
$$

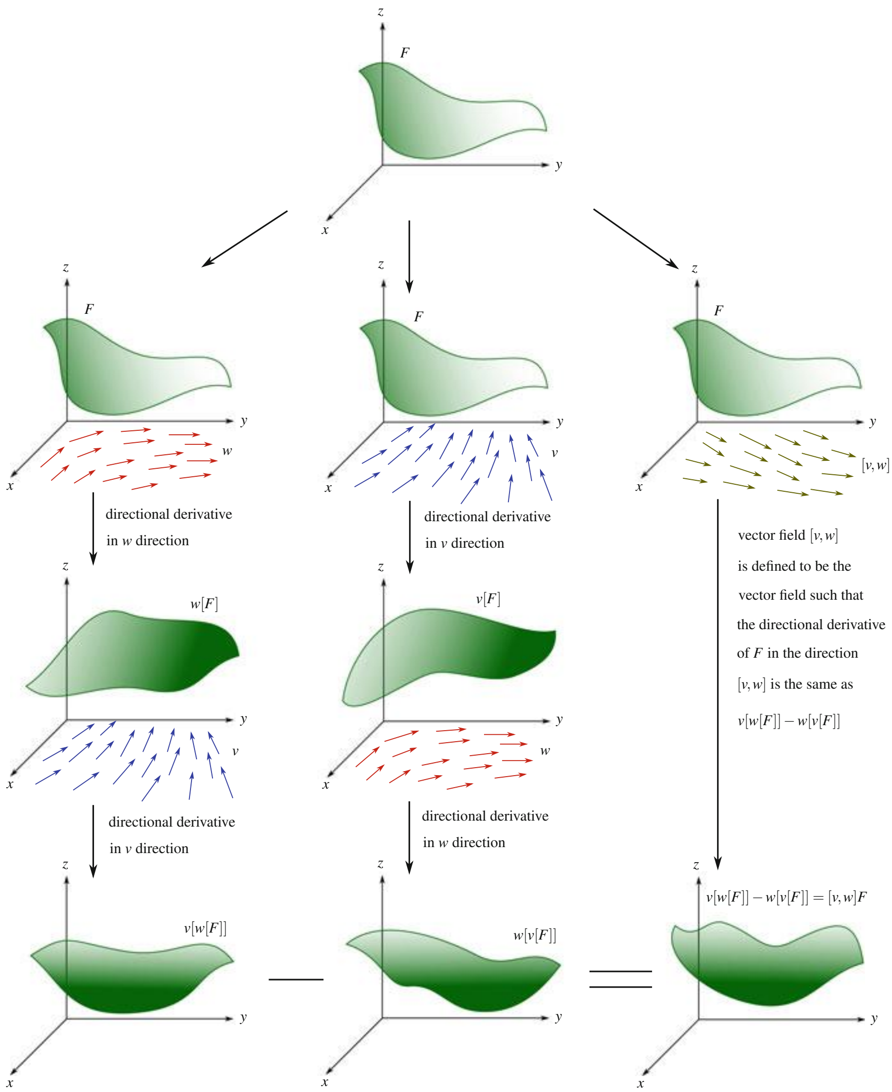


Fig. 4.3 On the left we show the directional derivative of $F : \mathbb{R}^2 \to \mathbb{R}$ taken in the direction of w to give the function $w[F]$, after which the directional derivative of $w[F]$ is taken in the direction of v to give the function $w\big[w[F]\big]$. Down the middle we show the directional derivative of F taken in the direction of v to give the function $v[F]$, after which the directional derivative of $v[F]$ is taken in the direction of w to give $w\big[v[F]\big]$. The difference $v\big[w[F]\big] - w\big[v[F]\big]$ of these two functions is taken to give a new function, shown along the bottom. The Lie bracket of v and w, denoted $[v, w]$, is the vector field which, when applied to the original function F, gives this new function. In other words, $[v, w]\big[F\big] = v\big[w[F]\big] - w\big[v[F]\big]$, as shown on the right

This process is all illustrated on the left side of Fig. 4.3. If we reverse the order of the vector fields and use v first we get the function $v[F] : \mathbb{R}^2 \to \mathbb{R}$. Applying w next we get the function $w\big[v[F]\big] : \mathbb{R}^2 \to \mathbb{R}$. This process is illustrated down the center of Fig. 4.3. Suppose we then took the difference of these two functions, as along the bottom of Fig. 4.3, then we would get a new function

$$v\big[w[F]\big] - w\big[v[F]\big] : \mathbb{R}^2 \longrightarrow \mathbb{R}.$$

It is in fact possible to find a single vector field that does this all in one step. In other words, we can find a single vector field which, when applied to F gives exactly this function $v\big[w[F]\big] - w\big[v[F]\big]$. That single vector field is called the Lie bracket of v and w and is denoted by $[v, w]$. Thus we have

$$[v, w]\big[F\big] = v\big[w[F]\big] - w\big[v[F]\big].$$

This is shown on the right side of Fig. 4.3. Despite the fact that the notation $[v, w]$ has both vector fields v and w in it, $[v, w]$ is simply a single vector field. Also notice that the square brackets are used in two different ways. However, by paying attention to what is inside the square brackets it should be easy to distinguish between the two uses. We will find $[v, w]$ computationally below, but we should note there is also another somewhat geometric way of viewing $[v, w]$. In fact, it turns out that $[v, w]$ is the Lie derivative of w in the direction v. The explanation for this is left until Sect. A.7 and Figs. A.5 and A.6.

Suppose we get rid of the function F in the notation to get

$$[v, w]\big[\,\cdot\,\big] = v\big[w[\cdot]\big] - w\big[v[\cdot]\big].$$

This explains the reason you will almost always see the Lie bracket of v and w somewhat lazily defined simply as

$$[v, w] = vw - wv.$$

Our goal now turns to finding out exactly what the vector field $[v, w]$ is,

$$\begin{aligned}
[v, w]F &= (vw - wv)F \\
&= v[w[F]] - w[v[F]] \\
&= v\left[h_1\frac{\partial F}{\partial x} + h_2\frac{\partial F}{\partial y}\right] - w\left[g_1\frac{\partial F}{\partial x} + g_2\frac{\partial F}{\partial y}\right] \\
&= (g_1\partial_x + g_2\partial_y)\left[h_1\frac{\partial F}{\partial x} + h_2\frac{\partial F}{\partial y}\right] - (h_1\partial_x + h_2\partial_y)\left[g_1\frac{\partial F}{\partial x} + g_2\frac{\partial F}{\partial y}\right] \\
&= g_1\frac{\partial}{\partial x}\left(h_1\frac{\partial F}{\partial x}\right) + g_1\frac{\partial}{\partial x}\left(h_2\frac{\partial F}{\partial y}\right) + g_2\frac{\partial}{\partial y}\left(h_1\frac{\partial F}{\partial x}\right) + g_2\frac{\partial}{\partial y}\left(h_2\frac{\partial F}{\partial y}\right) \\
&\quad - h_1\frac{\partial}{\partial x}\left(g_1\frac{\partial F}{\partial x}\right) - h_1\frac{\partial}{\partial x}\left(g_2\frac{\partial F}{\partial y}\right) - h_2\frac{\partial}{\partial y}\left(g_1\frac{\partial F}{\partial x}\right) - h_2\frac{\partial}{\partial y}\left(g_2\frac{\partial F}{\partial y}\right) \\
&\overset{\substack{\text{prod.}\\ \text{rule}}}{=} g_1\frac{\partial h_1}{\partial x}\frac{\partial F}{\partial x} + \cancel{g_1 h_1\frac{\partial}{\partial x}\left(\frac{\partial F}{\partial x}\right)} + g_1\frac{\partial h_2}{\partial x}\frac{\partial F}{\partial y} + \cancel{g_1 h_2\frac{\partial}{\partial x}\left(\frac{\partial F}{\partial y}\right)} \\
&\quad + g_2\frac{\partial h_1}{\partial y}\frac{\partial F}{\partial x} + \xcancel{g_2 h_1\frac{\partial}{\partial y}\left(\frac{\partial F}{\partial x}\right)} + g_2\frac{\partial h_2}{\partial y}\frac{\partial F}{\partial y} + \xcancel{g_2 h_2\frac{\partial}{\partial y}\left(\frac{\partial F}{\partial y}\right)} \\
&\quad - h_1\frac{\partial g_1}{\partial x}\frac{\partial F}{\partial x} - \cancel{h_1 g_1\frac{\partial}{\partial x}\left(\frac{\partial F}{\partial x}\right)} - h_1\frac{\partial g_2}{\partial x}\frac{\partial F}{\partial y} - \xcancel{h_1 g_2\frac{\partial}{\partial x}\left(\frac{\partial F}{\partial y}\right)} \\
&\quad - h_2\frac{\partial g_1}{\partial y}\frac{\partial F}{\partial x} - \cancel{h_2 g_1\frac{\partial}{\partial y}\left(\frac{\partial F}{\partial x}\right)} - h_2\frac{\partial g_2}{\partial y}\frac{\partial F}{\partial y} - \xcancel{h_2 g_2\frac{\partial}{\partial y}\left(\frac{\partial F}{\partial y}\right)} \\
&= g_1\frac{\partial h_1}{\partial x}\frac{\partial F}{\partial x} + g_1\frac{\partial h_2}{\partial x}\frac{\partial F}{\partial y} + g_2\frac{\partial h_1}{\partial y}\frac{\partial F}{\partial x} + g_2\frac{\partial h_2}{\partial y}\frac{\partial F}{\partial y} \\
&\quad - h_1\frac{\partial g_1}{\partial x}\frac{\partial F}{\partial x} - h_1\frac{\partial g_2}{\partial x}\frac{\partial F}{\partial y} - h_2\frac{\partial g_1}{\partial y}\frac{\partial F}{\partial x} - h_2\frac{\partial g_2}{\partial y}\frac{\partial F}{\partial y} \\
&= \left(\left(g_1\frac{\partial h_1}{\partial x} + g_2\frac{\partial h_1}{\partial y} - h_1\frac{\partial g_1}{\partial x} - h_2\frac{\partial g_1}{\partial y}\right)\frac{\partial}{\partial x}\right. \\
&\quad \left. + \left(g_1\frac{\partial h_2}{\partial x} + g_2\frac{\partial h_2}{\partial y} - h_1\frac{\partial g_2}{\partial x} - h_2\frac{\partial g_2}{\partial y}\right)\frac{\partial}{\partial y}\right)[F]
\end{aligned}$$

where the various terms cancel due to the equality of mixed patrials. Thus we have found the form the vector field $[v, w]$ takes, it is given by

$$[v, w] = \left(g_1\frac{\partial h_1}{\partial x} + g_2\frac{\partial h_1}{\partial y} - h_1\frac{\partial g_1}{\partial x} - h_2\frac{\partial g_1}{\partial y}\right)\partial_x$$

$$
\begin{aligned}
&+ \left(g_1 \frac{\partial h_2}{\partial x} + g_2 \frac{\partial h_2}{\partial y} - h_1 \frac{\partial g_2}{\partial x} - h_2 \frac{\partial g_2}{\partial y} \right) \partial_y \\
&= \begin{bmatrix} g_1 \frac{\partial h_1}{\partial x} + g_2 \frac{\partial h_1}{\partial y} - h_1 \frac{\partial g_1}{\partial x} - h_2 \frac{\partial g_1}{\partial y} \\ g_1 \frac{\partial h_2}{\partial x} + g_2 \frac{\partial h_2}{\partial y} - h_1 \frac{\partial g_2}{\partial x} - h_2 \frac{\partial g_2}{\partial y} \end{bmatrix}.
\end{aligned}
$$

Let us now see what α of the vector field $[v, w]$ is. We have

$$
\begin{aligned}
\alpha([v, w]) &= \langle \alpha, [v, w] \rangle \\
&= [f_1, f_2] \begin{bmatrix} g_1 \frac{\partial h_1}{\partial x} + g_2 \frac{\partial h_1}{\partial y} - h_1 \frac{\partial g_1}{\partial x} - h_2 \frac{\partial g_1}{\partial y} \\ g_1 \frac{\partial h_2}{\partial x} + g_2 \frac{\partial h_2}{\partial y} - h_1 \frac{\partial g_2}{\partial x} - h_2 \frac{\partial g_2}{\partial y} \end{bmatrix} \\
&= f_1 g_1 \frac{\partial h_1}{\partial x} + f_1 g_2 \frac{\partial h_1}{\partial y} - f_1 h_1 \frac{\partial g_1}{\partial x} - f_1 h_2 \frac{\partial g_1}{\partial y} \\
&\quad + f_2 g_1 \frac{\partial h_2}{\partial x} + f_2 g_2 \frac{\partial h_2}{\partial y} - f_2 h_1 \frac{\partial g_2}{\partial x} - f_2 h_2 \frac{\partial g_2}{\partial y}.
\end{aligned}
$$

Here notice the notation, $[v, w]$ is the lie bracket of two vector fields, which is itself a vector field while $[f_1, f_2]$ is the co-vector, or row vector, associated with the differential one-form $\alpha = f_1 dx + f_2 dy$. The way you tell these two mathematically distinct objects apart is by paying close attention to what is inside the brackets, vector fields or functions. But notice what we have gotten, these eight terms are exactly the negative of the eight terms left over from $v[\alpha(w)] - w[\alpha(v)]$. Thus, we define the global formula for the exterior derivative of a one-form α on the manifold $\mathbb{R}^2$ to be $d\alpha(v, w) = v[\alpha(w)] - w[\alpha(v)] - \alpha([v, w])$.

Of course, so far we have only done this for a one-form on the manifold $\mathbb{R}^2$. What about general one-form on the manifold $\mathbb{R}^n$? We will walk through the admittedly complicated calculations in this case as well, in order to see everything works out as expected. We start with

$$
\begin{aligned}
\alpha &= \sum \alpha_i dx_i, \\
v &= \sum g_i \partial_{x_i}, \\
w &= \sum h_i \partial_{x_i}.
\end{aligned}
$$

We will begin by computing $[v, w]$. Using a real-valued function F on the manifold, we first find

$$
\begin{aligned}
v\big[w[F]\big] &= v \left[\sum_i h_i \frac{\partial F}{\partial x_i} \right] \\
&= \left(\sum_j g_j \partial_{x_j} \right) \left[\sum_i h_i \frac{\partial F}{\partial x_i} \right] \\
&= \sum_j g_j \partial_{x_j} \left(\sum_i h_i \frac{\partial F}{\partial x_i} \right) \\
&= \sum_j g_j \sum_i \left(\frac{\partial h_i}{\partial x_j} \cdot \frac{\partial F}{\partial x_i} + h_i \frac{\partial}{\partial x_j} \left(\frac{\partial F}{\partial x_i} \right) \right) \\
&= \sum_j \sum_i \left(g_j \frac{\partial h_i}{\partial x_j} \cdot \frac{\partial F}{\partial x_i} + g_j h_i \frac{\partial}{\partial x_j} \left(\frac{\partial F}{\partial x_i} \right) \right).
\end{aligned}
$$

Finding the next term is similar and we get

$$
w\big[v[F]\big] = \sum_j \sum_i \left(h_j \frac{\partial g_i}{\partial x_j} \cdot \frac{\partial F}{\partial x_i} + h_j g_j \frac{\partial}{\partial x_j} \left(\frac{\partial F}{\partial x_i} \right) \right).
$$

Putting these together we get

$$
\begin{aligned}
v\big[w[F]\big]-w\big[v[F]\big] &= \sum_j\sum_i\left(g_j\frac{\partial h_i}{\partial x_j}\cdot\frac{\partial F}{\partial x_i}+g_jh_i\frac{\partial}{\partial x_j}\left(\frac{\partial F}{\partial x_i}\right)\right)-\sum_j\sum_i\left(h_j\frac{\partial g_i}{\partial x_j}\cdot\frac{\partial F}{\partial x_i}+h_jg_j\frac{\partial}{\partial x_j}\left(\frac{\partial F}{\partial x_i}\right)\right)\\
&= \sum_j\sum_i\left(g_j\frac{\partial h_i}{\partial x_j}\cdot\frac{\partial F}{\partial x_i}-h_j\frac{\partial g_i}{\partial x_j}\cdot\frac{\partial F}{\partial x_i}\right)+\underbrace{\sum_i\sum_j g_jh_i\frac{\partial}{\partial x_j}\left(\frac{\partial F}{\partial x_i}\right)-\underbrace{\sum_j\sum_i g_ih_j\frac{\partial}{\partial x_j}\left(\frac{\partial F}{\partial x_i}\right)}_{\text{switch dummy variables}}}_{=0 \text{ by equality of mixed partials}}\\
&= \left(\sum_i\left(\sum_j\left(g_j\frac{\partial h_i}{\partial x_j}-h_j\frac{\partial g_i}{\partial x_j}\right)\right)\frac{\partial}{\partial x_i}\right)F.
\end{aligned}
$$

Taking away the function F we get the actual vector field

$$
[v,w]=\sum_i\left(\sum_j\left(g_j\frac{\partial h_i}{\partial x_j}-h_j\frac{\partial g_i}{\partial x_j}\right)\right)\partial_{x_i}.
$$

Finally, we find

$$
\begin{aligned}
\alpha([v,w]) &= \left(\sum_k\alpha_k dx_k\right)\left(\sum_i\left(\sum_j\left(g_j\frac{\partial h_i}{\partial x_j}-h_j\frac{\partial g_i}{\partial x_j}\right)\right)\partial_{x_i}\right)\\
&= \sum_i\alpha_i\left(\sum_j\left(g_j\frac{\partial h_i}{\partial x_j}-h_j\frac{\partial g_i}{\partial x_j}\right)\right).
\end{aligned}
$$

Now we turn our attention to the expression $v[\alpha(w)]-w[\alpha(v)]-\alpha([v,w])$. Since $v[\alpha(w)]=d(\alpha(v))(w)=\big\langle d(\alpha(v)),w\big\rangle=\big\langle d\langle\alpha,v\rangle,w\big\rangle$ we start off with

$$
\begin{aligned}
\langle\alpha,v\rangle &= \left(\sum_i\alpha_i dx_i\right)\left(\sum_j g_j\partial_{x_j}\right)\\
&= \sum_i\alpha_i g_i.
\end{aligned}
$$

Taking the directional derivative we get

$$
\begin{aligned}
d\langle\alpha,v\rangle &= d\left(\sum_i\alpha_i g_i\right)\\
&= \sum_j\frac{\partial\left(\sum_i\alpha_i g_i\right)}{\partial x_j}dx_j\\
&= \sum_j\left(\sum_i\left(\frac{\partial\alpha_i}{\partial x_j}g_i+\alpha_i\frac{\partial g_i}{\partial x_j}\right)\right)dx_j,
\end{aligned}
$$

which we can then use to find

$$
w[\alpha(v)]=\sum_j h_j\left(\sum_i\left(\frac{\partial\alpha_i}{\partial x_j}g_i+\alpha_i\frac{\partial g_i}{\partial x_j}\right)\right).
$$

Similarly we have

$$v[\alpha(w)] = \sum_j g_j \left(\sum_i \left(\frac{\partial \alpha_i}{\partial x_j} h_i + \alpha_i \frac{\partial h_i}{\partial x_j} \right) \right).$$

Next, we combine everything

$$\begin{aligned}
d\alpha(v, w) &= v[\alpha(w)] - w[\alpha(v)] - \alpha([v, w]) \\
&= \sum_j g_j \left(\sum_i \left(\frac{\partial \alpha_i}{\partial x_j} h_i + \alpha_i \frac{\partial h_i}{\partial x_j} \right) \right) - \sum_j h_j \left(\sum_i \left(\frac{\partial \alpha_i}{\partial x_j} g_i + \alpha_i \frac{\partial g_i}{\partial x_j} \right) \right) - \sum_i \alpha_i \left(\sum_j \left(g_j \frac{\partial h_i}{\partial x_j} - h_j \frac{\partial g_i}{\partial x_j} \right) \right) \\
&= \sum_{i,j} g_j h_i \frac{\partial \alpha_i}{\partial x_j} + \sum_{i,j} g_j \alpha_i \frac{\partial h_i}{\partial x_j} - \sum_{i,j} h_j g_i \frac{\partial \alpha_i}{\partial x_j} - \sum_{i,j} h_j \alpha_i \frac{\partial g_i}{\partial x_j} - \sum_{i,j} \alpha_i g_j \frac{\partial h_i}{\partial x_j} + \sum_{i,j} \alpha_i h_j \frac{\partial g_i}{\partial x_j} \\
&= \sum_{i,j} g_j h_i \frac{\partial \alpha_i}{\partial x_j} - \sum_{i,j} h_j g_i \frac{\partial \alpha_i}{\partial x_j} \underbrace{+ \sum_{i,j} g_j \alpha_i \frac{\partial h_i}{\partial x_j} - \sum_{i,j} \alpha_i g_j \frac{\partial h_i}{\partial x_j}}_{=0} \underbrace{- \sum_{i,j} h_j \alpha_i \frac{\partial g_i}{\partial x_j} + \sum_{i,j} \alpha_i h_j \frac{\partial g_i}{\partial x_j}}_{=0} \\
&= \sum_{i,j} (g_j h_i - h_j g_i) \frac{\partial \alpha_i}{\partial x_j} \\
&= \sum_j \sum_i \frac{\partial \alpha_i}{\partial x_j} \begin{vmatrix} g_j & h_j \\ g_i & h_i \end{vmatrix} \\
&= \sum_j \sum_i \frac{\partial \alpha_i}{\partial x_j} \begin{vmatrix} dx_j(v) & dx_j(w) \\ dx_i(v) & dx_i(w) \end{vmatrix} \\
&= \sum_j \sum_i \frac{\partial \alpha_i}{\partial x_j} \, dx_j \wedge dx_i(v, w) \\
&= \sum_i \underbrace{\left(\sum_j \frac{\partial \alpha_i}{\partial x_j} dx_j \right)}_{=d\alpha_i} \wedge dx_i(v, w),
\end{aligned}$$

which is exactly the in-coordinates formula $d\alpha = \sum_i d\alpha_i \wedge dx_i$ that we want. Thus our global formula for a general one-form on the manifold $\mathbb{R}^n$ is actually what we want,

Global formula for exterior derivative of a one-form,	$d\alpha(v, w) = v[\alpha(w)] - w[\alpha(v)] - \alpha([v, w]).$

As in the case with constant vector fields, we will not make an attempt to provide a geometric argument for the global formula for the exterior derivative of a general k-form on the manifold $\mathbb{R}^n$. The process would be very computationally intensive and the geometry would not be at all clear. However, we now have all the components we need to have a general intuitive feel for why the formula is what it is. The exterior derivative of a k-form ω will be a $(k+1)$-form, which necessitates $k+1$ input vector fields $v_0, \ldots, v_k$. We start the labeling at 0 simply in order to have a nice clean $(-1)^{i+j}$ term,

Global formula for exterior derivative of k-form	$d\alpha(v_0, \ldots, v_k) = \sum_i (-1)^i v_i \big[\alpha(v_0, \ldots, \widehat{v_i}, \ldots, v_k)\big] + \sum_{i<j} (-1)^{i+j} \alpha\big([v_i, v_j], v_0, \ldots, \widehat{v_i}, \ldots, \widehat{v_j}, \ldots, v_k\big).$

As before, the hat in the notation means that those elements are omitted. This formula will actually be proved in Sect. A.7. The proof there will be entirely computational and require concepts and identities that will not be presented until that chapter.

Question 4.5 Show that in the case of a one-form the global formula for the exterior derivative of a k-form reduces to the global formula for the exterior derivative of a one-form.

4.5 Another Geometric Viewpoint

We begin this section by giving yet another definition for the exterior derivative of a k-form ω. We will then proceed to show that using this definition we arrive at exactly the same formula for $d\omega$ that we had in previous sections. Following that we will look at the geometric meaning of this definition. In many respects, this definition gives the most coherent and comprehensible geometric representation of the exterior derivative and it is a pity that more books do not emphasise this geometry. As mentioned in the introductory section, genuinely understanding this definition of the exterior derivative requires understanding integration of forms first. And generally integration is not introduced before differentiation. However, from a certain perspective differential forms can be viewed as "things one integrates." If differential forms are presented from this perspective integration would be discussed first.

Based on your understanding of integration from calculus you should be able to read this section to gain a general feel for the ideas and the underlying geometry. However, the various calculations rely on ideas, notations, and formulas developed and given in Chaps. 6–11 so you should not expect to understand the details. On a first read simply try to understand the big picture. We also highly recommend revisiting this section after Chap. 11.

We will motivate our new definition of the exterior derivative of a k-form ω by looking at the classical definition of the derivative of a function f. Recall that a function is nothing other than a zero-form. We have

$$\frac{df(x)}{dx} = \lim_{h\to 0}\frac{1}{h}\Big(f(x+h) - f(x)\Big).$$

We are evaluating f at the end points of the interval of length h, denoted by $hP = [x, x+h]$. Here the square brackets represent a closed interval on $\mathbb{R}$. A closed interval is an interval on $\mathbb{R}$ that includes the endpoints. For reasons explained in great detail in chapter 10 the boundary of the interval hP, which is denoted by $\partial(hP)$, is the point $\{(x+h)\}$ minus the point $\{(x)\}$. That is, $\partial(hP) = \partial[x, x+h] = \{(x+h)\} - \{(x)\}$. While this is something that we generally are not interested in calculus, and therefore generally is not done, we can define the integral of a function at a single point to be the function evaluated at that point. Thus

$$\int_{\{(x+h)\}} f \equiv f(x+h) \qquad \text{and} \qquad \int_{-\{(x)\}} f \equiv -f(x).$$

Notice what happens to the negative sign from $-\{(x)\}$. This allows us to rewrite our definition of the derivative of f as

$$\begin{aligned}
\frac{df(x)}{dx} &= \lim_{h\to 0}\frac{1}{h}\Big(f(x+h) - f(x)\Big)\\
&= \lim_{h\to 0}\frac{1}{h}\left(\int_{\{(x+h)\}} f + \int_{-\{(x)\}} f\right)\\
&= \lim_{h\to 0}\frac{1}{h}\int_{\{(x+h)\}-\{(x)\}} f\\
&= \lim_{h\to 0}\frac{1}{h}\int_{\partial(hP)} f.
\end{aligned}$$

Thus we have written the derivative of the zero-form f in terms of the integral of f evaluated at the boundary of the one-dimensional parallelepiped $[x, x+h]$. In calculus it would be absurd to use the formula that follows the final equality as a definition for the derivative of f, but for differential forms this definition leads to a clear understanding of what the exterior derivative means geometrically thus it makes sense to take this formula, applied to k-forms, as a definition. Thus we define for ω a k-form on the manifold $\mathbb{R}^n$, the exterior derivative of ω at the base point of the vectors $v_1, \ldots, v_{k+1}$ to be

Exterior derivative of a k-form	$$d\omega\big(v_1, \ldots, v_{k+1}\big) = \lim_{h \to 0} \frac{1}{h^{k+1}} \int_{\partial(hP)} \omega$$

where P is the parallelepiped spanned by the vectors $v_1, \ldots, v_{k+1}$, hP is that parallelepiped scaled by h, and $\partial(hP)$ is the boundary of the scaled parallelepiped. Again, the meaning of the boundary of a parallelepiped is explained in great detail in chapter 10.

Before exploring the geometric meaning of this definition we will proceed to show that this definition leads to exactly the same formula for $d\omega$ that we would expect from earlier sections. In showing this we are forced to use a number of concepts, notations, and formulas that are actually derived in later chapters. We also have to make use of a few ideas and techniques that are not in the purview of the book and which we therefore present without explanation or a great deal of rigor. The Taylor series of a function is one of these ideas. In the below we will simply assume that the Taylor series converges to the given function f in some neighborhood of the point a.

Suppose $f : \mathbb{R}^n \to \mathbb{R}$. The Taylor series of f about the point $a = (a_1, \ldots, a_n)$ is given by

$$\begin{aligned}
f(x_1, \ldots, x_n) &= \sum_{i_1=0}^{\infty} \cdots \sum_{i_n=0}^{\infty} \frac{(x_1 - a_1)^{i_1} \cdots (x_n - a_n)^{i_n}}{i_1! \cdots i_n!} \left(\frac{\partial^{i_1+\cdots+i_n} f}{\partial^{i_1} x_1 \cdots \partial^{i_n} x_n} \right)\Bigg|_{(a_1,\ldots,a_n)} \\
&= f(a_1, \ldots, a_n) + \sum_{j=1}^{n} (x_j - a_j) \left.\frac{\partial f}{\partial x_j}\right|_{(a_1,\ldots,a_n)} \\
&\quad + \frac{1}{2!} \sum_{j=0}^{n} \sum_{k=0}^{n} (x_j - a_j)(x_k - a_k) \left.\frac{\partial^2 f}{\partial x_j \partial x_k}\right|_{(a_1,\ldots,a_n)} \\
&\quad + \frac{1}{3!} \sum_{j=0}^{n} \sum_{k=0}^{n} \sum_{\ell=0}^{n} (x_j - a_j)(x_k - a_k)(x_\ell - a_\ell) \left.\frac{\partial^3 f}{\partial x_j \partial x_k \partial x_\ell}\right|_{(a_1,\ldots,a_n)} \\
&\quad + \cdots
\end{aligned}$$

In essence, for points $(x_1, \ldots, x_n)$ that are sufficiently close to the point $(a_1, \ldots, a_n)$ we can rewrite f as the above sum.

Question 4.6 Show that these two representations of the Taylor series of f are identical.

In finding the formula for $d\omega$ we will make two simplifying adjustments. First, we will write the Taylor series for f at the origin, hence we have $(a_1, \ldots, a_n) = (0, \ldots, 0)$. Second, we will group all of the second order terms and higher into a remainder function which we will simply write as R. Hence our Taylor series can be rewritten as

$$f(x_1, \ldots, x_n) = f(0, \ldots, 0) + \sum_{j=1}^{n} x_j \left.\frac{\partial f}{\partial x_j}\right|_{(0,\ldots,0)} + R(x_1, \ldots, x_n).$$

Next, we will identify the manifold $\mathbb{R}^n$ with the vector space $\mathbb{R}^n$. This will mean that the following argument will only apply to manifolds that are also vector spaces, though it is possible, with the appropriate technical details that we will not concern ourselves with here, to extend the results to general n-dimensional manifolds. Supplying these details will take us too far away from the big picture. Using our identification we have

$$\text{Manifold } \mathbb{R}^n \longleftrightarrow \text{Vector Space } \mathbb{R}^n$$

$$p = (x_1, \ldots, x_n) \Longleftrightarrow v = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}.$$

Notice that the second term of the Taylor series can be written as

$$\sum_{j=1}^{n} x_j \left.\frac{\partial f}{\partial x_j}\right|_{(0,\ldots,0)} = \underbrace{\left[\left.\frac{\partial f}{\partial x_1}\right|_{(0,\ldots,0)}, \cdots, \left.\frac{\partial f}{\partial x_n}\right|_{(0,\ldots,0)} \right]}_{\text{co-vector } df_0} \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} = df_0(v),$$

where df is the usual differential of f and df_0 is the differential of f taken at the origin $(0,\ldots,0) = 0$. With this we can write the Taylor series at the "point" $(x_1,\ldots,x_n) = v$ as

$$f(v) = f(0) + df_0(v) + R(v).$$

Now, suppose we are given the vectors $v_1,\ldots,v_{k+1}$. We will label the parallelepiped spanned by these vectors as P. Here the word span is being used differently than in linear algebra. Here we have

$$P = \operatorname{span}\Big\{v_1,\ldots,v_{k+1}\Big\} = \Big\{ t_1v_1 + \cdots + t_{k+1}v_{k+1} \ \Big|\ 0 \le t_i \le 1, i = 1,\ldots,k+1 \Big\}.$$

We are still identifying the manifold $\mathbb{R}^n$ with the vector space $\mathbb{R}^n$, thus the vector $t_1v_1 + \cdots + t_{k+1}v_{k+1}$ really means the point given by the vector's coordinates as per the identity above. We can scale this parallelepiped by the factor h to get a new parallelepiped

$$hP = \operatorname{span}\Big\{hv_1,\ldots,hv_{k+1}\Big\} = \Big\{ t_1v_1 + \cdots + t_{k+1}v_{k+1} \ \Big|\ 0 \le t_i \le h, i = 1,\ldots,k+1 \Big\}.$$

The various faces of hP are given by

$$hP_{(\ell,0)} = \Big\{ t_1v_1 + \cdots + t_{\ell-1}v_{\ell-1} + 0v_\ell + t_\ell v_{\ell+1} + \cdots + t_k v_{k+1} \ \Big|\ 0 \le t_i \le h, i = 1,\ldots,k \Big\}$$

or

$$hP_{(\ell,1)} = \Big\{ t_1v_1 + \cdots + t_{\ell-1}v_{\ell-1} + hv_\ell + t_\ell v_{\ell+1} + \cdots + t_k v_{k+1} \ \Big|\ 0 \le t_i \le h, i = 1,\ldots,k \Big\}$$

where $\ell = 1,\ldots,k+1$. Notice the labeling of the ts. In the faces of hP we now have t_1 through t_k. Continuing to write points as vectors, points of the faces $hP_{(\ell,1)}$ and $hP_{(\ell,0)}$ can be written as

$$\begin{aligned} & v_{(\ell,1)} = hv_\ell + t_1v_1 + \cdots + t_{\ell-1}v_{\ell-1} + v_\ell v_{\ell+1} + \cdots + t_k v_{k+1} \\ \text{and}\quad & v_{(\ell,0)} = \qquad t_1v_1 + \cdots + t_{\ell-1}v_{\ell-1} + t_\ell v_{\ell+1} + \cdots + t_k v_{k+1} \end{aligned}$$

respectively. Defining the scaled k-cube hI^k to be

$$hI^k = \Big\{ (t_1,\ldots,t_k) \ \Big|\ 0 \le t_i \le h, i = 1,\ldots,k \Big\},$$

as long as we know what the vectors $v_1,\ldots,v_{k+1}$ are, for each ℓ and any value of h we have a natural identification between each of the faces $hP_{(\ell,1)}$ and $hP_{(\ell,0)}$ to hI^k,

$$\begin{aligned} hP_{(\ell,1)} &\longleftrightarrow hI^k \\ v_{(\ell,1)} &\longleftrightarrow (t_1,\ldots,t_k) \end{aligned}$$

and

$$\begin{aligned} hP_{(\ell,0)} &\longleftrightarrow hI^k \\ v_{(\ell,0)} &\longleftrightarrow (t_1,\ldots,t_k). \end{aligned}$$

We are now going to focus in on the $hP_{(\ell,1)} \to hI^k$ and examine it closely. What we do in this particular case also holds for $hP_{(\ell,0)}$ and for all ℓ. We begin by naming the mapping Φ, thus we have

$$\begin{aligned} hP_{(\ell,1)} &\xrightarrow{\Phi} hI^k \\ v_{(\ell,1)} &\xmapsto{\Phi} (t_1, \ldots, t_k). \end{aligned}$$

Obviously Φ is invertible so we also have

$$\begin{aligned} hP_{(\ell,1)} &\xleftarrow{\Phi^{-1}} hI^k \\ v_{(\ell,1)} &\xleftarrow{\Phi^{-1}} (t_1, \ldots, t_k). \end{aligned}$$

If we are given a point $t = (t_1, \ldots, t_k) \in hI^k$ then we can write $\Phi^{-1}(t) = v_{(\ell,1)}(t)$. In other words, the point $v_{(\ell,1)}$ is written as a function of the point t.

Given the mapping Φ we also have the tangent mapping, also called the push-forward, $T\Phi$. When we are considering the tangent mapping then we are actually viewing the elements $v_{(\ell,1)}$ as vectors. The associated tangent mapping of Φ is

$$\begin{aligned} T(hP_{(\ell,1)}) &\xrightarrow{T\Phi} T(hI^k) \\ v_{(\ell,1)} &\longmapsto T\Phi \cdot v_{(\ell,1)} = \begin{bmatrix} t_1 \\ \vdots \\ t_k \end{bmatrix}. \end{aligned}$$

In particular we are interested in what happens to the original vectors $v_1, \ldots, v_{k+1}$ that were used to define the parallelepiped P,

$$\underbrace{v_m}_{m \neq \ell} \longmapsto T\Phi \cdot v_m = \begin{bmatrix} 0 \\ \vdots \\ 1 \\ 0 \\ \vdots \end{bmatrix}.$$

For $T\Phi \cdot v_m$ we have the 1 in the mth slot if $m < \ell$ and the $(m-1)$th slot if $m > \ell$.

Question 4.7 Find $T\Phi^{-1} \cdot e_i$, $i = 1, \ldots, k$, where e_i are the standard Euclidian vectors.

We also have the cotangent map, also call the pull-back, of Φ^{-1},

$$\begin{aligned} \bigwedge\nolimits^k \big(hP_{(\ell,1)}\big) &\xrightarrow{T^*\Phi^{-1}} \bigwedge\nolimits^k \big(hI^k\big) \\ \alpha &\longmapsto T^*\Phi^{-1} \cdot \alpha. \end{aligned}$$

In particular, we are interested in the k-forms on $hP_{(\ell,1)}$. Note, $hP_{(\ell,1)}$ is a k-dimensional submanifold of $\mathbb{R}^{k+1}$. Since $\mathbb{R}^{k+1}$ already has the Cartesian coordinate functions $x_1, \ldots, x_{k+1}$ we will simply use these as the coordinate functions of the embedded submanifold $hP_{(\ell,1)}$. This may seem slightly odd since the manifold $hP_{(\ell,1)}$ is k-dimensional yet we are using $k+1$ coordinate functions on it. But this is the simplest approach since the manifold in question is simply a parallelepiped embedded in Euclidian space, which we understand quite well. Also, we will use the $t_1, \ldots, t_k$ as the coordinate functions on hI^k. In essence the t_is are the Cartesian coordinate functions on $\mathbb{R}^k$, which hI^k is embedded in. We begin by looking at k-form basis elements on $hP_{(\ell,1)}$,

$$dx_{i_1} \wedge \cdots \wedge dx_{i_k}.$$

Since both $hP_{(\ell,1)}$ and hI^k are k-dimensional then we know that $T^*\Phi^{-1} \cdot \big(dx_{i_1} \wedge \cdots \wedge dx_{i_k}\big)$ must be of the form $F dt_1 \wedge \cdots \wedge dt_k$. Notice that the dimension of $\bigwedge^k \big(hI^k\big)$ is 1 so $dt_1 \wedge \cdots \wedge dt_k$ is the basis for this space and is hence the volume form on hI^k. We want to find F.

$$
\begin{aligned}
F &= F dt_1 \wedge \cdots \wedge dt_k(e_1, \ldots, e_k) \\
&= T^*\Phi^{-1} \cdot \big(dx_{i_1} \wedge \cdots \wedge dx_{i_k}\big)(e_1, \ldots, e_k) \\
&= dx_{i_1} \wedge \cdots \wedge dx_{i_k} \left(\underbrace{T\Phi^{-1} \cdot e_1}_{v_1}, \ldots, \underbrace{T\Phi^{-1} \cdot e_{\ell-1}}_{v_{\ell-1}}, \underbrace{T\Phi^{-1} \cdot e_\ell}_{v_{\ell+1}}, \ldots, \underbrace{T\Phi^{-1} \cdot e_1}_{v_{k+1}} \right) \\
&= dx_{i_1} \wedge \cdots \wedge dx_{i_k} \left(v_1, \ldots, \widehat{v_\ell}, \ldots, v_{k+1}\right).
\end{aligned}
$$

In other words, we have that $F = dx_{i_1} \wedge \cdots \wedge dx_{i_k} \left(v_1, \ldots, \widehat{v_\ell}, \ldots, v_{k+1}\right)$ and hence that

$$
T^*\Phi^{-1} \cdot \big(dx_{i_1} \wedge \cdots \wedge dx_{i_k}\big) = dx_{i_1} \wedge \cdots \wedge dx_{i_k} \left(v_1, \ldots, \widehat{v_\ell}, \ldots, v_{k+1}\right) dt_1 \wedge \cdots \wedge dt_k.
$$

In Chap. 7 we develop the following formula for integration that involves a change of variables,

$$
\int_R f(x_1, \ldots, x_n)\, dx_1 \wedge \cdots \wedge dx_n = \int_{\Phi(R)} f \circ \Phi^{-1}(\Phi_1, \ldots, \Phi_n)\, T^*\Phi^{-1} \cdot \big(dx_1 \wedge \cdots \wedge dx_n\big).
$$

We will now consider how this formula applies to the case where the region being integrated over is $hP_{(\ell,1)}$ and the change of variable mapping is $\Phi : hP_{(\ell,1)} \to hI^k$. Writing the point $(x_1, \ldots, x_k)$ as $v_{(\ell,1)}$ and the point $(t_1, \ldots, t_k)$ as t, recalling that $\Phi^{-1}(t) = v_{(\ell,1)}(t)$, and writing dx^I for $dx_{i_1} \wedge \cdots \wedge dx_{i_k}$, we have

$$
\begin{aligned}
&\int_{hP_{(\ell,1)}} f(v_{(\ell,1)})\, dx_{i_1} \wedge \cdots \wedge dx_{i_k} \\
&= \int_{\Phi(hP_{(\ell,1)})} f \circ \Phi^{-1}(t)\, T^*\Phi^{-1} \cdot \big(dx_{i_1} \wedge \cdots \wedge dx_{i_k}\big) \\
&= \int_{hI^k} f\big(v_{(\ell,1)}(t)\big)\, dx^I(v_1, \ldots, \widehat{v_\ell}, \ldots, v_{k+1})\, dt_1 \wedge \cdots \wedge dt_k \\
&= \underbrace{\int_0^1 \cdots \int_0^1}_{k \text{ times}} f\big(v_{(\ell,1)}(t)\big)\, dx^I(v_1, \ldots, \widehat{v_\ell}, \ldots, v_{k+1})\, \underbrace{dt \cdots dt}_{k \text{ times}}.
\end{aligned}
$$

The computation for the region $hP_{(\ell,0)}$ and for all other values of ℓ are exactly analogous.

Now we have all the ingredients in place to consider the definition of the exterior derivative of ω that was given at the beginning of this section. For ω a k-form on the manifold $\mathbb{R}^n$, the exterior derivative of ω at the base point of the vectors $v_1, \ldots, v_{k+1}$ is defined to be

$$
d\omega\big(v_1, \ldots, v_{k+1}\big) = \lim_{h \to 0} \frac{1}{h^{k+1}} \int_{\partial(hP)} \omega.
$$

To simplify our computations we have already assumed this base point is the origin of the manifold $\mathbb{R}^n$ when we discussed the Taylor series expansion. If the point we were interested in were not the origin we would have to translate the k-form so the point we were interested in coincided with the origin, do the necessary computation, and then translate back. This would add a layer of complexity to the argument we are about to present, but would not fundamentally change anything. In essence we are simplifying our argument by only looking at a single special case which can, with a little extra work, be extended to the general case. The phrase that one often uses in mathematics in situations like this is *without loss of generality*. Thus we will say without loss of generality assume the base point is the origin.

Consider the unit $(k+1)$-cube I^{k+1}. In the Chap. 11 we find the boundary I^{k+1} to be

$$\partial I^{k+1} = \sum_{i=1}^{k+1}\sum_{a=0}^{1}(-1)^{i+a} I^{k+1}_{(i,a)}$$

where $I^{k+1}_{(i,a)}$ is the face of I^{k+1} obtained by holding the ith variable fixed at a, which is either 0 or 1. The $(-1)^{i+a}$ indicate the orientation of that face. This way the orientations of faces opposite each other are opposite. The boundary of any parallelepiped can be found similarly, thus we have

$$\begin{aligned}\partial(hP) &= \sum_{\ell=1}^{k+1}\sum_{a=0}^{1}(-1)^{\ell+a} hP_{(\ell,a)}\\ &= \sum_{\ell=1}^{k+1}\Big((-1)^{\ell} hP_{(\ell,0)} + (-1)^{\ell+1} hP_{(\ell,1)}\Big).\end{aligned}$$

Finally, recall the Taylor series for a function f about the origin. For points v close to the origin we have $f(v) = f(0) + df_0(v) + R(v)$. Now suppose we were given the k-form $\omega = f dx_{i_1} \wedge \cdots \wedge dx_{i_k}$ then the integral on the right hand side of our definition of the exterior derivative of ω can be written as

$$\begin{aligned}\int_{\partial(hP)} \omega &= \int_{\partial(hP)} f\ dx_{i_1} \wedge \cdots \wedge dx_{i_k}\\ &= \int_{\sum_{\ell=1}^{k+1}\left((-1)^{\ell} hP_{(\ell,0)} + (-1)^{\ell+1} hP_{(\ell,1)}\right)} f\ \underbrace{dx_{i_1} \wedge \cdots \wedge dx_{i_k}}_{dx^I}\\ &= \sum_{\ell=1}^{k+1}\left((-1)^{\ell}\int_{hP_{(\ell,0)}} f\ dx^I + (-1)^{\ell+1}\int_{hP_{(\ell,1)}} f\ dx^I\right)\\ &= \sum_{\ell=1}^{k+1}(-1)^{\ell-1}\left(\int_{hP_{(\ell,1)}} f\ dx^I - \int_{hP_{(\ell,0)}} f\ dx^I\right)\\ &= \sum_{\ell=1}^{k+1}(-1)^{\ell-1}\left(\int_{hP_{(\ell,1)}} \Big(f(0) + df_0(v) + R(v)\Big)\, dx^I - \int_{hP_{(\ell,0)}} \Big(f(0) + df_0(v) + R(v)\Big)\, dx^I\right)\\ &= \sum_{\ell=1}^{k+1}(-1)^{\ell-1}\left(\int_{hP_{(\ell,1)}} f(0)\ dx^I - \int_{hP_{(\ell,0)}} f(0)\ dx^I \right. && (1)\\ &\qquad + \int_{hP_{(\ell,1)}} df_0(v)\ dx^I - \int_{hP_{(\ell,0)}} df_0(v)\ dx^I && (2)\\ &\qquad \left. + \int_{hP_{(\ell,1)}} R(v)\ dx^I - \int_{hP_{(\ell,0)}} R(v)\ dx^I \right). && (3)\end{aligned}$$

Consider the first and second terms that appear in (1). For each $\ell = 1, \ldots, k+1$ we have

$$\begin{aligned}&\int_{hP_{(\ell,1)}} f(0)\ dx^I - \int_{hP_{(\ell,0)}} f(0)\ dx^I\\ &\quad = \int_{hI^k} f(0) dx^I(v_1, \ldots, \widehat{v_\ell}, \ldots, v_{k+1})\ dt_1 \wedge \cdots \wedge dt_k\\ &\qquad - \int_{hI^k} f(0) dx^I(v_1, \ldots, \widehat{v_\ell}, \ldots, v_{k+1})\ dt_1 \wedge \cdots \wedge dt_k\end{aligned}$$

$$= \int_{hI^k} \underbrace{\Big(\underbrace{f(0)}_{\text{a constant}} - \underbrace{f(0)}_{\text{the same constant}} \Big)}_{=0} dx^I(v_1, \ldots, \widehat{v_\ell}, \ldots, v_{k+1})\, dt_1 \wedge \cdots \wedge dt_k$$

$$= 0.$$

Now consider the third and fourth terms that appear in (2). For each $\ell = 1, \ldots, k+1$ we have

$$\begin{aligned}
&\int_{hP_{(\ell,1)}} df_0(v)\, dx^I - \int_{hP_{(\ell,0)}} df_0(v)\, dx^I \\
&\quad = \int_{hI^k} df_0\big(v_{(\ell,1)}(t)\big) dx^I(v_1, \ldots, \widehat{v_\ell}, \ldots, v_{k+1})\, dt_1 \wedge \cdots \wedge dt_k \\
&\qquad - \int_{hI^k} df_0\big(v_{(\ell,0)}(t)\big) dx^I(v_1, \ldots, \widehat{v_\ell}, \ldots, v_{k+1})\, dt_1 \wedge \cdots \wedge dt_k \\
&\quad = \int_{hI^k} \Big(df_0\big(v_{(\ell,1)}(t)\big) - df_0\big(v_{(\ell,0)}(t)\big) \Big) dx^I(v_1, \ldots, \widehat{v_\ell}, \ldots, v_{k+1})\, dt_1 \wedge \cdots \wedge dt_k \\
&\quad = \int_{hI^k} \Big(df_0\big(hv_\ell + v_{(\ell,0)}(t)\big) - df_0\big(v_{(\ell,0)}(t)\big) \Big) dx^I(v_1, \ldots, \widehat{v_\ell}, \ldots, v_{k+1})\, dt_1 \wedge \cdots \wedge dt_k \\
&\quad = \int_{hI^k} \Big(df_0\big(hv_\ell\big) + \underbrace{df_0\big(v_{(\ell,0)}(t)\big) - df_0\big(v_{(\ell,0)}(t)\big)}_{=0} \Big) dx^I(v_1, \ldots, \widehat{v_\ell}, \ldots, v_{k+1})\, dt_1 \wedge \cdots \wedge dt_k \\
&\quad = \int_{hI^k} \underbrace{df_0\big(hv_\ell\big)\, dx^I(v_1, \ldots, \widehat{v_\ell}, \ldots, v_{k+1})}_{\text{does not depend on } t \in hI^k}\, dt_1 \wedge \cdots \wedge dt_k \\
&\quad = df_0\big(hv_\ell\big)\, dx^I(v_1, \ldots, \widehat{v_\ell}, \ldots, v_{k+1}) \int_{hI^k} dt_1 \wedge \cdots \wedge dt_k \\
&\quad = df_0\big(hv_\ell\big)\, dx^I(v_1, \ldots, \widehat{v_\ell}, \ldots, v_{k+1}) \underbrace{\int_0^h \cdots \int_0^h}_{k \text{ times}} \underbrace{dt \cdots dt}_{k \text{ times}} \\
&\quad = h\, df_0\big(v_\ell\big)\, dx^I(v_1, \ldots, \widehat{v_\ell}, \ldots, v_{k+1}) \cdot h^k \\
&\quad = h^{k+1} df_0\big(v_\ell\big)\, dx^I(v_1, \ldots, \widehat{v_\ell}, \ldots, v_{k+1}).
\end{aligned}$$

And now we consider the fifth and six terms that appear in (3),

$$\begin{aligned}
&\int_{hP_{(\ell,1)}} R(v)\, dx^I - \int_{hP_{(\ell,0)}} R(v)\, dx^I \\
&\quad = \int_{hI^k} R\big(v_{(\ell,1)}(t)\big)\, dx^I(v_1, \ldots, \widehat{v_\ell}, \ldots, v_{k+1})\, dt_1 \wedge \cdots \wedge dt_k \\
&\qquad - \int_{hI^k} R\big(v_{(\ell,0)}(t)\big)\, dx^I(v_1, \ldots, \widehat{v_\ell}, \ldots, v_{k+1})\, dt_1 \wedge \cdots \wedge dt_k \\
&\quad = \int_{hI^k} \Big(R\big(v_{(\ell,1)}(t)\big) - R\big(v_{(\ell,0)}(t)\big) \Big)\, dx^I(v_1, \ldots, \widehat{v_\ell}, \ldots, v_{k+1})\, dt_1 \wedge \cdots \wedge dt_k.
\end{aligned}$$

We will only sketch the next part of the argument in enough detail to give you an intuitive understanding. To fill in the details of the argument would require some background in analysis, something we are not assuming in this book. Consider what the remainder term actually is in this case,

$$\begin{aligned}
R(x_1, \dots, x_n) &= \frac{1}{2!} \sum_{j=0}^{n} \sum_{k=0}^{n} (x_j)(x_k) \left. \frac{\partial^2 f}{\partial x_j \partial x_k} \right|_{(0,\dots,0)} \\
&\quad + \frac{1}{3!} \sum_{j=0}^{n} \sum_{k=0}^{n} \sum_{m=0}^{n} (x_j)(x_k)(x_m) \left. \frac{\partial^3 f}{\partial x_j \partial x_k \partial x_m} \right|_{(0,\dots,0)} \\
&\quad + \cdots .
\end{aligned}$$

We are writing the point $(x_1, \dots, x_n)$ as a vector, which is either

$$\begin{aligned}
&v_{(\ell,1)} = hv_\ell + t_1 v_1 + \cdots + t_{\ell-1} v_{\ell-1} + v_\ell v_{\ell+1} + \cdots + t_k v_{k+1} \\
\text{or} \quad &v_{(\ell,0)} = \phantom{hv_\ell +{}} t_1 v_1 + \cdots + t_{\ell-1} v_{\ell-1} + t_\ell v_{\ell+1} + \cdots + t_k v_{k+1}.
\end{aligned}$$

There are $k+1$ vectors each consisting of n components. Take the absolute value of these $n(k+1)$ components and find the maximum value and call it M. The absolute value of each component in the vectors $v_{(\ell,1)}$ or $v_{(\ell,0)}$ is less than $h(k+1)M$, so we have

$$\begin{aligned}
\left| R(x_1, \dots, x_n) \right| &\le \frac{1}{2!} \sum_{j=0}^{n} \sum_{k=0}^{n} \left| \big(h(k+1)M\big)\big(h(k+1)M\big) \left. \frac{\partial^2 f}{\partial x_j \partial x_k} \right|_{(0,\dots,0)} \right| \\
&\quad + \frac{1}{3!} \sum_{j=0}^{n} \sum_{k=0}^{n} \sum_{m=0}^{n} \left| \big(h(k+1)M\big)\big(h(k+1)M\big)\big(h(k+1)M\big) \left. \frac{\partial^3 f}{\partial x_j \partial x_k \partial x_m} \right|_{(0,\dots,0)} \right| \\
&\quad + \cdots .
\end{aligned}$$

Notice that we can pull out at least an h^2 from every term in this sum. It requires some effort to prove, which we will not go to, but there exists some number K such that

$$\left| R(x_1, \dots, x_n) \right| \le h^2 K.$$

Using the triangle inequality we have

$$\begin{aligned}
\left| R\big(v_{(\ell,1)}(t)\big) - R\big(v_{(\ell,0)}(t)\big) \right| &\le \left| R\big(v_{(\ell,1)}(t)\big) \right| + \left| R\big(v_{(\ell,0)}(t)\big) \right| \\
&\le h^2 K + h^2 K \\
&= 2h^2 K.
\end{aligned}$$

Similarly, for some value of C we have

$$\begin{aligned}
\left| dx^I(v_1, \dots, \widehat{v_\ell}, \dots, v_{k+1}) \right| &= \left| \sum_{\sigma \in S_k} sgn(\sigma) \prod_{\substack{j=1 \\ j \ne \ell}}^{k+1} dx_{\sigma(i_j)}(v_j) \right| \\
&\le \sum_{\sigma \in S_k} \prod_{\substack{j=1 \\ j \ne \ell}}^{k+1} \left| dx_{\sigma(i_j)}(v_j) \right| \\
&\le C.
\end{aligned}$$

Putting this together

$$
\begin{aligned}
&\left| \int_{hI^k} \Big(R\big(v_{(\ell,1)}(t)\big) - R\big(v_{(\ell,0)}(t)\big) \Big) \, dx^I(v_1, \ldots, \widehat{v_\ell}, \ldots, v_{k+1}) \, dt_1 \wedge \cdots \wedge dt_k \right| \\
&\quad \leq \int_{hI^k} \big(2h^2 K\big) C \, dt_1 \wedge \cdots \wedge dt_k \\
&\quad = h^2 \big(2KC\big) \underbrace{\int_0^h \cdots \int_0^h}_{k \text{ times}} \underbrace{dt \cdots dt}_{k \text{ times}} \\
&\quad = h^{k+2} \big(2KC\big).
\end{aligned}
$$

Returning to our original calculation where we had left off,

$$
\begin{aligned}
\int_{\partial(hP)} \omega = \sum_{\ell=1}^{k+1} (-1)^{\ell-1} \Bigg(&\int_{hP_{(\ell,1)}} f(0) \, dx^I - \int_{hP_{(\ell,0)}} f(0) \, dx^I \\
&+ \int_{hP_{(\ell,1)}} df_0(v) \, dx^I - \int_{hP_{(\ell,0)}} df_0(v) \, dx^I \\
&+ \int_{hP_{(\ell,1)}} R(v) \, dx^I - \int_{hP_{(\ell,0)}} R(v) \, dx^I \Bigg) \\
\text{``}=\text{''} \sum_{\ell=1}^{k+1} (-1)^{\ell-1} \Bigg(&\, 0 \\
&+ h^{k+1} df_0\big(v_\ell\big) \, dx^I(v_1, \ldots, \widehat{v_\ell}, \ldots, v_{k+1}) \\
&+ h^{k+2} \big(2KC\big) \Bigg).
\end{aligned}
$$

Technically, the last equality is not an equality because of how we obtained the $h^{k+2}\big(2KC\big)$ term, but as we will see in a moment this does not matter. Using this in our definition of $d\omega$ we have

$$
\begin{aligned}
&d\omega\big(v_1, \ldots, v_{k+1}\big) \\
&\quad = \lim_{h \to 0} \frac{1}{h^{k+1}} \int_{\partial(hP)} \omega \\
&\quad = \lim_{h \to 0} \frac{1}{h^{k+1}} \left(\sum_{\ell=1}^{k+1} (-1)^{\ell-1} \Big(0 + h^{k+1} df_0\big(v_\ell\big) \, dx^I(v_1, \ldots, \widehat{v_\ell}, \ldots, v_{k+1}) + h^{k+2}\big(2KC\big) \Big) \right) \\
&\quad = \underbrace{\lim_{h \to 0} \left(\sum_{\ell=1}^{k+1} (-1)^{\ell-1} \frac{0}{h^{k+1}} \right)}_{=0} + \underbrace{\lim_{h \to 0} \left(\sum_{\ell=1}^{k+1} (-1)^{\ell-1} \cancelto{1}{\frac{h^{k+1}}{h^{k+1}}} df_0\big(v_\ell\big) \, dx^I(v_1, \ldots, \widehat{v_\ell}, \ldots, v_{k+1}) \right)}_{\text{no } h} \\
&\qquad + \underbrace{\lim_{h \to 0} \left(\sum_{\ell=1}^{k+1} (-1)^{\ell-1} \frac{h^{k+2}}{h^{k+1}} \big(2KC\big) \right)}_{\longrightarrow 0}
\end{aligned}
$$

$$
\begin{aligned}
&= \sum_{\ell=1}^{k+1} (-1)^{\ell-1} df_0\big(v_\ell\big)\, dx_{i_1} \wedge \cdots \wedge dx_{i_k}(v_1, \ldots, \widehat{v_\ell}, \ldots, v_{k+1}) \\
&= df \wedge dx_{i_1} \wedge \cdots \wedge dx_{i_k}(v_1, \ldots, v_{k+1}),
\end{aligned}
$$

where the base point of 0 was left off the final line. You are asked to show the final equality in the next question.

Question 4.8 Using the definition of the wedgeproduct of a k-form α and an ℓ-form β in terms of a $(k+\ell)$-shuffle,

$$
(\alpha \wedge \beta)\big(v_1, \ldots, v_{k+\ell}\big) = \sum_{\substack{\sigma \text{ is a} \\ (k+\ell)-\text{shuffle}}} \operatorname{sgn}(\sigma)\alpha\big(v_{\sigma(1)}, \ldots, v_{\sigma(k)}\big)\beta\big(v_{\sigma(k+1)}, \ldots, v_{\sigma(k+\ell)}\big),
$$

show the final equality in the proceeding calculation.

Thus we have shown that this definition of the exterior derivative of $\omega = f\, dx_{i_1} \wedge \cdots \wedge dx_{i_k}$ leads to the same computational formula we had in previous sections. It is then of course easy to show that given a general k-form $\alpha = \sum \alpha_{i_1 \cdots i_k} dx_{i_1} \wedge \cdots \wedge dx_{i_k}$ that

$$
d\alpha = \sum d\alpha_{i_1 \cdots i_k} \wedge dx_{i_1} \wedge \cdots \wedge dx_{i_k}.
$$

We now turn our attention to exploring the geometry of this definition. We will look at a two-form on the manifold $\mathbb{R}^3$ in order to allow us to draw pictures. We can write

$$
\begin{aligned}
d\omega(v_1, v_2, v_3) &= \lim_{h \to 0} \frac{1}{h^3} \int_{\partial(hP)} \omega \\
&= \lim_{h \to 0} \frac{1}{h^3} \left(\int_{hP_{(1,1)}} \omega - \int_{hP_{(1,0)}} \omega - \int_{hP_{(2,1)}} \omega + \int_{hP_{(2,0)}} \omega + \int_{hP_{(3,1)}} \omega - \int_{hP_{(3,0)}} \omega \right) \\
&= \lim_{h \to 0} \frac{1}{h^3} \underbrace{\left(\int_{hP_{(1,1)}} \omega - \int_{hP_{(1,0)}} \omega \right)}_{\text{see Fig. 4.4}} - \lim_{h \to 0} \frac{1}{h^3} \underbrace{\left(\int_{hP_{(2,1)}} \omega - \int_{hP_{(2,0)}} \omega \right)}_{\text{see Fig. 4.5}} \\
&\quad + \lim_{h \to 0} \frac{1}{h^3} \underbrace{\left(\int_{hP_{(3,1)}} \omega - \int_{hP_{(3,0)}} \omega \right)}_{\text{see Fig. 4.5}}.
\end{aligned}
$$

Figure 4.4 attempts to show what is going on with the first two terms above. The two faces $hP_{(1,0)}$ and $hP_{(1,1)}$ of the parallelepiped hP spanned by hv_1, hv_2, hv_3 are shown. The faces are defined as

$$
\begin{aligned}
&hP_{(1,0)} = \left\{ 0v_1 + t_2 v_2 + t_3 v_2 \;\middle|\; 0 \le t_i \le h, i = 2, 3 \right\} \\
\text{and} \qquad &hP_{(1,1)} = \left\{ hv_1 + t_2 v_2 + t_3 v_2 \;\middle|\; 0 \le t_i \le h, i = 2, 3 \right\}.
\end{aligned}
$$

The two-form ω, which is defined on $\mathbb{R}^3$, is integrated over each of these two-dimensional spaces. Using our intuition of what integration does, we can very roughly think of this as the "amount" of ω there is on each of these two spaces. When we take the difference of these integrals, $\int_{hP_{(1,1)}} \omega - \int_{hP_{(1,0)}} \omega$, we are trying to measure how much the "amount" of ω changes as we move in the v_1 direction. Of course, the "amount" of ω that we have we have obtained depends on the spaces that we are integrating over, which are faces of the parallelepiped hP, which is itself just a scaled down version of P. Thus, both the direction v_1 and the faces we are integrating over depend on the original parallelepiped P. We will say that the difference of these two integrals gives a measure of *how much the "amount" of ω is changing in the v_1 direction with respect to the parallelepiped P.*

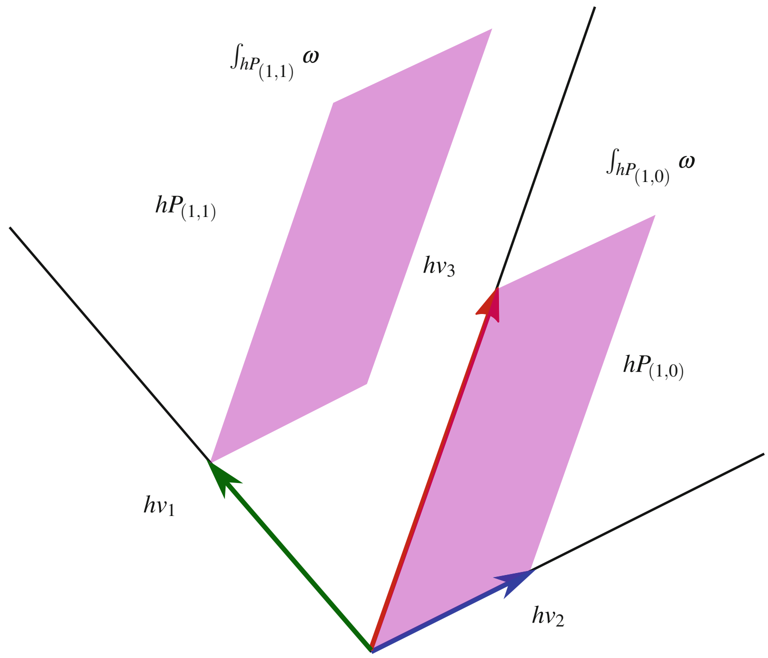


Fig. 4.4 Here the two faces $hP_{(1,0)}$ and $hP_{(1,1)}$ of the parallelepiped hP spanned by hv_1, hv_2, hv_3 are shown. The two-form ω is integrated over each of these faces. The difference of these two integrals, $\int_{hP_{(1,1)}} \omega - \int_{hP_{(1,0)}} \omega$, gives a measure of the "amount" ω is changing in the v_1 direction. Since clearly both the faces of hP we are integrating over and the direction v_1 in which the change is being measured depend on P, we say that we are measuring the change in "amount" of ω in the direction v_1 with respect to P

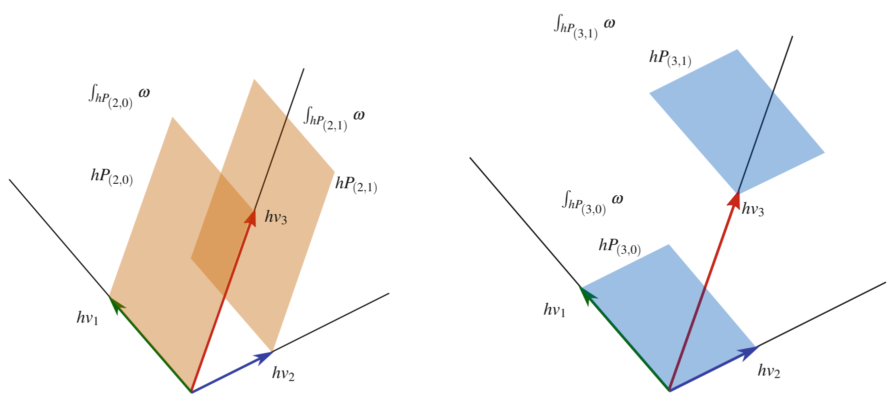


Fig. 4.5 On the left we are finding the difference between two integrals, $\int_{hP_{(2,1)}} \omega - \int_{hP_{(2,0)}} \omega$, which gives a measure on how much the "amount" of ω is changing in the v_2 direction with respect to the P. Similarly, the right we are finding the difference between two integrals, $\int_{hP_{(3,1)}} \omega - \int_{hP_{(3,0)}} \omega$, which gives a measure on how much the "amount" of ω is changing in the v_3 direction with respect to the P

In a similar manner the difference $\int_{hP_{(2,1)}} \omega - \int_{hP_{(2,0)}} \omega$ gives a measure of how much the "amount" of ω is changing in the v_2 direction with respect to the parallelepiped P and the difference $\int_{hP_{(3,1)}} \omega - \int_{hP_{(3,0)}} \omega$ gives a measure how much the "amount" of ω is changing in the v_3 direction with respect to the parallelepiped P. See Fig. 4.5.

By taking the limit as $h \to 0$ of $\frac{1}{h^3}\left(\int_{hP_{(1,1)}} \omega - \int_{hP_{(1,0)}} \omega\right)$ we are in effect finding the rate of change of the "amount" of ω in the direction v_1 with respect to P at the base point. (Recall, in the computation above we had assumed without loss of generality that the base point of the vectors $v_1, \ldots, v_k$ was the origin, but in general the base point could be any point on the manifold.) The limits as $h \to 0$ of $\frac{1}{h^3}\left(\int_{hP_{(2,1)}} \omega - \int_{hP_{(2,0)}} \omega\right)$ and $\frac{1}{h^3}\left(\int_{hP_{(3,1)}} \omega - \int_{hP_{(3,0)}} \omega\right)$ also find the rates of change on the "amount" of ω in the directions v_2 and v_3, respectively, with respect to the parallelepiped P which is spanned by the vectors v_1, v_2, v_3 at the base point. And then on top of that we add/subract these various rates of change. Whether we add or subtract is done in accordance the orientations of the various faces of P. This is explained in detail in Chap. 11.

The number we actually get for $d\omega(v_1, v_2, v_3)$ at a particular point is a somewhat strange combination of the various rates of change in the "amount" of ω in the directions v_1, v_2, and v_3, where those "amounts" are calculated using faces of the parallelepiped spanned by the vectors v_1, v_2, v_3. Thus the number $d\omega(v_1, v_2, v_3)$ is an overall rate of change of the "amount" of ω at the base point that depends on v_1, v_2, v_3. The final number may seem a little odd to you, but all the components that go into finding it have very clear and reasonable geometric meanings. Of course $d\omega$ itself is a three-form that allows us to compute this overall rate of change at a point by plugging in three vectors v_1, v_2, v_3 at that point.

Other than having a feeling for what we mean when we say "amount" of ω, using this definition for the exterior derivative gives us the most concrete geometric meaning for the exterior derivative of a k-form. We have not attempted to explain here what we mean when we say the "amount" of ω, other than simply saying it is the integral of ω over a face of P. The integrals of forms is explained in Chaps. 7–11. In particular, we look at the integration of forms from the perspective of Riemann sums in Sect. 7.2, which gives a very nice idea of what the integration of a form is achieving. However, in three dimensions the integration of forms have some very nice geometrical pictures that can be utilized to get an intuitive idea of what is going on. These pictures rely on how one-, two-, and three-forms can be visualized, which is covered in Chap. 5, and on Stokes' theorem, covered in Chap. 11. We will return to exploring the geometry of the exterior derivatives of one-, two-, and three-forms at the end of Sect. 11.6. Though there are no longer nice pictures to help us visualize differential forms for $k > 3$, the geometric meaning for the exterior derivative of a k-form on an n-dimensional manifold is analogous; there are just more directions and faces of P to be concerned with and all the faces of P are k-dimensional.

This particular definition for the exterior derivative of an n-form ω also presents the clearest intuitive idea of why the generalized Stokes' theorem is true. Using this definition,

$$d\omega\big(v_1, \ldots, v_{k+1}\big) = \lim_{h \to 0} \frac{1}{h^{k+1}} \int_{\partial(hP)} \omega$$

$$\Big\Updownarrow \quad \text{When } h \text{ is very small.}$$

$$d\omega\big(v_1, \ldots, v_{k+1}\big) \approx \frac{1}{h^{k+1}} \int_{\partial(hP)} \omega$$

$$\Big\Updownarrow \quad \text{Multiply both sides by } h^{k+1}.$$

$$h^{k+1} \underbrace{d\omega\big(v_1, \ldots, v_{k+1}\big)}_{\text{a number}} \approx \int_{\partial(hP)} \omega$$

$$\Big\Updownarrow \quad \text{Volume of } hP \text{ is approximately } h^{k+1}.$$

$$\underbrace{d\omega\big(v_1, \ldots, v_{k+1}\big)}_{\text{a number}} \int_{hP} \underbrace{dx_1 \wedge \cdots \wedge dx_{k+1}}_{\text{volume form over } hP} \approx \int_{\partial(hP)} \omega$$

$$\Big\Updownarrow \quad \text{Pull number under integral sign.}$$

$$\int_{hP} \underbrace{d\omega\big(v_1, \ldots, v_{k+1}\big)}_{\text{a number}} \underbrace{dx_1 \wedge \cdots \wedge dx_{k+1}}_{\text{volume form over } hP} \approx \int_{\partial(hP)} \omega$$

$$\Big\Updownarrow \quad \text{Write LHS in abstract notation.}$$

$$\int_{hP} d\omega \approx \int_{\partial(hP)} \omega$$

$$\Big\Updownarrow \quad \text{Letting } h \to 0.$$

$$\int_{hP} d\omega = \int_{\partial(hP)} \omega.$$

The final implication gives us the generalized Stokes' theorem for the region hP. This is of course not a rigorous proof, that is done in Chap. 11, but it does provide a very nice big-picture idea of what the generalized Stokes' theorem is and a reasonable geometric argument for why it is true.

Question 4.9 Analogous to above, explain the geometric meaning of $d\omega(v_1, v_2, v_3, v_4)$ for a three-form ω on an n-dimensional manifold.

Question 4.10 Analogous to above, explain the geometric meaning of $d\omega(v_1, \ldots, v_k)$ for a k-form ω on an n-dimensional manifold.

4.6 Exterior Differentiation Examples

We now turn our attention to utilizing the various formulas that we have obtained in the previous sections. Probably the simplest way to get used to doing computations involving exterior differentiation is simply to see and work through a series of examples. When doing these computations we will be going into quite a bit of detail just to get you used to and familiar with everything. Once you are familiar with computations of this nature you will be able to skip a number of the steps without any problem.

Suppose we have the two functions $f, g : \mathbb{R}^2 \longrightarrow \mathbb{R}$. We want to find $df \wedge dg$. First we note that we have

$$df = \frac{\partial f}{\partial x}dx + \frac{\partial f}{\partial y}dy \quad \text{and} \quad dg = \frac{\partial g}{\partial x}dx + \frac{\partial g}{\partial y}dy.$$

We then take the wedgeproduct of these two one-forms and use the algebraic properties of the wedgeproduct presented in Sect. 3.3.1,

$$\begin{aligned}
df \wedge dg &= \left(\frac{\partial f}{\partial x}dx + \frac{\partial f}{\partial y}dy\right) \wedge \left(\frac{\partial g}{\partial x}dx + \frac{\partial g}{\partial y}dy\right) \\
&= \frac{\partial f}{\partial x}dx \wedge \left(\frac{\partial g}{\partial x}dx + \frac{\partial g}{\partial y}dy\right) + \frac{\partial f}{\partial y}dy \wedge \left(\frac{\partial g}{\partial x}dx + \frac{\partial g}{\partial y}dy\right) \\
&= \frac{\partial f}{\partial x}\frac{\partial g}{\partial x}\underbrace{dx \wedge dx}_{=0} + \frac{\partial f}{\partial x}\frac{\partial g}{\partial y}dx \wedge dy + \frac{\partial f}{\partial y}\frac{\partial g}{\partial x}dy \wedge dx + \frac{\partial f}{\partial y}\frac{\partial g}{\partial y}\underbrace{dy \wedge dy}_{=0} \\
&= \frac{\partial f}{\partial x}\frac{\partial g}{\partial y}dx \wedge dy + \frac{\partial f}{\partial y}\frac{\partial g}{\partial x}dy \wedge dx \\
&= \frac{\partial f}{\partial x}\frac{\partial g}{\partial y}dx \wedge dy - \frac{\partial f}{\partial y}\frac{\partial g}{\partial x}dx \wedge dy \\
&= \left(\frac{\partial f}{\partial x}\frac{\partial g}{\partial y} - \frac{\partial f}{\partial y}\frac{\partial g}{\partial x}\right)dx \wedge dy \\
&= \begin{vmatrix} \frac{\partial f}{\partial x} & \frac{\partial f}{\partial y} \\ \frac{\partial g}{\partial x} & \frac{\partial g}{\partial y} \end{vmatrix} dx \wedge dy.
\end{aligned}$$

So, the wedgeproduct of df and dg involves the Jacobian matrix from vector calculus in front of the two-dimensional volume form (area form) $dx \wedge dy$. In other words, we have found the identity

$$\boxed{df \wedge dg = \underbrace{\begin{vmatrix} \frac{\partial f}{\partial x} & \frac{\partial f}{\partial y} \\ \frac{\partial g}{\partial x} & \frac{\partial g}{\partial y} \end{vmatrix}}_{Jacobian} \underbrace{dx \wedge dy.}_{area form}}$$

It is perhaps a little tedious, but not difficult to show that for $f, g, h : \mathbb{R}^3 \to \mathbb{R}$ we have

$$\boxed{df \wedge dg \wedge dh = \underbrace{\begin{vmatrix} \frac{\partial f}{\partial x} & \frac{\partial f}{\partial y} & \frac{\partial f}{\partial z} \\ \frac{\partial g}{\partial x} & \frac{\partial g}{\partial y} & \frac{\partial g}{\partial z} \\ \frac{\partial h}{\partial x} & \frac{\partial h}{\partial y} & \frac{\partial h}{\partial z} \end{vmatrix}}_{Jacobian} \underbrace{dx \wedge dy \wedge dz}_{volumeform}.}$$

The obvious generalization holds as well. However, the proof of the general formula using the properties of the wedgeproduct is rather tedious. Instead, we will defer the proof of this formula until later when more mathematical machinery has been introduced and we are able to give a much slicker proof.

Question 4.11 Given functions $f, g, h : \mathbb{R}^3 \to \mathbb{R}$ find the above identity for $df \wedge dg \wedge dh$.

Now we look at an example where we have two explicitly given functions. For example, suppose we had a change of coordinates given by $f : \mathbb{R}^2 \to \mathbb{R}$ by $f(x, y) = x + y$ and $g : \mathbb{R}^2 \to \mathbb{R}$ by $g(x, y) = x - y$. To find $df \wedge dg$ we could take one of two approaches. For example, we could first find df and dg and then use the algebraic properties of differential forms to simplify it or we could simply skip to the identity we found above. We will show both approaches. Following the first approach and first finding df and dg we have

$$\begin{aligned} df &= \frac{\partial f}{\partial x}dx + \frac{\partial f}{\partial y}dy = \frac{\partial(x+y)}{\partial x}dx + \frac{\partial(x+y)}{\partial y}dy = dx + dy \\ dg &= \frac{\partial g}{\partial x}dx + \frac{\partial g}{\partial y}dy = \frac{\partial(x-y)}{\partial x}dx + \frac{\partial(x-y)}{\partial y}dy = dx - dy. \end{aligned}$$

Then using our algebraic properties of differential forms to simplify we find

$$\begin{aligned} df \wedge dg &= (dx + dy) \wedge (dx - dy) \\ &= dx \wedge (dx - dy) + dy \wedge (dx - dy) \\ &= \underbrace{dx \wedge dx}_{=0} - dx \wedge dy + dy \wedge dx - \underbrace{dy \wedge dy}_{=0} \\ &= -dx \wedge dy - dx \wedge dy \\ &= -2dx \wedge dy. \end{aligned}$$

In the second approach we just use the identity we found earlier to get

$$\begin{aligned} df \wedge dy &= \begin{vmatrix} \frac{\partial f}{\partial x} & \frac{\partial f}{\partial y} \\ \frac{\partial g}{\partial x} & \frac{\partial g}{\partial y} \end{vmatrix} dx \wedge dy \\ &= \begin{vmatrix} \frac{\partial(x+y)}{\partial x} & \frac{\partial(x+y)}{\partial y} \\ \frac{\partial(x-y)}{\partial x} & \frac{\partial(x-y)}{\partial y} \end{vmatrix} dx \wedge dy \\ &= \begin{vmatrix} 1 & 1 \\ 1 & -1 \end{vmatrix} dx \wedge dy \\ &= -2dx \wedge dy. \end{aligned}$$

It is not at all surprising that our two answers agree.

Let's go into more depth with another example involving more complicated functions. Consider $f : \mathbb{R}^2 \to \mathbb{R}$ given by $f(x, y) = x^2y + x$ and $g : \mathbb{R}^2 \to \mathbb{R}$ given by $g(x, y) = x^2y^3 + 2xy$. We use our identity,

$$df \wedge dg = \begin{vmatrix} (2xy + 1) & x^2 \\ (2xy^3 + 2y) & (3x^2y^2 + 2x) \end{vmatrix} dx \wedge dy = \Big[(2xy + 1)(3x^2y^2 + 2x) - x^2(2xy^3 + 2y)\Big] dx \wedge dy.$$

This is $df \wedge dg$ on all of our manifold $\mathbb{R}$. Suppose we want to know what $df \wedge dg$ were at a specific point of the manifold? We would have to substitute in that point's values. For example, suppose we wanted to find $df \wedge dg$ at the point $p = (1, 1)$ of manifold $\mathbb{R}^2$. We would substitute to get

$$\begin{aligned}\Big(df \wedge dg\Big)_{p=(1,1)} &= \Big[\big(2\cdot 1\cdot 1+1\big)\big(3\cdot 1^2\cdot 1^2+2\cdot 1\big) - 1^2\big(2\cdot 1\cdot 1^3+2\cdot 1\big)\Big]dx\wedge dy\\ &= \big[(3)(5)-(4)\big]dx\wedge dy\\ &= 11dx\wedge dy.\end{aligned}$$

Similarly, if we wanted to find $df \wedge dg$ at the point $p = (-1, 2)$ of manifold $\mathbb{R}^2$ we would substitute to get

$$\begin{aligned}\Big(df \wedge dg\Big)_{p=(-1,2)} &= \Big[\Big(2\cdot(-1)\cdot 2+1\Big)\Big(3\cdot(-1)^2\cdot 2^2+2\cdot(-1)\Big)\\ &\qquad -(-1)^2\Big(2\cdot(-1)\cdot 2^3+2\cdot(-1)\Big)\Big]dx\wedge dy\\ &= \big[(-3)(10)-(-18)\big]dx\wedge dy\\ &= -12dx\wedge dy.\end{aligned}$$

Now suppose we had the two vectors $v_p = \begin{bmatrix} -2 \\ 3 \end{bmatrix}_{(1,1)}$ and $w_p = \begin{bmatrix} -4 \\ -2 \end{bmatrix}_{(1,1)}$ and wanted to find $(df \wedge dg)(v_p, w_p)$. Omitting writing the base point $(1, 1)$ after the first equality we get

$$\begin{aligned}\big(df\wedge dg\big)_{(1,1)}\left(\begin{bmatrix} -2 \\ 3 \end{bmatrix}_{(1,1)}, \begin{bmatrix} -4 \\ -2 \end{bmatrix}_{(1,1)}\right) &= 11dx\wedge dy\left(\begin{bmatrix} -2 \\ 3 \end{bmatrix}, \begin{bmatrix} -4 \\ -2 \end{bmatrix}\right)\\ &= 11\cdot\begin{vmatrix} dx\left(\begin{bmatrix} -2 \\ 3 \end{bmatrix}\right) & dx\left(\begin{bmatrix} -4 \\ -2 \end{bmatrix}\right) \\ dy\left(\begin{bmatrix} -2 \\ 3 \end{bmatrix}\right) & dy\left(\begin{bmatrix} -4 \\ -2 \end{bmatrix}\right)\end{vmatrix}\\ &= 11\cdot\begin{vmatrix} -2 & 4 \\ 3 & -2 \end{vmatrix} = 11(4-12)\\ &= 88.\end{aligned}$$

And if we had the vectors $v_p = \begin{bmatrix} 3 \\ 1 \end{bmatrix}_{(-1,2)}$ and $w_p = \begin{bmatrix} 1 \\ 2 \end{bmatrix}_{(-1,2)}$ and wanted to find $(df \wedge dg)(v_p, w_p)$ we would have

$$\big(df\wedge dg\big)_{(-1,2)}\left(\begin{bmatrix} 3 \\ 1 \end{bmatrix}_{(-1,2)}, \begin{bmatrix} 1 \\ 2 \end{bmatrix}_{(-1,2)}\right) = -12\begin{vmatrix} 3 & 1 \\ 1 & 2 \end{vmatrix} = -60.$$

Now we do one more example that may look vaguely familiar to you. Write $dx \wedge dy$ in terms of $d\theta \wedge dr$ if $x = r\cos(\theta)$ and $y = r\sin(\theta)$. Doing this gives us

$$\begin{aligned}dx\wedge dy &= \begin{vmatrix} \frac{dx}{d\theta} & \frac{dx}{dr} \\ \frac{dy}{d\theta} & \frac{dy}{dr}\end{vmatrix} d\theta\wedge dr\\ &= \begin{vmatrix} \frac{d\big(r\cos(\theta)\big)}{d\theta} & \frac{d\big(r\cos(\theta)\big)}{dr} \\ \frac{d\big(r\sin(\theta)\big)}{d\theta} & \frac{d\big(r\sin(\theta)\big)}{dr}\end{vmatrix} d\theta\wedge dr\end{aligned}$$

$$
\begin{aligned}
&= \begin{vmatrix} -r\sin(\theta) & \cos(\theta) \\ r\cos(\theta) & \sin(\theta) \end{vmatrix} d\theta \wedge dr \\
&= \big(-r\sin^2(\theta) - r\cos^2(\theta)\big) d\theta \wedge dr \\
&= -r d\theta \wedge dr.
\end{aligned}
$$

If you think this looks a little like a polar change of coordinates formula then you would be absolutely correct. We will discuss this in much greater detail soon.

Question 4.12 Suppose $f, g, h : \mathbb{R}^3 \to \mathbb{R}$ are defined by $f(x, y, z) = x^2 y$, $g(x, y, z) = y^2 z^2$, and $h(x, y, z) = xyz$.

(a) Find $df \wedge dg \wedge dh$.
(b) Find $df \wedge dg \wedge dh$ at the point $p = (-1, 1, -1)$.
(c) Find $df \wedge dg \wedge dh_{(-1,1,-1)}\big(u_p, v_p, w_p\big)$ where

$$
u_p = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}_{(-1,1,-1)}, \quad v_p = \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix}_{(-1,1,-1)}, \quad w_p = \begin{bmatrix} -2 \\ -1 \\ 0 \end{bmatrix}_{(-1,1,-1)}.
$$

Over the last few examples and questions we have dealt with a very specific sort of problem. Notice that the number of functions we had (two or three) was the same as the dimension of the manifold the functions were defined on. For example, on the manifold $\mathbb{R}^2$ we had two functions, f and g. On the manifold $\mathbb{R}^3$ we had three functions, f, g, h. Thus, in the first case $df \wedge dg$ gave us a function multiplied by the two dimensional volume form $dx \wedge dy$ and in the second case $df \wedge dg \wedge dh$ gave us a functions multiplied by the three dimensional volume form. Now we will look at a few more general examples.

Question 4.13 Suppose $f, g : \mathbb{R}^3 \to \mathbb{R}$ are given by $f(x, y, z) = x^2 y^3 z^2$ and $g(x, y, z) = xyz^2$.

(a) Find df and dg.
(b) Find $df \wedge dg$. (Notice, you can no longer use the Jacobian matrix, you must use df and dg that you found above and the algebraic properties of the wedgeproduct.)
(c) Find $df \wedge dg$ at the point $(3, 2, 1)$.
(d) Suppose $v_p = \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix}$ and $w_p = \begin{bmatrix} -2 \\ 1 \\ -1 \end{bmatrix}$. Find $df \wedge dg(v_p, w_p)$.

Now we turn to some examples where we find the exterior derivative of a one-form. Suppose that $\phi = x^2 yz dx$ is a one-form on $\mathbb{R}^3$. We will find $d\phi$ using our formula from the last section.

$$
\begin{aligned}
d\phi &= d(x^2 yz) \wedge dx \\
&= \Big(\frac{\partial(x^2yz)}{\partial x} dx + \frac{\partial(x^2yz)}{\partial y} dy + \frac{\partial(x^2yz)}{\partial z} dz\Big) \wedge dx \\
&= \big(2xyz dx + x^2 z dy + x^2 y dz\big) \wedge dx \\
&= 2xyz \underbrace{dx \wedge dx}_{=0} + x^2 z dy \wedge dx + x^2 y dz \wedge dx \\
&= -x^2 z dx \wedge dy - x^2 y dx \wedge dz.
\end{aligned}
$$

For our next example suppose $\phi = \sin(x)dx + e^x \cos(y)dy + 3xydz$. We will find $d\phi$ in three steps, one for each term of ϕ,

$$
d\phi = \underbrace{d\big(\sin(x)\big) \wedge dx}_{(1)} + \underbrace{d\big(e^x \cos(y)\big) \wedge dy}_{(2)} + \underbrace{d\big(3xy\big) \wedge dz}_{(3)}.
$$

The first term gives us

$$
\begin{aligned}
d\big(\sin(x)\big) \wedge dx &= \Big(\frac{\partial \sin(x)}{\partial x}dx + \frac{\partial \sin(x)}{\partial y}dy + \frac{\partial \sin(x)}{\partial z}dz\Big) \wedge dx \\
&= \big(\cos(x)dx + 0dy + 0dz\big) \wedge dx \\
&= \cos(x)\underbrace{dx \wedge dx}_{=0} \\
&= 0.
\end{aligned}
$$

The second term gives us

$$
\begin{aligned}
d\big(e^x\cos(y)\big) \wedge dy &= \Big(\frac{\partial(e^x\cos(y))}{\partial x}dx + \frac{\partial(e^x\cos(y))}{\partial y}dy + \frac{\partial(e^x\cos(y))}{\partial z}dz\Big) \wedge dy \\
&= e^x\cos(y)dx \wedge dy - e^x\sin(y)\underbrace{dy \wedge dy}_{=0} + 0dz \wedge dy \\
&= e^x\cos(y)dx \wedge dy.
\end{aligned}
$$

The third term gives us

$$
\begin{aligned}
d\big(3xy\big) \wedge dz &= \Big(\frac{\partial 3xy}{\partial x}dx + \frac{\partial 3xy}{\partial y}dy + \frac{\partial 3xy}{\partial z}dz\Big) \wedge dz \\
&= 3ydx \wedge dz + 3xdy \wedge dz + 0dz \wedge dz \\
&= 3xdy \wedge dz - 3ydz \wedge dx.
\end{aligned}
$$

Combining everything we have

$$
d\phi = e^x\cos(y)dx \wedge dy + 3xdy \wedge dz - 3ydz \wedge dx.
$$

Question 4.14 Find a formula for $d\phi$ if $\phi = f_1dx_1 + f_2dx_2 + f_3dx_3$.

Your answer for the last question should have been

$$
d\phi = \Big(\frac{\partial f_2}{\partial x_1} - \frac{\partial f_1}{\partial x_2}\Big)dx_1 \wedge dx_2 + \Big(\frac{\partial f_3}{\partial x_1} - \frac{\partial f_1}{\partial x_3}\Big)dx_1 \wedge dx_3 + \Big(\frac{\partial f_3}{\partial x_2} - \frac{\partial f_2}{\partial x_3}\Big)dx_2 \wedge dx_3.
$$

Next we show that if f, g are functions on $\mathbb{R}^n$ that $d(f \cdot g) = df \cdot g + f \cdot dg$.

$$
\begin{aligned}
d(f \cdot g) &= \sum_i \frac{\partial f \cdot g}{\partial x_i}dx_i \\
&\underset{\text{prod. rule}}{=} \sum_i \Big(\frac{\partial f}{\partial x_i} \cdot g + f \cdot \frac{\partial g}{\partial x_i}\Big)dx_i \\
&= f \cdot \sum_i \frac{\partial f}{\partial x_i}dx_i + f\sum_i \frac{\partial g}{\partial x_i}dx_i \\
&= g \cdot df + f \cdot dg.
\end{aligned}
$$

This is really nothing more than the product rule in somewhat different notation.

Question 4.15 Show that the exterior derivative d has the linearity property, that is, if $\phi = \sum f_i dx_i$ and $\psi = \sum g_i dx_i$ then for $a, b \in \mathbb{R}$,

$$
d(a\phi + b\psi) = ad\phi + bd\psi.
$$

Question 4.16 If f is a function and ϕ a one-form, show

$$d(f\phi) = df \wedge \phi + f d\phi.$$

Question 4.17 If ϕ, ψ are one-forms, show

$$d(\phi \wedge \psi) = d\phi \wedge \psi + \phi \wedge d\psi.$$

Note, in order of operations multiplication and wedgeproducts take precedence over addition and subtraction.

Next we show that if f is a function on $\mathbb{R}^n$ that $ddf = 0$. This means that no matter what f is that taking the exterior derivative twice gives 0. First, we know $df = \sum_i \frac{\partial f}{\partial x_i} dx_i$, so

$$\begin{aligned}
ddf &= d\Big(\sum_i \frac{\partial f}{\partial x_i} dx_i\Big) \\
&= \sum_i d\Big(\frac{\partial f}{\partial x_i}\Big) \wedge dx_i \\
&= \sum_i \sum_j \frac{\partial}{\partial x_j}\Big(\frac{\partial f}{\partial x_i}\Big) dx_j \wedge dx_i \\
&= \sum_{i,j} \frac{\partial^2 f}{\partial x_j \partial x_i} \qquad \underbrace{dx_j \wedge dx_i}_{\substack{\text{If } i<j \text{ then} \\ dx_j \wedge dx_i = -dx_i \wedge dx_j}} \\
&= \sum_{i<j} \underbrace{\Big(\frac{\partial^2 f}{\partial x_j \partial x_i} - \frac{\partial^2 f}{\partial x_i \partial x_j}\Big)}_{\substack{=0 \\ \text{by equality of mixed partials}}} dx_i \wedge dx_j \\
&= 0.
\end{aligned}$$

Question 4.18 For the functions f, g, show $d(f dg) = df \wedge dg$.

Question 4.19 Simplify

(a) $d(f dg + f df)$,
(b) $d\big((f - g)(df + dg)\big)$,
(c) $d\big((f dg) \wedge (g df)\big)$,
(d) $d(g f df) + d(f dg)$.

As we mentioned earlier, one of our simplifications so far has been to stick with the standard Cartesian coordinate system. Given how familiar most students are with the Cartesian coordinate system it makes sense to introduce a complex idea like differential forms, wedgeproducts, and exterior differentiation in an arena that most students are completely comfortable with. We don't want you to lose sight of the big ideas because you were getting mired down in the details of other coordinate systems. However, other coordinate systems will soon play an increasingly important role for us.

4.7 Summary, References, and Problems

4.7.1 Summary

In this chapter four different approaches to exterior differentiation were explored. In the first approach the local formula, that is a formula that uses coordinates, for the exterior derivative was given,

Exterior derivative of an n-form	$d\Big(\sum \alpha_{i_1\cdots i_n}\, dx_{i_1}\wedge\cdots\wedge dx_{i_n}\Big) = \sum\sum_{j=1}^{n} \frac{\partial\, \alpha_{i_1\cdots i_n}}{\partial x_{i_j}}\, dx_{i_j}\wedge dx_{i_1}\wedge\cdots\wedge dx_{i_n}.$

Using this local formula one can then show the following algebraic properties hold,

(1) $d(\alpha+\beta) = d\alpha + d\beta$,
(2) $d(\alpha\wedge\omega) = d\alpha\wedge\omega + (-1)^n\alpha\wedge d\omega$,
(3) for each n-form α, $d(d\alpha) = 0$.

The second approach to defining the exterior derivative basically goes backwards from this. The three properties are given as defining axioms and then a unique formula in coordinates is derived from these properties.

The third approach to exterior differentiation is an attempt to find a global formula. A global formula differs from a local formula in that it does not rely on coordinates. That is, the coordinates do not show up in the formula. First a global formula for the exterior derivative of a k-form is found using constant vector fields,

Global formula for exterior derivative of a k-form, constant vector fields	$d\omega(v_0,\ldots,v_k) = \sum_i (-1)^i v_i\big[\omega(v_0,\ldots,\widehat{v_i},\ldots,v_k)\big].$

Using the definition of the lie bracket of two vector fields,

$$[v,w]\big[F\big] = v\big[w[F]\big] - w\big[v[F]\big],$$

a global formula for the exterior derivative of a two-form is found for vector fields that are not constant,

Global formula for exterior derivative of a one-form,	$d\alpha(v,w) = v[\alpha(w)] - w[\alpha(v)] - \alpha([v,w]).$

The general global formula for the exterior derivative of a k-form with non-constant vector fields is then given,

Global formula for exterior derivative of k-form	$d\alpha(v_0,\ldots,v_k) = \sum_i (-1)^i v_i\big[\alpha(v_0,\ldots,\widehat{v_i},\ldots,v_k)\big] + \sum_{i<j}(-1)^{i+j}\alpha\big([v_i,v_j],v_0,\ldots,\widehat{v_i},\ldots,\widehat{v_j},\ldots,v_k\big).$

This formula is not actually proved until Sect. A.7.

The fourth approach to exterior differentiation is a very geometrical approach where the exterior derivative of k-form at a point is defined by

Exterior derivative of a k-form	$d\omega\big(v_1,\ldots,v_{k+1}\big) = \lim_{h\to 0}\frac{1}{h^{k+1}}\int_{\partial(hP)}\omega.$

This geometrical definition is revisited in Sect. 11.6 in order to obtain a deeper understanding of the geometry in the three-dimensional case. This geometrical definition of the exterior derivative also gives the most geometrical argument (but not proof) for why Stokes' theorem is true.

4.7.2 References and Further Reading

As the overview explains, in most books exterior differentiation is typically approached in one of four ways. As with the wedgeproduct, this is understandable since exterior differentiation is simply one of many different topics needed to be covered. However, we have tried, as much as is possible, to bring everything together and explore all of these approaches to exterior differentiation so the reader will be completely comfortable with any approach to exterior differentiation that they may encounter in the future. For example, O'Neill [36] uses a local coordinates approach while Tu [46] covers exterior differentiation from a local coordinates perspective early on in the book and then from a global perspective toward the middle of the book. Martin [33], Darling [12], and Flanders [19] all use an axiom-based approach. The geometric viewpoint of the exterior derivative is quite rare, the only place we are aware of where it is actually used as an approach to introducing the exterior derivative is in Hubbard and Hubbard [27] though this geometric meaning of the exterior derivative is also alluded to briefly in Darling [12] and Arnold [3]. Our exposition largely follows that of Hubbard and Hubbard, though it is not as complete or detailed as theirs is.

4.7.3 Problems

Question 4.20 Let $f_1(x,y,z) = x^3y^2z - 2xyz^3$, $f_2(x,y,z) = x^2 - 3y^2 + 6z^4$, $f_3(x,y,z) = xy^2 + yz^2 + zx^2$, and $f_4(x,y,z) = 3x - 2y + 5z - 4$. Find df_1, df_2, df_3, and df_4.

Question 4.21 Let $g_1(x,y,z) = \frac{x^2+y^2}{z}$, $g_2(x,y,z) = \sin(x^2+y^2) + \tan(2z)$, $g_3(x,y,z) = xz + ye^y + 3x - 2$, and $g(x,y,z) = \cos(xy^2z) + \frac{x^2}{yz}$. Find dg_1, dg_2, and dg_3.

Question 4.22 For the functions in Question 4.20 and 4.21 find $d(2f_1+3g_1)$, $d(5f_2+3g_2)$, $d(-2f_3-5g_3)$, and $d(7f_4-3g_3)$.

Question 4.23 For the functions in Question 4.20 find $d(f_1 \cdot f_2)$, $d(f_1 \cdot f_3)$, $d(f_1 \cdot f_4)$, $d(f_2 \cdot f_3)$, $d(f_2 \cdot f_4)$, and $d(f_3 \cdot f_4)$.

Question 4.24 Let $\alpha_1 = 3x^3\,dx + \frac{x+y}{z}\,dy$, $\alpha_2 = -x\,dx + (x-3y)\,dy$, $\alpha_3 = \frac{x^3}{y}\,dx + xyz\,dx + (x^2+y^2+z^2)\,dz$, and $\alpha_4 = y^2z\,dx - xz\,dy + (3x+2)\,dz$. Find $d\alpha_1$, $d\alpha_2$, $d\alpha_3$, and $d\alpha_4$.

Question 4.25 Let $\beta_1 = \sin(x)\,dx + \sin(y)\,dy - \cos(z)\,dz$, $\beta_2 = \sin(xyz)\,dx - \cos(x^2)\,dy + e^y\,dz$, $\beta_3 = \sin^2(xz)\,dx + \cos^3(yz)\,dy - e^{xy}\,dz$, and $\beta_4 = (x^2+y^2+z^2)\,dx - x^z\,dy - xze^y\,dz$. Find $d\beta_1$, $d\beta_2$, $d\beta_3$, and $d\beta_4$.

Question 4.26 For the one-forms in Questions 4.24 and 4.25 find $d(2\alpha_1+3\beta_1)$, $d(5\alpha_2+3\beta_2)$, $d(-2\alpha_3-5\beta_3)$, and $d(7\alpha_4-3\beta_3)$.

Question 4.27 For the functions in Question 4.20 and the one-forms in Question 4.24 find $d(f_1\alpha_1)$, $d(f_2\alpha_2)$, $d(f_3\alpha_3)$, and $d(f_4\alpha_4)$.

Question 4.28 For the one-forms in Questions 4.24 and 4.25 find $d(\alpha_1 \wedge \beta_1)$, $d(\alpha_2 \wedge \beta_2)$, $d(\alpha_3 \wedge \beta_3)$, and $d(\alpha_4 \wedge \beta_4)$.

Question 4.29 For the functions in Question 4.20 find d^2f_1, d^2f_2, d^2f_3, and d^2f_4.

Question 4.30 For the one-forms in Questions 4.24 find $d^2\alpha_1$, $d^2\alpha_2$, $d^2\alpha_3$, and $d^2\alpha_4$.

Question 4.31 Let $\gamma_1 = (x^2+y^4)\,dx \wedge dy + s^3y^2z^5\,dy \wedge dz - (3x+2y-4z+7)\,dz \wedge dx$ and $\gamma_2 = xy\sin(z)\,dx \wedge dy - \cos(xyz)\,dy \wedge dz + \sin(xz)\cos(yz)dy \wedge dx$. Find $d\gamma_1$, $d\gamma_2$, and $d(3\gamma_1 - 5\gamma_2)$. Then find $d^2\gamma_1$ and $d^2\gamma_2$.

Question 4.32 If $f : \mathbb{R}^2 \to \mathbb{R}$ and $g : \mathbb{R} \to \mathbb{R}$, show that $v_p\big[g(f)\big] = g'(f)\ v_p\big[f\big]$. Use this to deduce that $d\big(g(f)\big) = g'(f)\,df$.

Question 4.33 If f, g, and h are real-valued functions on $\mathbb{R}^2$ and α is a one-form on $\mathbb{R}^2$, show $d(fgh) = gh\,df + fh\,dg + fg\,dh$ and $\big(df \wedge dg\big)(v,w) = v[f]w[g] - v[g]w[f]$.

Question 4.34 Given the two-forms $\alpha = x_1x_4x_6\,dx_1 \wedge dx_5$, $\beta = x_2^3\sin(x_4)\,dx_3 \wedge dx_5$, and $\gamma = (x_1^2+x_3^2+x_5^2)\,dx_2 \wedge dx_6$ on $\mathbb{R}^6$ find $d\alpha$, $d\beta$, $d\gamma$, $d^2\alpha$, $d^2\beta$, and $d^2\gamma$. Now assume α, β, and γ are two-forms on $\mathbb{R}^8$ and find the same exterior derivatives. Are they the same or different? Explain why or why not.

Use the definition for the integral of a zero-form at a single point given in Sect. 4.5 to answer the following questions.

Question 4.35 Let $P = \big\{(3,2,1)\big\} \subset \mathbb{R}^3$ and let $f = x^3y^2 - z^3$ be a zero-form on $\mathbb{R}^3$. Evaluate $\int_P f$.

Question 4.36 Let $P = -\big\{(-1,2,-1)\big\} \subset \mathbb{R}^3$ and $f = x(y-z+2)$. Evaluate $\int_P f$.

Question 4.37 Let $P = \big\{(5,-5,3)\big\} \subset \mathbb{R}^3$ and $f = x^3(xy^2+yz^2+3)$. Evaluate $\int_P f$.

Question 4.38 Let $P = -\big\{(-4,2,7)\big\} \subset \mathbb{R}^3$ and $f = 2^xy^z$. Evaluate $\int_P f$.

Chapter 5
Visualizing One-, Two-, and Three-Forms

In this chapter we will introduce and discuss at length one of the ways that physicists sometimes visualize "nice" differential forms. In essence, we will be considering ways of visualizing one-forms, two-forms, and three-forms on a vector space. That is, we will find a "cartoon picture" of $\alpha_p \in T_p^*\mathbb{R}^2$ and $\alpha_p \in T_p^*\mathbb{R}^3$. Our picture of α_p will be superimposed on the vector space $T_p\mathbb{R}^3$. This perspective is developed extensively in Misner, Thorne, and Wheeler's giant book *Gravitation*. In fact, they make some efforts to develop the four-dimensional picture (for space-time) as well, however we will primarily stick to the two and three-dimensional cases here. Section one focuses on the two-dimensional case and sections two through four focuses on the three-dimensional case. Then in section five we expand our cartoon picture to general two and three-dimensional manifolds. Again, this cartoon picture really only applies to "nice" differential forms, of the kind physicists are more likely to encounter, but it is still useful for forms and manifolds that are not overly complicated. Finally, in section six we introduce the Hodge star operator. Our visualizations of forms in three dimensions provide a nice way to visualize what the Hodge star operator does, which makes this a nice place to introduce it. Despite the power and usefulness of the way of visualizing differential forms in physics developed in this chapter, it is rarely encountered in mathematics. One of the reasons for this is that in reality it is not a completely general way of considering forms; when dealing with dimensions greater than four or with more abstract manifolds or with forms that are not "nice" in some sense it breaks down.

5.1 One- and Two-Forms in $\mathbb{R}^2$

We will start out by considering the one-forms $\alpha_p \in T_p^*\mathbb{R}^2$. Recall that a one-form α_p "eats" vectors v_p at the point p on the manifold $\mathbb{R}^2$. We will take the particular one-form dx_p at some point p. For the moment we will suppress the p in the notation and assume everything happens at one particular point p. Recall, the one-form dx is the exterior derivative of the zero-form x, which is itself nothing more than the coordinate function x. But in the case of zero-forms exterior derivatives are simply differentials of functions. Recall, the differential of a function f is defined by $df(v) = v[f]$, the directional derivative of f in the direction v.

But what is the rate of change of the coordinate function x in the x direction? It is one of course. In other words, the slope of the graph of coordinate function x in the x direction is one. We think of the slope as being the "rise" over "run", the vertical change divided by the horizontal change. But when we defined directional derivatives we made a slight alteration in our definition to accommodate vectors that were not unit vectors. Instead of directional derivatives giving us slopes they give us the "rise" portion of the slope. So, for the coordinate function x the differential dx measures the "rise" that occurs in the x direction. For example, consider a vector

$$v = \begin{bmatrix} 3 \\ 2 \end{bmatrix}.$$

As one moves along the vector v from the point $(0, 0)$ to the point $(3, 2)$ the graph of the coordinate function x "rises" three units, hence

$$dx\left(\begin{bmatrix} 3 \\ 2 \end{bmatrix}\right) = 3.$$

J. P. Fortney, *A Visual Introduction to Differential Forms and Calculus on Manifolds*,
https://doi.org/10.1007/978-3-319-96992-3_5

Question 5.1 Sketch the coordinate function x on $\mathbb{R}^2$ and find the "rise" of the coordinate function x along the vectors

(a) $\begin{bmatrix} 1 \\ 0 \end{bmatrix}$, that is, from the point $(0, 0)$ to the point $(1, 0)$,

(b) $\begin{bmatrix} -1 \\ 0 \end{bmatrix}$, that is, from the point $(0, 0)$ to the point $(-1, 0)$,

(c) $\begin{bmatrix} 2 \\ 2 \end{bmatrix}$, that is, from the point $(0, 0)$ to the point $(2, 2)$,

(d) $\begin{bmatrix} -3 \\ -1 \end{bmatrix}$, that is, from the point $(0, 0)$ to the point $(-3, -1)$.

Question 5.2 Sketch the coordinate function y on $\mathbb{R}^2$ and find the "rise" of the coordinate function y along the vectors

(a) $\begin{bmatrix} 0 \\ 1 \end{bmatrix}$, that is, from the point $(0, 0)$ to the point $(0, 1)$,

(b) $\begin{bmatrix} 1 \\ 1 \end{bmatrix}$, that is, from the point $(0, 0)$ to the point $(1, 1)$,

(c) $\begin{bmatrix} -1 \\ 2 \end{bmatrix}$, that is, from the point $(0, 0)$ to the point $(-1, 2)$,

(d) $\begin{bmatrix} 2 \\ -3 \end{bmatrix}$, that is, from the point $(0, 0)$ to the point $(2, -3)$.

This explains the way we originally introduced the one-forms. The way we introduced the one-form dx was as a projection of $v \in T_p\mathbb{R}^2$ onto the ∂_x-axis of $T_p\mathbb{R}^2$. This is exactly the amount of "rise" of the coordinate function x from the point p to the endpoint of v_p. Thus we had

$$
\begin{array}{llll}
dx\left(\begin{bmatrix} 1 \\ 4 \end{bmatrix}\right) = 1, & dx\left(\begin{bmatrix} -1 \\ 4 \end{bmatrix}\right) = -1, & dx\left(\begin{bmatrix} 2 \\ 5 \end{bmatrix}\right) = 2, & dx\left(\begin{bmatrix} 3 \\ 6 \end{bmatrix}\right) = 3, \\
dx\left(\begin{bmatrix} 1 \\ 2 \end{bmatrix}\right) = 1, & dx\left(\begin{bmatrix} -1 \\ 3 \end{bmatrix}\right) = -1, & dx\left(\begin{bmatrix} 2 \\ 2 \end{bmatrix}\right) = 2, & dx\left(\begin{bmatrix} 3 \\ 1 \end{bmatrix}\right) = 3, \\
dx\left(\begin{bmatrix} 1 \\ -2 \end{bmatrix}\right) = 1, & dx\left(\begin{bmatrix} -1 \\ -2 \end{bmatrix}\right) = -1, & dx\left(\begin{bmatrix} 2 \\ -2 \end{bmatrix}\right) = 2, & dx\left(\begin{bmatrix} 3 \\ -1 \end{bmatrix}\right) = 3.
\end{array}
$$

Consider how this is illustrated in the Fig. 5.1. The vertical lines given by the equations $x = -1, x = 0, x = 1, x = 2, x = 3$, etc. act as illustrations of the projections. For example, the vectors $\begin{bmatrix} 2 \\ 5 \end{bmatrix}$, $\begin{bmatrix} 2 \\ 2 \end{bmatrix}$, and $\begin{bmatrix} 2 \\ -2 \end{bmatrix}$ all terminate on the line $x = 2$. Thus each of these vectors "pierces" the $x = 1$ and $x = 2$ lines. If the vector terminates on a line we consider that the vector 'pierces" the line. Instead of thinking of dx as a projection onto the ∂_x-axis we instead picture dx as the infinite series of lines $x = \pm n$, $n = 0, 1, 2, 3, \ldots$. Using this image we interpret $dx(v)$ as the number of lines that the vector v pierces.

Along with thinking about dx as the lines $x = \pm n$ we also have to attach an orientation. That is shown in the picture as the arrow pointing to the right indicating a positive orientation. Thus the vectors $\begin{bmatrix} -1 \\ 4 \end{bmatrix}$, $\begin{bmatrix} -1 \\ 3 \end{bmatrix}$, and $\begin{bmatrix} -1 \\ -2 \end{bmatrix}$ all pierce one line, the $x = -1$ line. But since the vectors go "backwards" (in the negative direction) we view it as "piercing" negative one times. If we wanted to consider the differential one-form $-dx$ we would have the same picture, the lines given by $x = \pm n$, but we would draw the orientation in the opposite direction.

Now consider the following,

$$
\begin{array}{lll}
dx\left(\begin{bmatrix} 1.42 \\ 4 \end{bmatrix}\right) = 1.42, & dx\left(\begin{bmatrix} 1.97 \\ 2 \end{bmatrix}\right) = 1.97, & dx\left(\begin{bmatrix} 1.1 \\ -2 \end{bmatrix}\right) = 1.1, \\
dx\left(\begin{bmatrix} -1.21 \\ 4 \end{bmatrix}\right) = -1.21, & dx\left(\begin{bmatrix} -1.5 \\ 3 \end{bmatrix}\right) = -1.5, & dx\left(\begin{bmatrix} -1.92 \\ -2 \end{bmatrix}\right) = -1.92,
\end{array}
$$

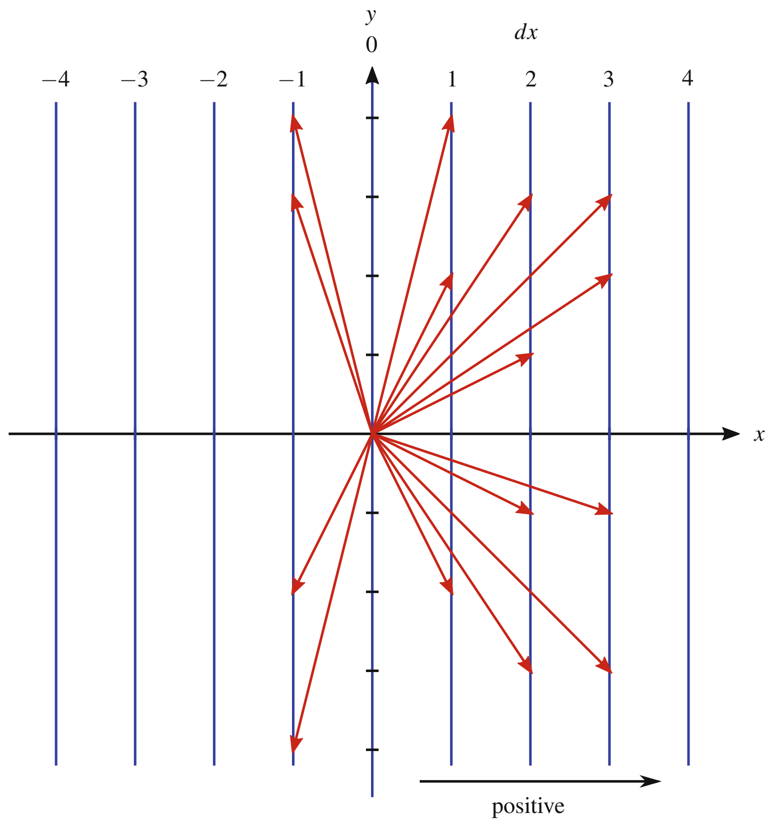


Fig. 5.1 Vectors for which the one-form dx gives outputs of -1, 1, 2, and 3. A positive orientation is shown

$$dx\left(\begin{bmatrix}2.23\\5\end{bmatrix}\right) = 2.23, \qquad dx\left(\begin{bmatrix}2.74\\2\end{bmatrix}\right) = 2.74, \qquad dx\left(\begin{bmatrix}2.01\\-2\end{bmatrix}\right) = 2.01,$$
$$dx\left(\begin{bmatrix}3.69\\6\end{bmatrix}\right) = 3.69, \qquad dx\left(\begin{bmatrix}3.91\\1\end{bmatrix}\right) = 3.91, \qquad dx\left(\begin{bmatrix}3.34\\-1\end{bmatrix}\right) = 3.34.$$

Clearly $dx\left(\begin{bmatrix}1.42\\4\end{bmatrix}\right) = 1.42$ is correct, yet using our picture of dx as the infinite series of lines $x = \pm n$, $n = 0, 1, 2, 3, \ldots$ along with an orientation, the vector $\begin{bmatrix}1.42\\4.84\end{bmatrix}$ only pierces one line $x = 1$. So, using the "piercing" picture for dx we have $dx\left(\begin{bmatrix}1.42\\4\end{bmatrix}\right) = 1$. Clearly this is not exact, but it does give an approximation. Using this idea of piercing we would have

$$dx\left(\begin{bmatrix}1.42\\4\end{bmatrix}\right) = 1, \qquad dx\left(\begin{bmatrix}1.97\\2\end{bmatrix}\right) = 1, \qquad dx\left(\begin{bmatrix}1.1\\-2\end{bmatrix}\right) = 1,$$
$$dx\left(\begin{bmatrix}-1.21\\4\end{bmatrix}\right) = -1, \qquad dx\left(\begin{bmatrix}-1.5\\3\end{bmatrix}\right) = -1, \qquad dx\left(\begin{bmatrix}-1.92\\-2\end{bmatrix}\right) = -1.$$

In an identical manor we can represent the differential form $dy_p \in T_p^*\mathbb{R}^2$ as an infinite series of lines $y = \pm n$, $n = 0, 1, 2, 3, \ldots$, along with an orientation, in the vector space $T_p\mathbb{R}^2$. The value of $dy_p(v_p)$ is the number of lines the vector v_p pierces. Again, if the tip of the vector touches a line, even if it does not go through the line, we consider it to pierce the line.

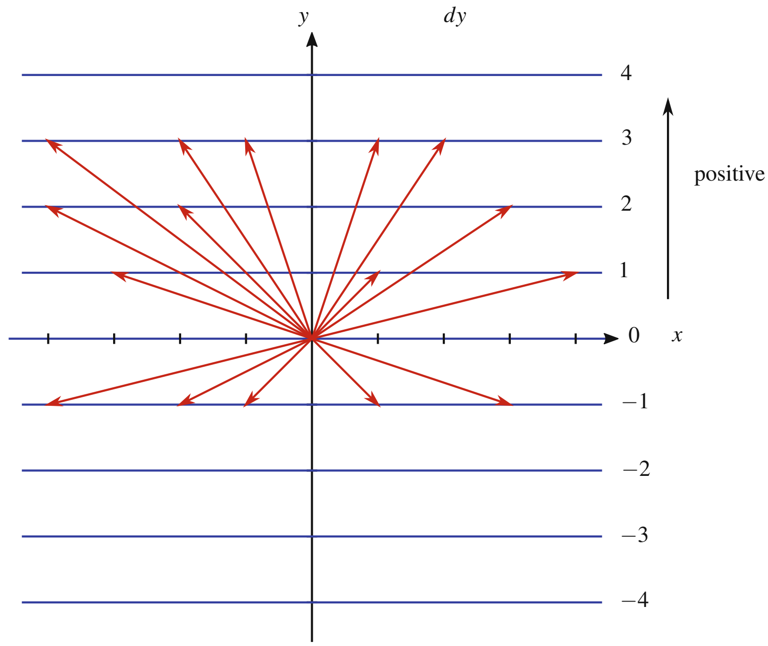


Fig. 5.2 Vectors for which the one-form dy gives outputs of -1, 1, 2, and 3. A positive orientation is shown

So for Fig. 5.2 we have

$$dy\left(\begin{bmatrix}-2\\3\end{bmatrix}\right)=3,\quad dy\left(\begin{bmatrix}-2\\2\end{bmatrix}\right)=2,\quad dy\left(\begin{bmatrix}-3/2\\1\end{bmatrix}\right)=1,\quad dy\left(\begin{bmatrix}-2\\-1\end{bmatrix}\right)=-1,$$
$$dy\left(\begin{bmatrix}1\\3\end{bmatrix}\right)=3,\quad dy\left(\begin{bmatrix}1/2\\2\end{bmatrix}\right)=2,\quad dy\left(\begin{bmatrix}1\\1\end{bmatrix}\right)=1,\quad dy\left(\begin{bmatrix}-1\\-1\end{bmatrix}\right)=-1,$$
$$dy\left(\begin{bmatrix}4\\3\end{bmatrix}\right)=3,\quad dy\left(\begin{bmatrix}3\\2\end{bmatrix}\right)=2,\quad dy\left(\begin{bmatrix}3\\1\end{bmatrix}\right)=1,\quad dy\left(\begin{bmatrix}1\\-1\end{bmatrix}\right)=-1.$$

The same comments regarding approximations apply. For example, the vector $\begin{bmatrix}1.42\\4.84\end{bmatrix}$ pierces four lines of our picture of dy so we would have $dy\left(\begin{bmatrix}1.42\\4.84\end{bmatrix}\right)=4$, though the exact value would of course be 4.84.

But the differential forms dx and dy are not the only elements of $T_p^*\mathbb{R}^2$. How would the differential form $2dx_p \in T_p^*\mathbb{R}^2$ be represented in the vector space $T_p\mathbb{R}^2$? Considering the same vectors from above we have

$$2dx\left(\begin{bmatrix}1\\4\end{bmatrix}\right)=2,\quad 2dx\left(\begin{bmatrix}-1\\4\end{bmatrix}\right)=-2,\quad 2dx\left(\begin{bmatrix}2\\5\end{bmatrix}\right)=4,\quad 2dx\left(\begin{bmatrix}3\\6\end{bmatrix}\right)=6,$$
$$2dx\left(\begin{bmatrix}1\\2\end{bmatrix}\right)=2,\quad 2dx\left(\begin{bmatrix}-1\\3\end{bmatrix}\right)=-2,\quad 2dx\left(\begin{bmatrix}2\\2\end{bmatrix}\right)=4,\quad 2dx\left(\begin{bmatrix}3\\1\end{bmatrix}\right)=6,$$
$$2dx\left(\begin{bmatrix}1\\-2\end{bmatrix}\right)=2,\quad 2dx\left(\begin{bmatrix}-1\\-2\end{bmatrix}\right)=-2,\quad 2dx\left(\begin{bmatrix}2\\-2\end{bmatrix}\right)=4,\quad 2dx\left(\begin{bmatrix}3\\-1\end{bmatrix}\right)=6.$$

So the vectors are piercing twice as many lines in the $2dx$ case as in the dx case, implying that the lines in the $2dx$ case are twice as "dense" as in the dx case. We would draw that as in Fig. 5.3.

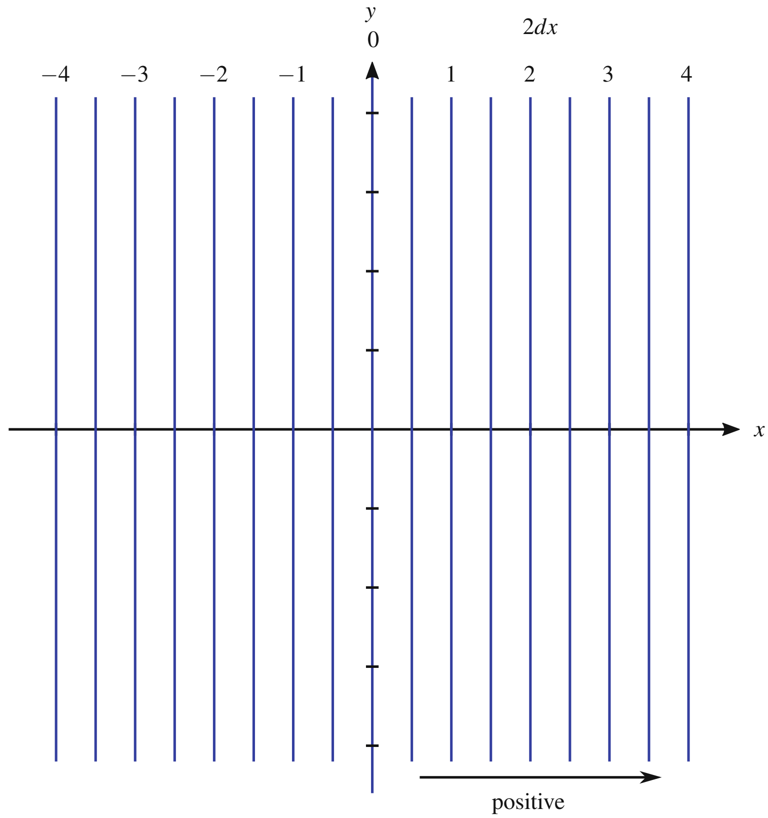


Fig. 5.3 The one-form $2dx$ shown as lines-to-be-pierced

We could perform exactly the same analysis in the $2dy$ case. Again, using the same vectors as above we would have

$$
\begin{array}{llll}
2dy\left(\begin{bmatrix} -2 \\ 3 \end{bmatrix}\right) = 6, & 2dy\left(\begin{bmatrix} -2 \\ 2 \end{bmatrix}\right) = 4, & 2dy\left(\begin{bmatrix} -3/2 \\ 1 \end{bmatrix}\right) = 2, & 2dy\left(\begin{bmatrix} -2 \\ -1 \end{bmatrix}\right) = -2, \\
2dy\left(\begin{bmatrix} 1 \\ 3 \end{bmatrix}\right) = 6, & 2dy\left(\begin{bmatrix} 1/2 \\ 2 \end{bmatrix}\right) = 4, & 2dy\left(\begin{bmatrix} 1 \\ 1 \end{bmatrix}\right) = 2, & 2dy\left(\begin{bmatrix} -1 \\ -1 \end{bmatrix}\right) = -2, \\
2dy\left(\begin{bmatrix} 4 \\ 3 \end{bmatrix}\right) = 6, & 2dy\left(\begin{bmatrix} 3 \\ 2 \end{bmatrix}\right) = 4, & 2dy\left(\begin{bmatrix} 3 \\ 1 \end{bmatrix}\right) = 2, & 2dy\left(\begin{bmatrix} 1 \\ -1 \end{bmatrix}\right) = -2.
\end{array}
$$

Again, the vectors are piercing twice as many lines in the $2dy$ case as in the dy case, so again the lines in the $2dy$ case are twice as "dense" as in the dy case. We would draw that as in Fig. 5.4.

Clearly if you performed the same analysis on $3dx$ and $3dy$ you would find the lines in the picture representations three times as "dense" as in the dx and dy case, and for ndx and ndy they would be n times as dense.

Question 5.3 Draw a picture representing the following differential forms. $3dx$, $5dy$, $\frac{1}{2}dx$, $\frac{2}{3}dx$, $\frac{1}{3}dy$, and $\frac{3}{5}dy$.

Now we turn our attention to more general elements of $T_p^*\mathbb{R}^2$. We will start off with something simple, $dx + dy$. Consider how the one-form $dx + dy$ acts on the following vectors,

$$
\begin{array}{ll}
(dx+dy)\left(\begin{bmatrix} -1 \\ 3 \end{bmatrix}\right) = -1 + 3 = 2, & (dx+dy)\left(\begin{bmatrix} 1 \\ 2 \end{bmatrix}\right) = 1 + 2 = 3, \\
(dx+dy)\left(\begin{bmatrix} -1 \\ 1 \end{bmatrix}\right) = -1 + 1 = 0, & (dx+dy)\left(\begin{bmatrix} 0 \\ 1 \end{bmatrix}\right) = 0 + 1 = 1,
\end{array}
$$

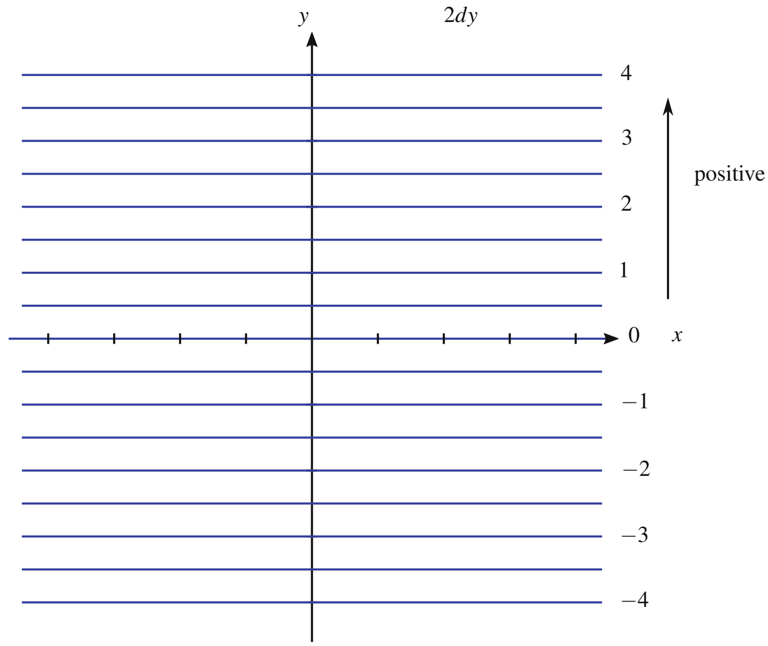


Fig. 5.4 The one-form $2dy$ shown as lines-to-be-pierced

$$(dx+dy)\left(\begin{bmatrix}1\\1\end{bmatrix}\right)=1+1=2, \qquad (dx+dy)\left(\begin{bmatrix}-1\\0\end{bmatrix}\right)=-1+0=-1,$$

$$(dx+dy)\left(\begin{bmatrix}1\\0\end{bmatrix}\right)=1+0=1, \qquad (dx+dy)\left(\begin{bmatrix}0\\-1\end{bmatrix}\right)=0-1=-1.$$

Figure 5.5 is a picture for $dx + dy$ that works. The fact that it works can be seen just by inspecting the vectors and counting how many lines they pierce, taking into account the orientation depicted. The picture for $dx + dy$ clearly requires slanted lines. By inspection it should be apparent that the lines are given by $y = -x + n$ for $n = 0, \pm 1, \pm 2, \ldots$.

Now suppose we want to draw a picture for the differential one-form $2dx+dy$ or $dx-2dy$ or something more complicated still. We need a systematic way to decide what the lines are. This is actually fairly easy to do. Consider the one-form $2dx+dy$. To find the lines we need to draw to make a picture of this differential we simply start with the equation $2x + y = n$ where $n = 0, \pm 1, \pm 2, \ldots$ and solve for y which gives $y = -2x + n$, depicted in Fig. 5.6. Consider the vector $v = \begin{bmatrix}1\\2\end{bmatrix}$ in Fig. 5.6. It "pierces" four lines in the positive direction, so according to the picture $(2dx + dy)(v) = 4$. Computationally we have

$$(2dx+dy)\left(\begin{bmatrix}1\\2\end{bmatrix}\right)=2dx\left(\begin{bmatrix}1\\2\end{bmatrix}\right)+dy\left(\begin{bmatrix}1\\2\end{bmatrix}\right)=2(1)+2=4.$$

Similarly, to draw the picture for the one-form $dx - 2dy$ we use the equation $x - 2y = n$, which becomes $y = \frac{x}{2} - \frac{n}{2}$. See Fig. 5.7.

Question 5.4 Check that the various vectors pictured in the images for $2dx + dy$ and $dx - 2dy$ give the same answers both computationally and pictorially.

Question 5.5 Explain why the above procedure for finding the lines in the pictorial representation works. To do this consider how a generic two-form $adx + bdy$ acts on a generic vector $v = \begin{bmatrix}x\\y\end{bmatrix}$. Which values of this vector give the value $(adx + bdy)(v) = n$ for any $n = 0, \pm 1, \pm 2, \ldots$?

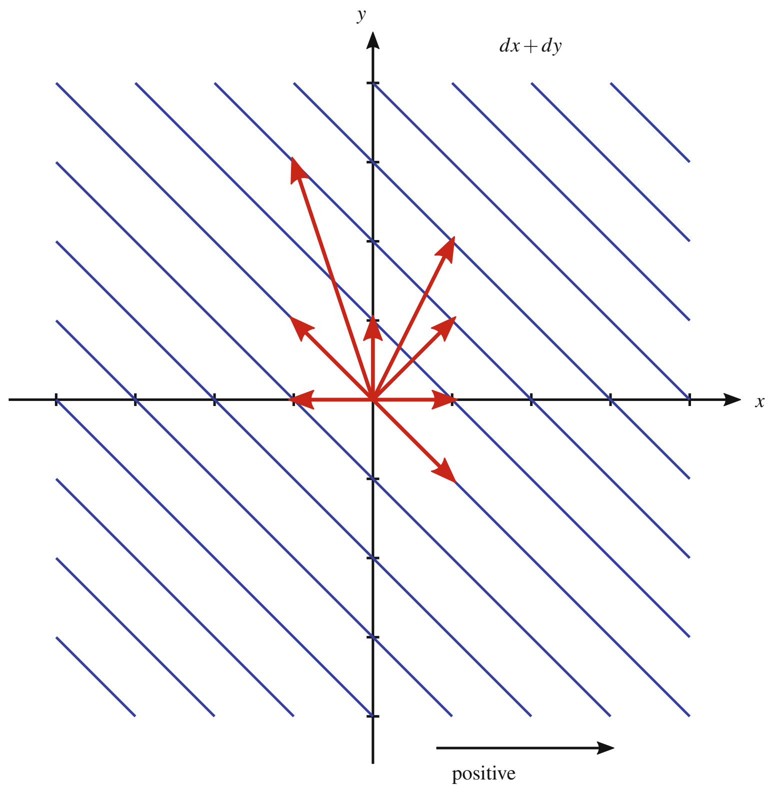


Fig. 5.5 The lines-to-be-pieced for the one-form $dx + dy$. The positive orientation is depicted

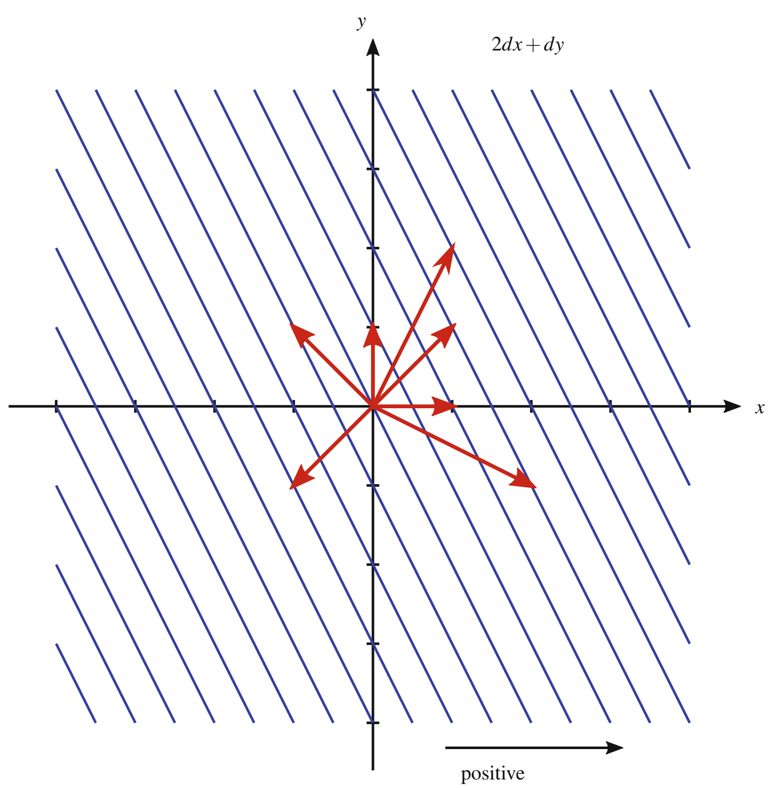


Fig. 5.6 The lines-to-be-pieced for the one-form $2dx + dy$. The positive orientation is depicted

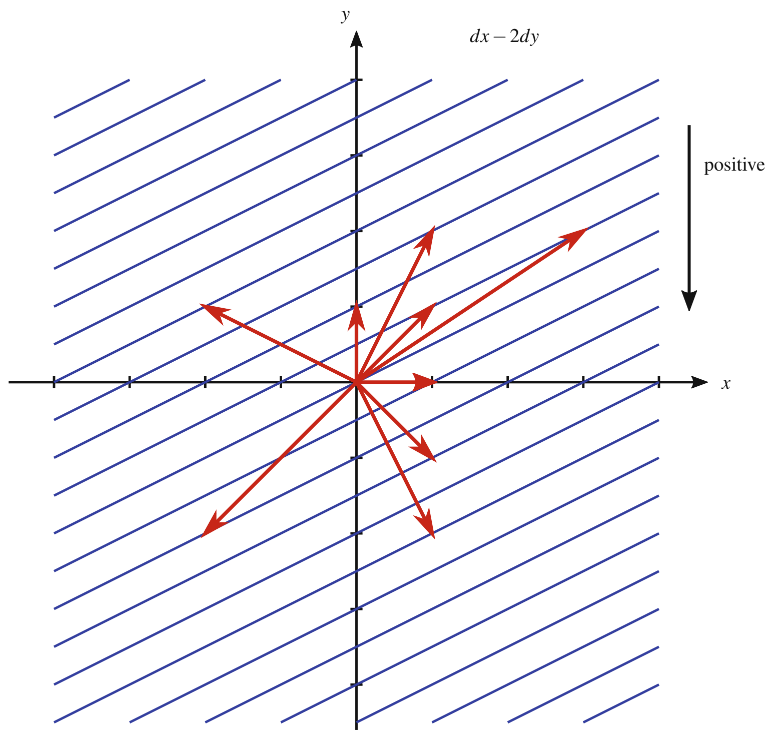


Fig. 5.7 The lines-to-be-pierced for the one-form $dx - 2dy$. The positive orientation is depicted

Question 5.6 Draw the graphical representations of the following differential forms,

(a) $3dx + dy$,
(b) $2dx - dy$,
(c) $-dx + 3dy$.

Differential one-forms in $T_p^*\mathbb{R}^2$ are not the only kind of forms that can act on vectors in $T_p\mathbb{R}^2$. Differential two-forms in $\bigwedge_p^2(\mathbb{R})$ can also act on vectors in $T_p\mathbb{R}^2$. As we know, differential two-forms act on two vectors in $T_p\mathbb{R}^2$. We begin by considering the vectors

$$v_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \quad v_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

and the differential two-form $dx \wedge dy$. We have

$$(dx \wedge dy)(v_1, v_2) = \begin{vmatrix} dx(v_1) & dx(v_2) \\ dy(v_1) & dy(v_2) \end{vmatrix} = \begin{vmatrix} 1 & 0 \\ 0 & 1 \end{vmatrix} = 1,$$

$$(dx \wedge dy)(v_2, v_1) = \begin{vmatrix} dx(v_2) & dx(v_1) \\ dy(v_2) & dy(v_1) \end{vmatrix} = \begin{vmatrix} 0 & 1 \\ 1 & 0 \end{vmatrix} = -1.$$

Like before, we want to draw a picture in $T_p\mathbb{R}^2$ that represents the differential two-form $dx \wedge dy$. For differential one-forms we used a picture that consisted of lines and then we counted how many lines our vector "pierced" in order to get an approximate value for the differential form acting on the vector. Here our picture has to somehow incorporate both input vectors. We represent the differential two-form $dx \wedge dy$ in $T_p\mathbb{R}^2$ as horizontal and vertical lines along with an orientation.

Figure 5.8 shows the picture of $dx \wedge dy$, consisting of blue horizontal lines and green vertical lines, which produces a grid of squares. The parallelepiped spanned by the vectors v_1 and v_2 exactly covers one of the grid's squares, thus we say $dx \wedge dy(v_1, v_2)$ is one. Now we only have to consider orientation. Since $(dx \wedge dy)(v_1, v_2)$ is positive we use that to decide on the orientation that is needed, which is depicted in the picture as a small circular arrow that goes in the counter-clockwise direction. Going from v_1 to v_2, traversing the smallest angle between the two vectors, is counter-clockwise, so that is the

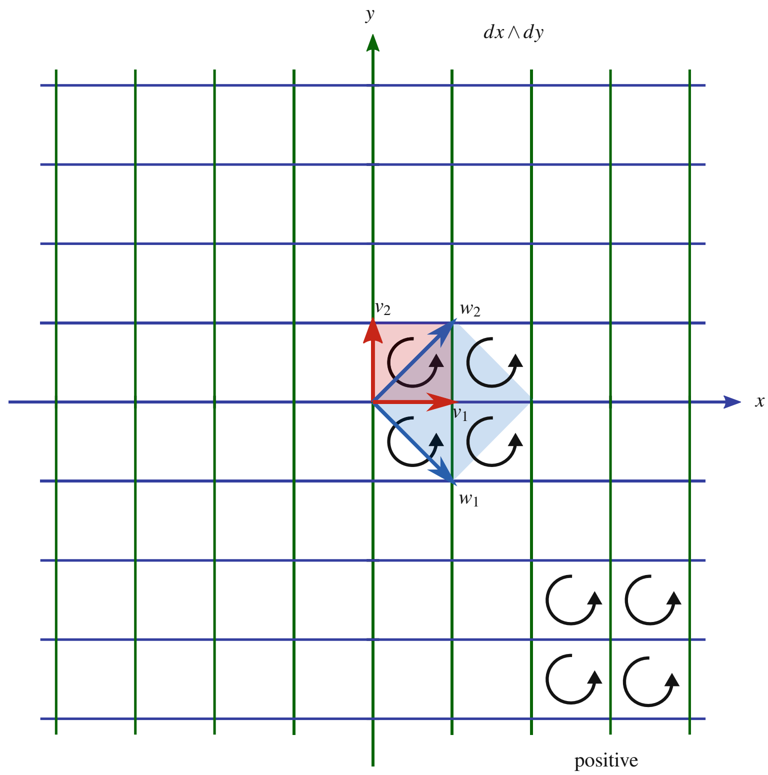


Fig. 5.8 The two-form $dx \wedge dy$ consists of both horizontal and vertical lines. The counter-clockwise circular arrows depicts a positive orientation

orientation that is chosen. Thus, if we traverse the small angle between the first and second vectors in a counter-clockwise direction we choose positive, and if we traverse the small angle between the first and second vectors in a clockwise direction we choose negative. Thus for $(dx \wedge dy)(v_2, v_1)$ we have the first vector being v_2 and the second vector being v_1. Traversing from the first vector to the second we go in a clockwise direction, so we would choose negative. The parallelepiped spanned by the vectors v_2 and v_1 covers exactly one square so the answer is -1.

Now consider the vectors

$$w_1 = \begin{bmatrix} 1 \\ -1 \end{bmatrix}, \; w_2 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}.$$

Computationally we have that

$$(dx \wedge dy)(w_1, w_2) = \begin{vmatrix} dx(w_1) & dx(w_2) \\ dy(w_1) & dy(w_2) \end{vmatrix} = \begin{vmatrix} 1 & -1 \\ 1 & 1 \end{vmatrix} = 2,$$

$$(dx \wedge dy)(w_2, w_1) = \begin{vmatrix} dx(w_2) & dx(w_1) \\ dy(w_2) & dy(w_1) \end{vmatrix} = \begin{vmatrix} 1 & 1 \\ 1 & -1 \end{vmatrix} = -2.$$

Using Fig. 5.8 we can see that the parallelepiped spanned by w_1 and w_2 covers a total of two grid squares and if we traverse the small angle from w_1 to w_2 it is in the counter-clockwise direction so we choose the positive sign to get $dx \wedge dy(w_1, w_2) = 2$. But if we traverse the small angle from w_2 to w_1 it is in the clockwise direction so we choose the negative sign to get $dx \wedge dy(w_2, w_1) = -2$.

Now we turn our attention to the differential two-form $2dx \wedge dy$. We have already discussed how this way of visualizing differential forms has the drawback that the answers it gives are only approximate. Now we will illustrate another drawback. There can be some ambiguity in what picture or image we use for some differential forms. Again, we will consider the vectors

$$v_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \; v_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix},$$

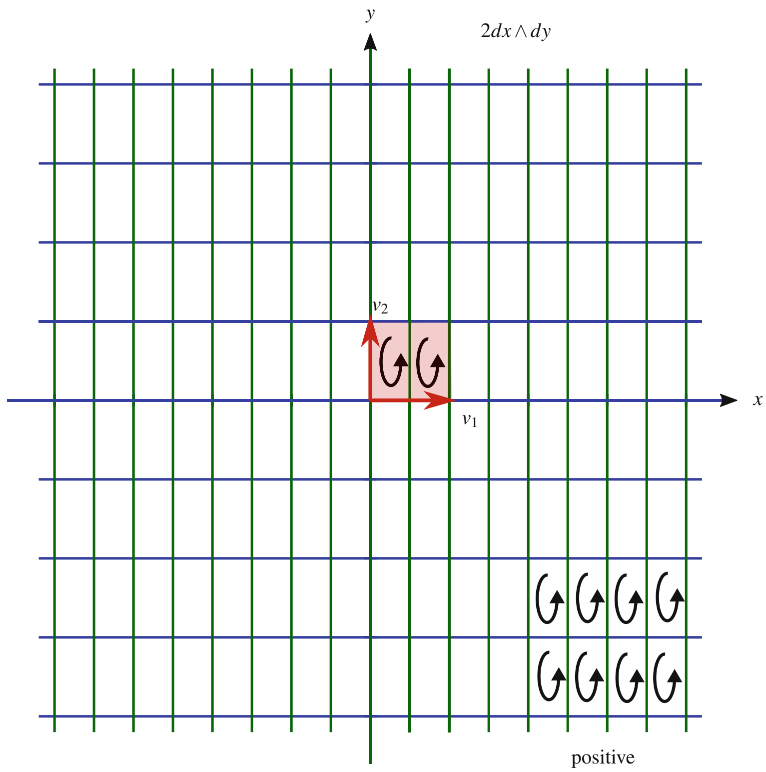


Fig. 5.9 One possible way to draw the two-form $2dx \wedge dy$. Here the lines are $x = n/2$ and $y = n$. The counter-clockwise circular arrows depicts a positive orientation

which computationally give us

$$(2dx \wedge dy)(v_1, v_2) = 2 \begin{vmatrix} 1 & 0 \\ 0 & 1 \end{vmatrix} = 2.$$

So the parallelepiped spanned by v_1 and v_2 has to cover two boxes from the grid that represents the two-form $2dx \wedge dy$. There is no unique way of doing that. For example, choosing the lines $y = n$ and $x = \frac{n}{2}$ for $n = 0, \pm 1, \pm 2, \ldots$ accomplishes this; see Fig. 5.9. Another way of accomplishing this is choosing the lines $y = \frac{n}{2}$ and $x = n$ for $n = 0, \pm 1, \pm 2, \ldots$, as in Fig. 5.10. And in fact, there are other choices as well.

Question 5.7 Find another graphical representation for the differential two-form $2dx \wedge dy$.

Question 5.8 Find three graphical representations for each of the following differential two-forms

(a) $3dx \wedge dy$,
(b) $5dx \wedge dy$,
(c) $\frac{1}{2} dx \wedge dy$.

5.2 One-Forms in $\mathbb{R}^3$

As before we will fix some point p on the manifold $\mathbb{R}^3$ and only consider one-forms in $T_p^*\mathbb{R}^3$. The "picture" of the one-forms that we will draw are actually superimposed on the vector space $T_p\mathbb{R}^3$. We will start out with the one-form $dx \in T_p^*\mathbb{R}^3$.

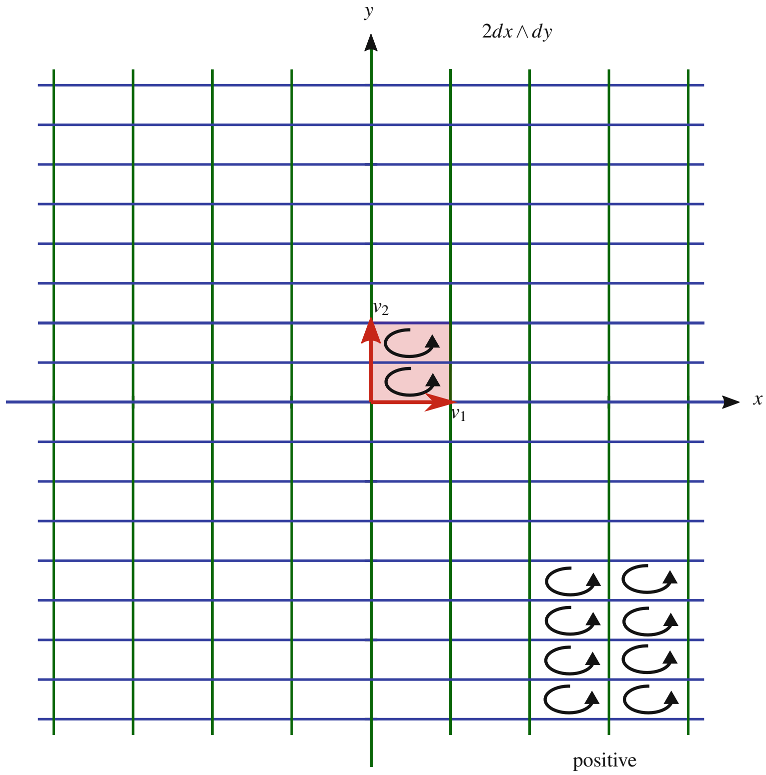


Fig. 5.10 Another possible way to draw the two-form $2dx \wedge dy$. Here the lines are $x = n$ and $y = n/2$. The counter-clockwise circular arrows depicts a positive orientation

Consider what dx does to the following vectors,

$$dx\left(\begin{bmatrix}1\\0\\5\end{bmatrix}\right) = 1, \qquad dx\left(\begin{bmatrix}1\\0\\0\end{bmatrix}\right) = 1, \qquad dx\left(\begin{bmatrix}1\\2\\1\end{bmatrix}\right) = 1,$$

$$dx\left(\begin{bmatrix}1\\2\\-1\end{bmatrix}\right) = 1, \qquad dx\left(\begin{bmatrix}1\\-3\\-2\end{bmatrix}\right) = 1, \qquad dx\left(\begin{bmatrix}1\\-4\\3\end{bmatrix}\right) = 1.$$

Any vector $v \in T_p\mathbb{R}^3$ that terminates anywhere on the $x = 1$ plane in $T_p\mathbb{R}^3$ has the value 1. See Fig. 5.11.

Similarly, consider what dx does to these vectors,

$$dx\left(\begin{bmatrix}3\\0\\0\end{bmatrix}\right) = 3, \qquad dx\left(\begin{bmatrix}2\\-2\\-1\end{bmatrix}\right) = 2, \qquad dx\left(\begin{bmatrix}0\\3\\4\end{bmatrix}\right) = 0,$$

$$dx\left(\begin{bmatrix}3\\-1\\2\end{bmatrix}\right) = 3, \qquad dx\left(\begin{bmatrix}2\\-2\\-4\end{bmatrix}\right) = 2, \qquad dx\left(\begin{bmatrix}0\\-3\\-1\end{bmatrix}\right) = 0,$$

$$dx\left(\begin{bmatrix}3\\2\\-5\end{bmatrix}\right) = 3, \qquad dx\left(\begin{bmatrix}2\\5\\-3\end{bmatrix}\right) = 2, \qquad dx\left(\begin{bmatrix}0\\-2\\5\end{bmatrix}\right) = 0.$$

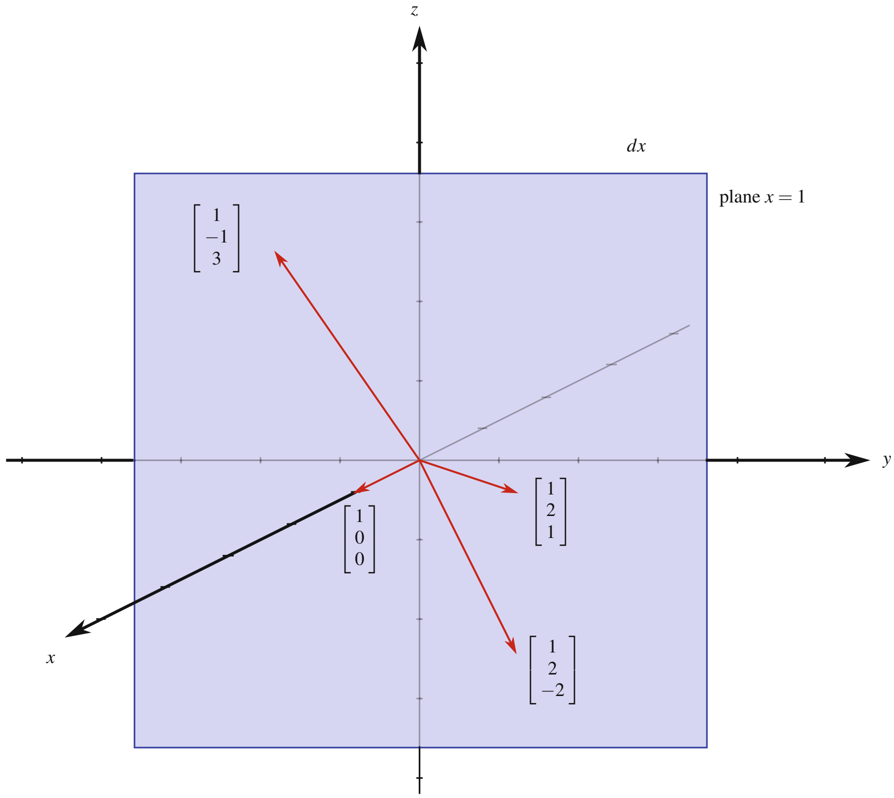


Fig. 5.11 The plane $x = 1$ in $\mathbb{R}^3$. The one-form dx sends all of these vectors that terminate on this plane to the value 1. That is, if v terminates on this plane then $dx(v) = 1$

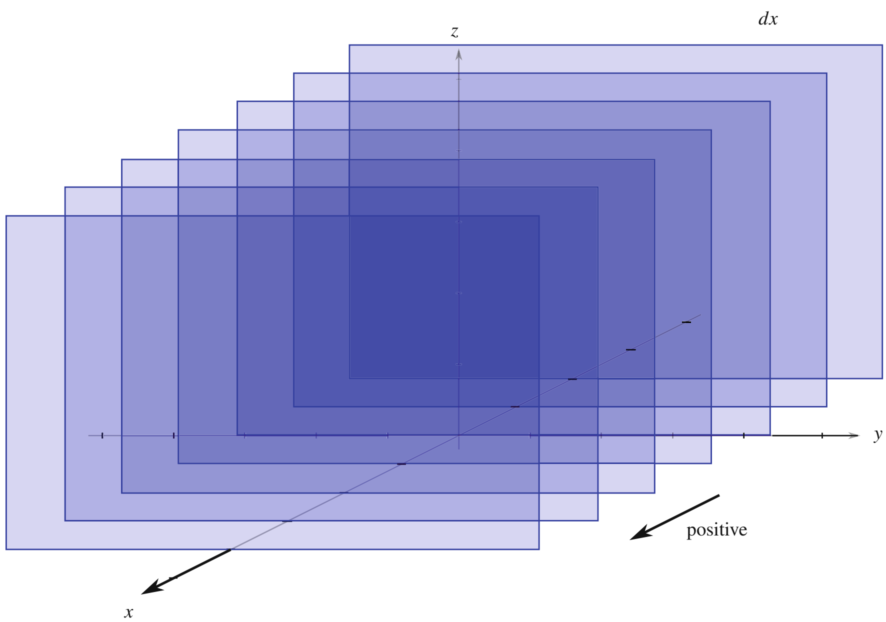


Fig. 5.12 The one-form dx depicted as planes in $\mathbb{R}^3$

The one-form sends any vector that terminates on the plane $x = 3$ to 3, any vector that terminates on the plane $x = 2$ to 2, and any vector that terminates on the plane $x = 0$ to 0. Based on this we will graphically represent the differential one-form $dx \in T_p^*\mathbb{R}^3$ as the set of planes $x = n$, where $n = 0, \pm 1, \pm 2, \ldots$, in the vector space $T_p\mathbb{R}^3$. In Fig. 5.12 we show dx by drawing the seven planes $x = 4$, $x = 3$, $x = 2$, $x = 1$, $x = 0$, $x = -1$, and $x = -2$.

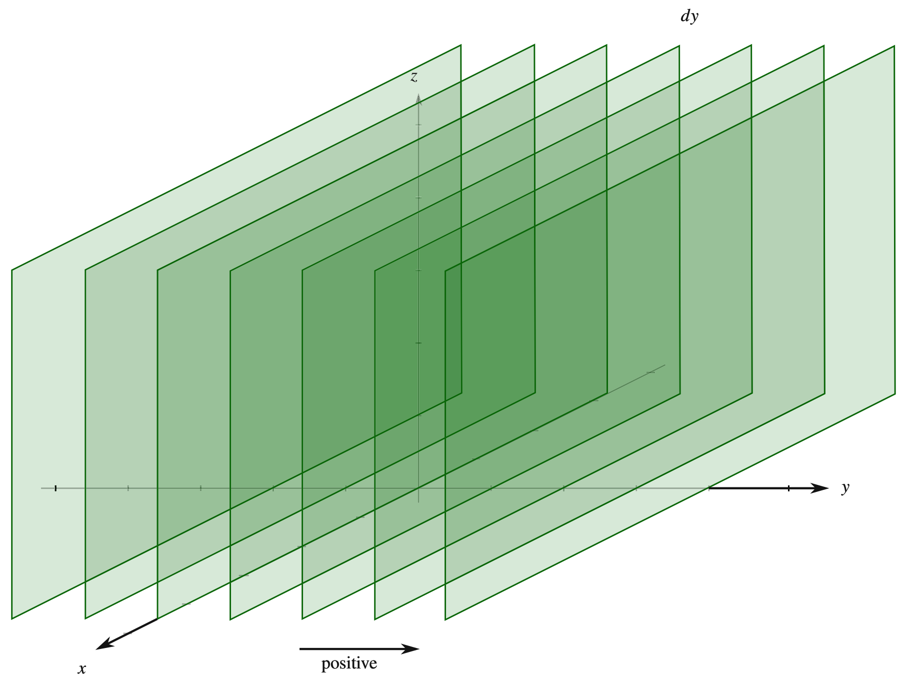


Fig. 5.13 The one-form dy depicted as planes in $\mathbb{R}^3$

Using this picture we say that dx of a vector v is the number of planes that v "pierces." And as before, the answer that this gives is only an approximate answer. For example, the vectors

$$v_1 = \begin{bmatrix} 1 \\ -1 \\ 7 \end{bmatrix}, \quad v_2 = \begin{bmatrix} 1.4 \\ 1 \\ -2 \end{bmatrix}, \quad v_3 = \begin{bmatrix} 1.999 \\ -2 \\ 5 \end{bmatrix}$$

all pierce only one plane, so we would say they all have the same value, 1, though clearly the real answers are $dx(v_1) = 1$, $dx(v_2) = 1.4$, and $dx(v_3) = 1.999$. Again, an orientation needs to be assigned to the picture.

A similar line of reasoning results in dy being graphically represented as the set of planes $y = n$, where $n = 0, \pm 1, \pm 2, \ldots$, see Fig. 5.13, and dz being graphically represented as the set of planes $z = n$, where $n = 0, \pm 1, \pm 2, \ldots$, see Fig. 5.14.

Question 5.9 Show that the positive orientations depicted in Figs. 5.12, 5.13, and 5.14 are correct.

Let us take a moment to compare between the way we view a differential one-form as an element of the cotangent space $T_p^*\mathbb{R}^3$ and our way of visualizing a differential form on $T_p\mathbb{R}^3$. See Fig. 5.15. A differential one-form in $T_p^*\mathbb{R}^3$ was also called a co-vector which is drawn in the cotangent space exactly as a vector would be drawn in the tangent space. For example, the one-form dx is the unit vector along the dx-axis, the one-form dy is the unit vector along the dy-axis, and the one-form dz is the unit vector along the dz-axis. Recall that we chose to always write co-vectors as row vectors, thus in $T_p^*\mathbb{R}^3$ we have $dx = [1, 0, 0]$, $dy = [0, 1, 0]$, and $dz = [0, 0, 1]$ so the one-form $dx + dy + dz = [1, 1, 1]$. In the top of Fig. 5.15 we show each of the co-vectors dx in blue, dy in green, and dz in red.

Using the method of visualizing differential forms that we have introduced in this chapter, the differential one-form dx is regarded as an infinite number of planes in $T_p\mathbb{R}^3$ that are perpendicular to the $\frac{\partial}{\partial x}$-axis, also denoted the ∂_x-axis. Likewise, the differential one-form dy is an infinite number of planes perpendicular to the ∂_y-axis and dz is an infinite number of planes perpendicular to the ∂_z-axis. In the bottom of Fig. 5.15 we show a few planes of dx in blue, dy in green, and dz in red.

Like the one-forms dx, dy, and dz, general one-forms $adx + bdy + cdz$, where $a, b, c \in \mathbb{R}$, can also be visualized as an infinite number of planes. As an example we will consider the one-form $dx + dy$. In essence, we want to somehow add the

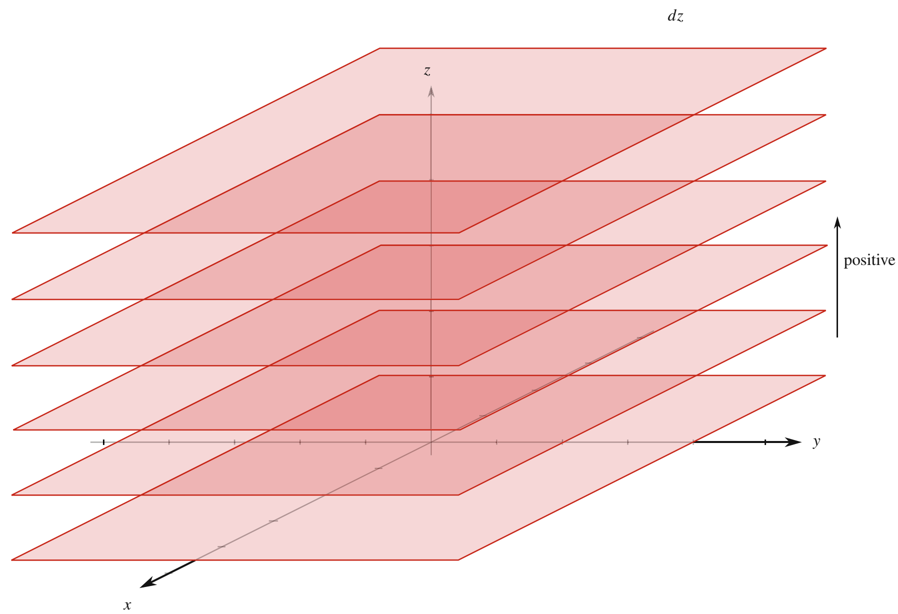


Fig. 5.14 The one-form dz depicted as planes in $\mathbb{R}^3$

dx planes with dy planes. See Fig. 5.16. In essence the planes of dx are perpendicular to the vector

$$\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$$

and the planes of dy are perpendicular to the vector

$$\begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$$

so the sum of dx and dy, or $dx + dy$ will be planes perpendicular to the sum of the vectors, that is, the planes of $dx + dy$ are perpendicular to the vector

$$\begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}.$$

The answer is shown in Fig. 5.17.

Misner, Thorne, and Wheeler's book *Gravitation* describes graphical method to find the planes associated with $dx + dy$ by choosing appropriate vectors (like v_1, v_2, w_1, w_2 in the next question), counting the number of planes they pierce, and then using the points of the vectors to determine a plane of $dx + dy$. The following questions leads you through their process.

Question 5.10 Consider the vectors

$$v_1 = \begin{bmatrix} 3 \\ 0 \\ 0 \end{bmatrix}, v_2 = \begin{bmatrix} 3 \\ 0 \\ 2 \end{bmatrix}, w_1 = \begin{bmatrix} 0 \\ 3 \\ 0 \end{bmatrix}, w_2 = \begin{bmatrix} 0 \\ 3 \\ 2 \end{bmatrix}$$

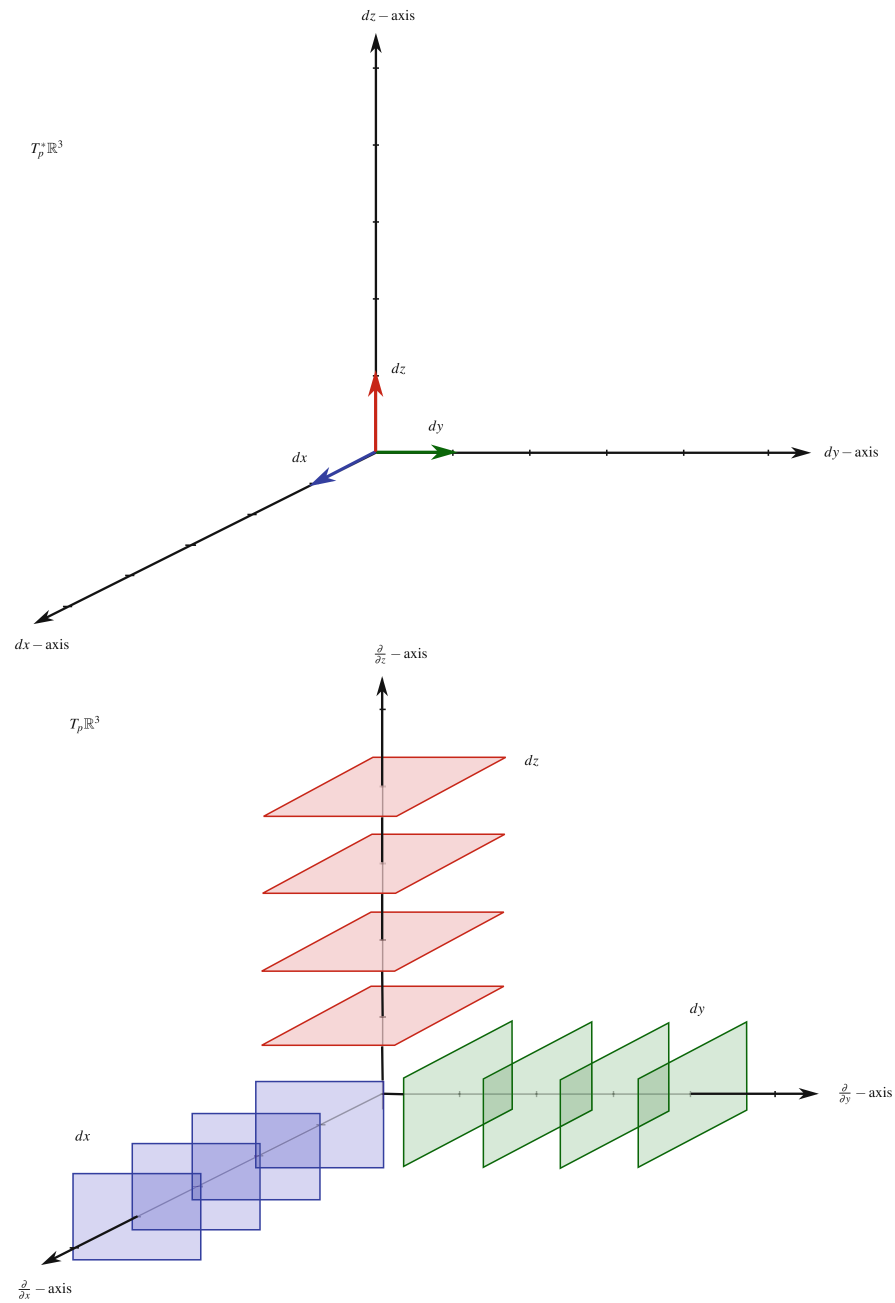


Fig. 5.15 The one-forms dx (blue), dy (green), and dz (red) pictured as co-vectors in $T_p^*\mathbb{R}^3$ (top) and as planes in $T_p\mathbb{R}^3$ (bottom)

that are shown in Fig. 5.16.

(a) Find $dx(v_1)$, $dx(v_2)$, $dx(w_1)$, and $dx(w_2)$ both computationally and graphically from the above picture.
(b) Find $dy(v_1)$, $dy(v_2)$, $dy(w_1)$, and $dy(w_2)$ both computationally and graphically from the above picture.
(c) Find $(dx + dy)(v_1)$, $(dx + dy)(v_2)$, $(dx + dy)(w_1)$, and $(dx + dy)(w_2)$ both computationally and graphically from the above picture.
(d) How many dx planes do v_1 and v_2 pierce? How many dy planes? How many dy planes do w_1 and w_2 pierce? How many dx planes? Now how many $dx + dy$ planes does each of these vectors pierce?
(e) Based on (a)–(d) determine how you would use the vectors v_1, v_2, w_1, w_2 to find the planes of $dx + dy$.

Question 5.11 Sketch the differential one-forms $dy + dz$ and $dx + dz$ on $\mathbb{R}^3$.

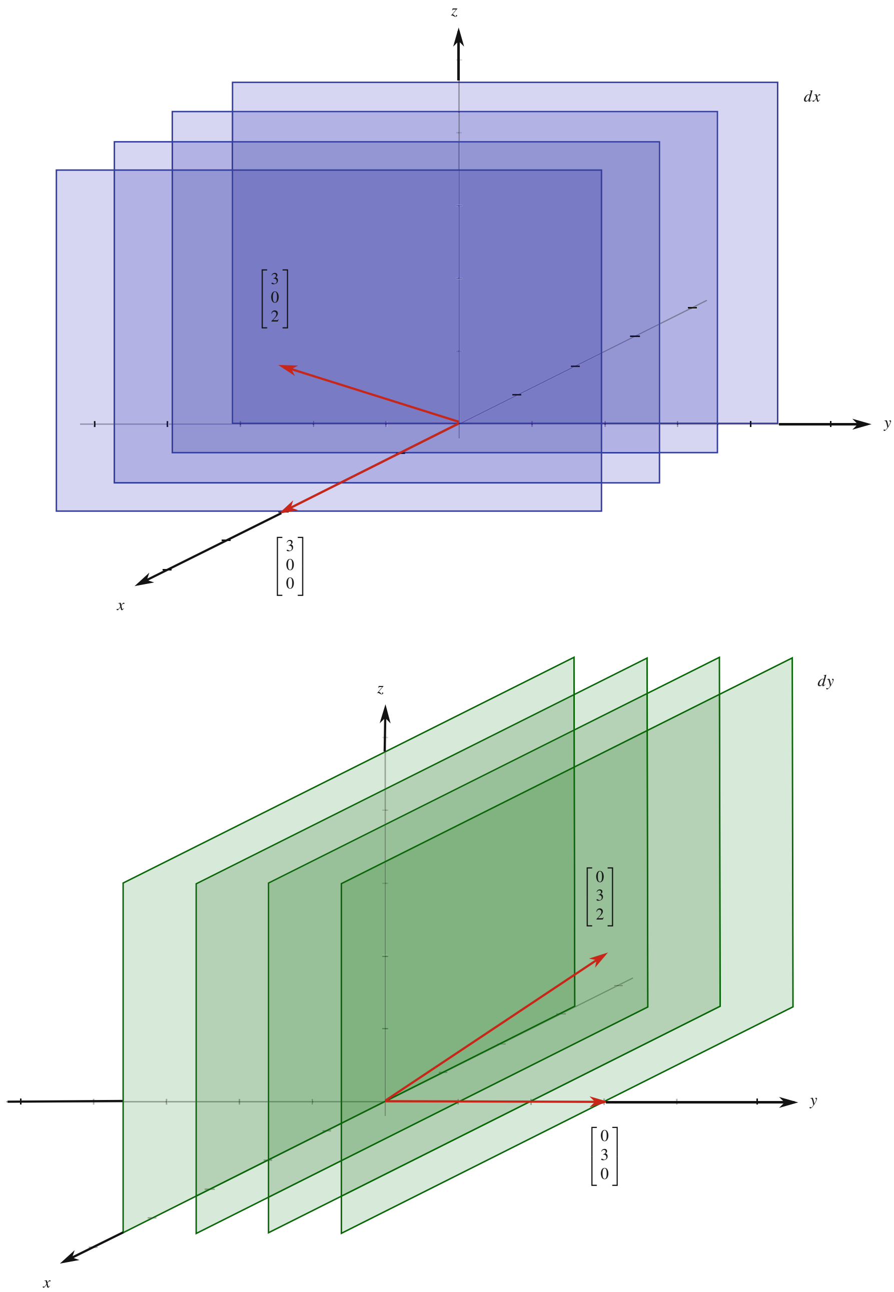


Fig. 5.16 To find the picture for $dx + dy$ we want to somehow "add" the planes for dx (top) with the planes for dy (bottom). This "addition" gives the planes depicted in Fig. 5.17

Now we have all the pieces necessary to compare the representation of $dx + dy + dx$ in $T_p^*\mathbb{R}^3$ and the way we are visualizing this same differential form in $T_p\mathbb{R}^3$, as in Fig. 5.18.

5.3 Two-Forms in $\mathbb{R}^3$

Similar to how we drew two-forms in $T_p\mathbb{R}^2$ we can draw two-forms in $T_p\mathbb{R}^3$. In fact, our graphical representation of two-forms in $T_p\mathbb{R}^3$ will essentially be a three-dimensional version of our grid-lines from the two-dimensional case. Instead of trying to walk through the development of these pictures we will simply present the pictures of the two-forms $dx \wedge dy$, $dy \wedge dz$, and $dz \wedge dx$ and illustrate how they work. We begin by a picture of the two-form $dx \wedge dy$ in Fig. 5.19.

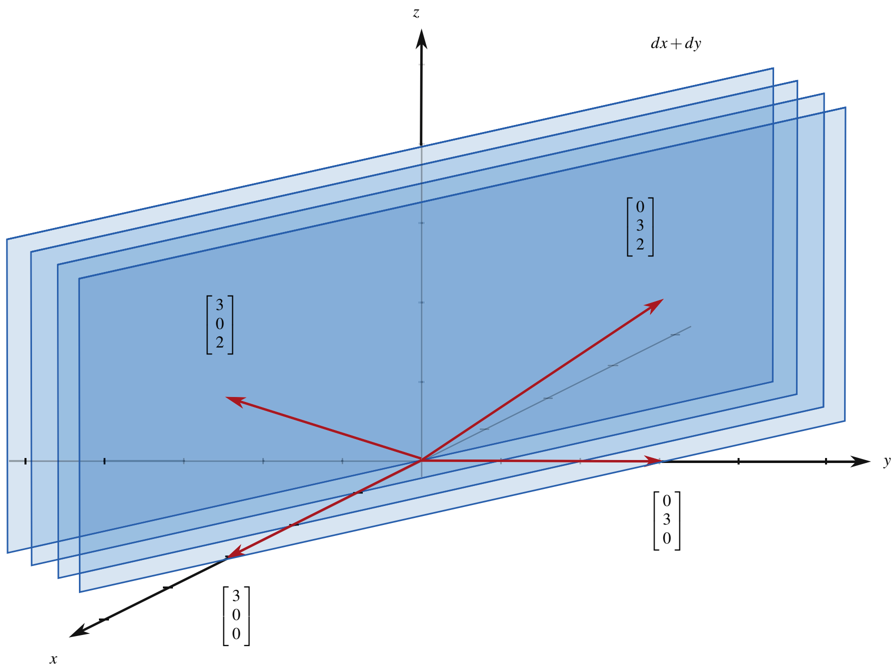


Fig. 5.17 The planes that depict the one-form $dx + dy$

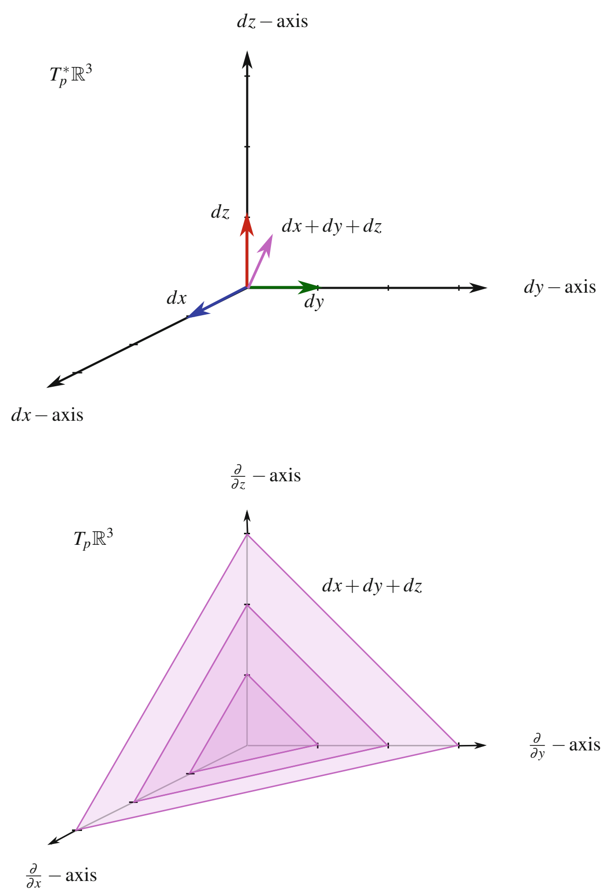


Fig. 5.18 The one-form $dx + dy + dz$ as a co-vector in $T_p^*\mathbb{R}^3$ (top) and the planes that depict $dx + dy + dz$ in $T_p\mathbb{R}^3$ (bottom)

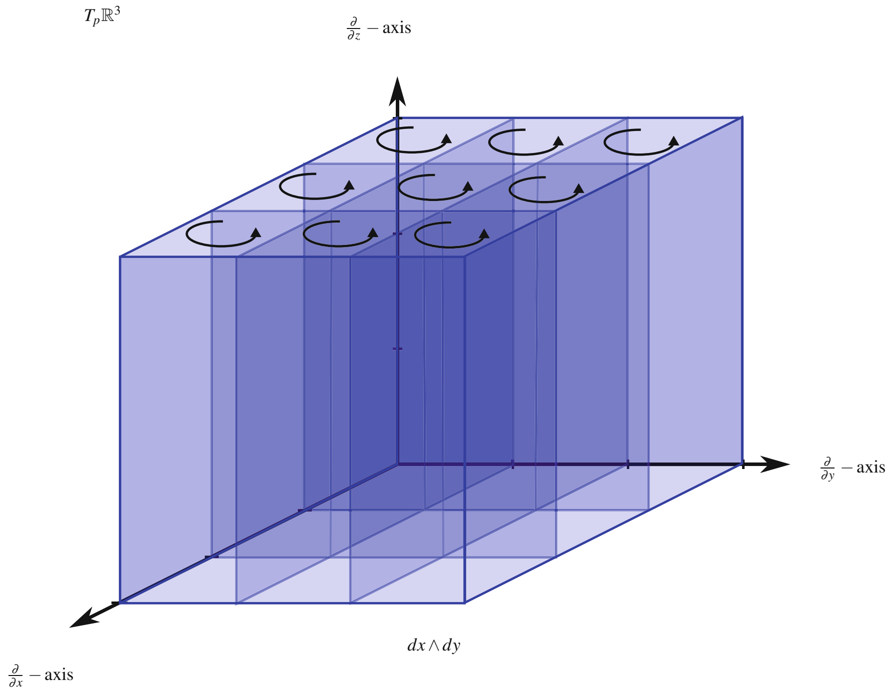


Fig. 5.19 The two-form $dx \wedge dy$ depicted in $T_p\mathbb{R}^3$ as a series of "tubes" made up of planes parallel to the ∂_x and ∂_y axes. The positive orientation is illustrated

Notice that graphically $dx \wedge dy$ is nothing more than both the dx and dy planes being drawn simultaneously which results in "tubes" that go in the ∂_z direction. The orientation is depicted by the small circular arrows in the tubes. Looking down on the picture these arrows are turning in the counter-clockwise direction. Rather than trying to prove that this is an the correct picture we will show that it works with several examples. First consider the vectors

$$u_1 = \begin{bmatrix} 3 \\ 0 \\ 0 \end{bmatrix} \text{ and } u_2 = \begin{bmatrix} 0 \\ 3 \\ 0 \end{bmatrix}.$$

We have

$$(dx \wedge dy)(u_1, u_2) = \begin{vmatrix} 3 & 0 \\ 0 & 3 \end{vmatrix} = 9.$$

By imagining the vectors u_1 and u_2 imposed on Fig. 5.19 you can see that the parallelepiped spanned by them has nine "tubes" of $dx \wedge dy$ passing through it. Furthermore, when traversing from u_1 to u_2 via the smallest angle between the vectors, we are moving in a direction that matches the orientation imposed on the tubes, so the nine is positive. If we traversed from u_2 to u_1 via the smallest angle between the vectors we are moving in a direction counter to the orientation imposed on the tubes, so in this case the nine is negative, just as we would expect from the computation

$$(dx \wedge dy)(u_2, u_1) = \begin{vmatrix} 0 & 3 \\ 3 & 0 \end{vmatrix} = -9.$$

Now imagine the vectors

$$v_1 = \begin{bmatrix} 3 \\ 0 \\ 0 \end{bmatrix} \text{ and } v_2 = \begin{bmatrix} 0 \\ 0 \\ 3 \end{bmatrix}$$

imposed on Fig. 5.19. The parallelepiped spanned by these two vectors is parallel with the tubes and so does not cut through any of the tubes, hence graphically we have $dx \wedge dy(v_1, v_2) = 0$. Computationally we get the same thing

$$(dx \wedge dy)(v_2, v_1) = \begin{vmatrix} 3 & 0 \\ 0 & 0 \end{vmatrix} = 0.$$

Finally, imagine the vectors

$$w_1 = \begin{bmatrix} 0 \\ 3 \\ 0 \end{bmatrix} \text{ and } w_2 = \begin{bmatrix} 0 \\ 0 \\ 3 \end{bmatrix}$$

imposed on Fig. 5.19. Again, the parallelepiped spanned by w_1 and w_2 is parallel with the tubes and so does not cut through any of them, giving $dx \wedge dy(w_1, w_2) = 0$, which again matches the computation

$$(dx \wedge dy)(w_2, w_1) = \begin{vmatrix} 0 & 0 \\ 3 & 0 \end{vmatrix} = 0.$$

The graphical representation for $dy \wedge dz$ is very similar to that of $dx \wedge dy$ and basically consists of both the dy and dz planes resulting in tubes that are parallel with the ∂_x-axis, as in Fig. 5.20. Again we will consider how many tubes the parallelepipeds spanned by u_1 and u_2, by v_1 and v_2, and by w_1 and w_2, intersect. Computationally we have

$$(dy \wedge dz)(u_1, u_2) = \begin{vmatrix} 0 & 3 \\ 0 & 0 \end{vmatrix} = 0,$$

$$(dy \wedge dz)(v_1, v_2) = \begin{vmatrix} 0 & 0 \\ 0 & 3 \end{vmatrix} = 0,$$

$$(dy \wedge dz)(w_1, w_2) = \begin{vmatrix} 3 & 0 \\ 0 & 3 \end{vmatrix} = 9.$$

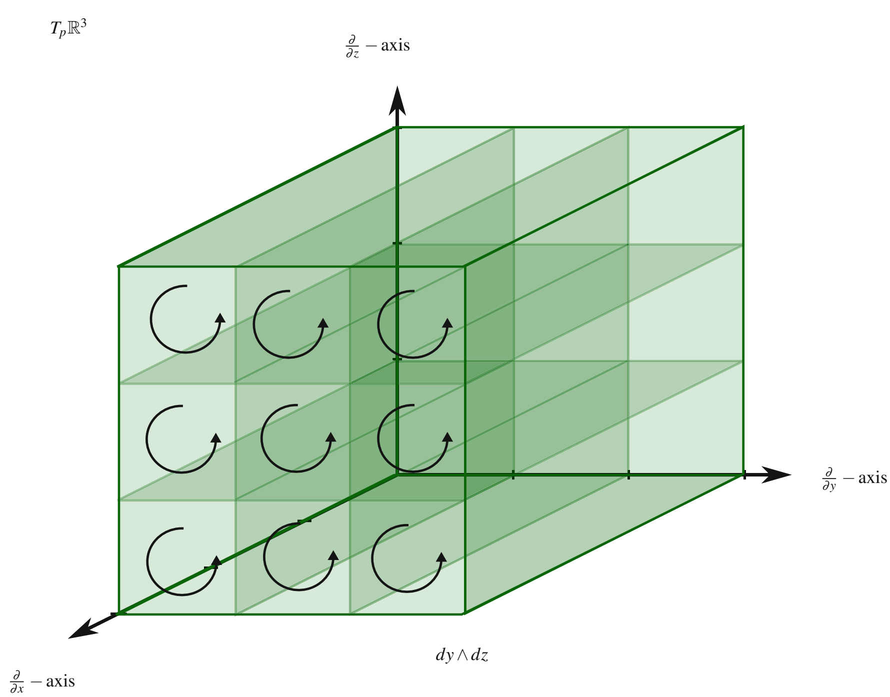


Fig. 5.20 The two-form $dy \wedge dz$ depicted in $T_p\mathbb{R}^3$ as a series of "tubes" made up of planes parallel to the ∂_y and ∂_z axes. The positive orientation is illustrated

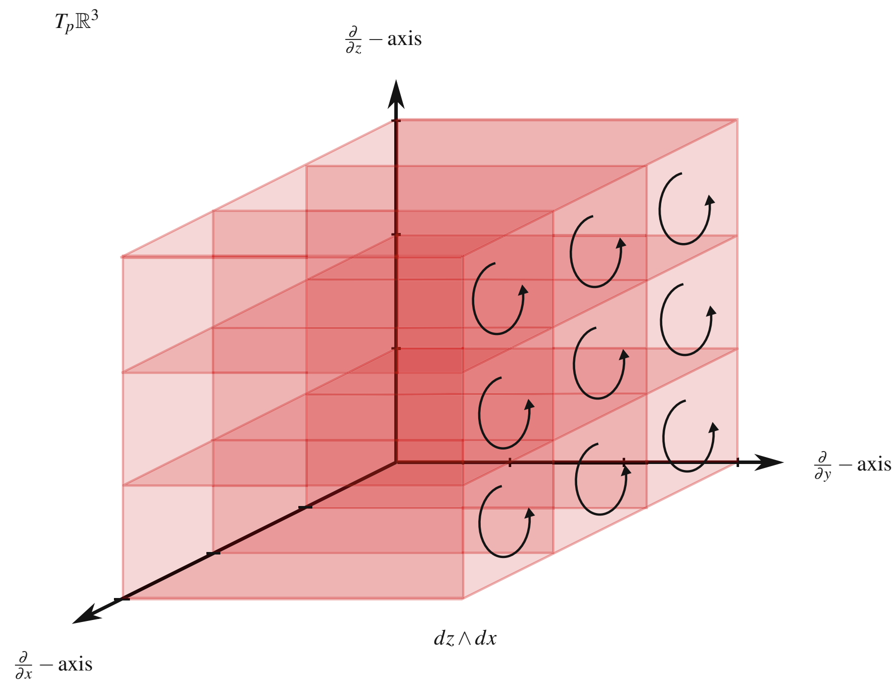


Fig. 5.21 The two-form $dz \wedge dx$ depicted in $T_p\mathbb{R}^3$ as a series of "tubes" made up of planes parallel to the ∂_z and ∂_x axes. The positive orientation is illustrated

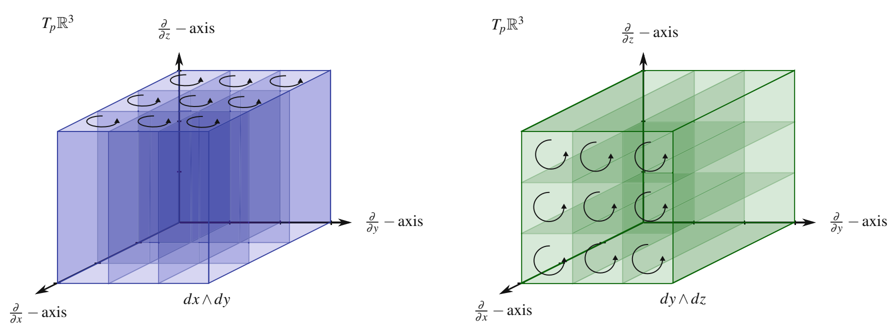


Fig. 5.22 To find the graphical representation of the two-form $dx \wedge dy + dy \wedge dz$ we have to somehow "add" the tubes of $dx \wedge dy$ (left) with those of $dy \wedge dz$ (right). The "addition" of these tubes is shown in Fig. 5.23

By imagining the vectors u_1 and u_2 imposed on the picture of $dy \wedge dz$. We can see that the parallelepiped spanned by these two vectors is parallel with the tubes and so does not cut through any of them. Similarly, the parallelepiped spanned by v_1 and v_2 does not cut through any tubes, but the parallelepiped spanned by w_1 and w_2 cuts through nine tubes.

Finally, the graphical representation for $dz \wedge dx$ consists of both the dz and dx planes resulting in tubes that are parallel with the ∂_y-axis. See Fig. 5.21. A similar analysis with the above vectors gives the expected results.

When finding a graphical representation for the two-forms $adx \wedge dy$, $bdy \wedge dz$, and $cdz \wedge dx$, where $a, b, c \in \mathbb{R}$, we have exactly the same ambiguities that we had in the-two dimensional case.

Question 5.12 Find at least three different graphical representations for the differential two-form $4dx \wedge dy$.

Now we consider general differential two-forms $adx \wedge dy + bdy \wedge dz + cdz \wedge dx$, where $a, b, c \in \mathbb{R}$. To make matters simple we first consider the two-form $dx \wedge dy + dy \wedge dz$. In essence we want to find a graphical representation that is the sum of these two pictures, see Fig. 5.22. The tubes of the differential two-form $dx \wedge dy$ are parallel to the ∂_z-axis and the tubes of the differential two-form $dy \wedge dz$ are parallel to the ∂_x-axis. Consider the unit vector in the ∂_x direction and the unit

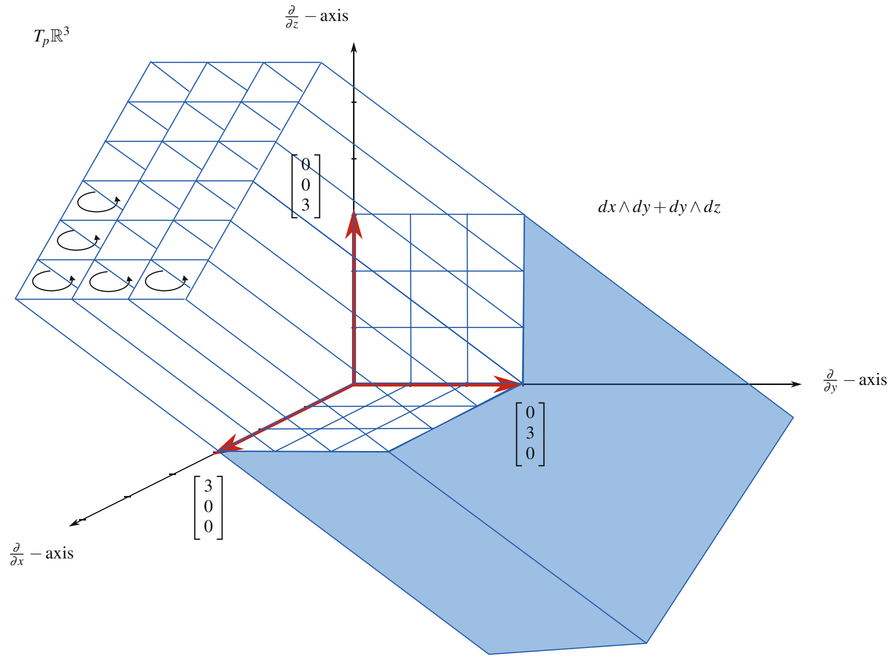


Fig. 5.23 The graphical representation of the two-form $dx \wedge dy + dy \wedge dz$ we get from "adding" the tubes of $dx \wedge dy$ and $dy \wedge dz$ shown in Fig. 5.22

vector in the ∂_z direction. When we add these two unit vectors we get

$$\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}.$$

The tubes of $dx \wedge dy + dy \wedge dz$ are parallel to the line determined by this vector, as illustrated in Fig. 5.23.

Question 5.13 Let

$$u = \begin{bmatrix} 3 \\ 0 \\ 0 \end{bmatrix}, v = \begin{bmatrix} 0 \\ 3 \\ 0 \end{bmatrix}, w = \begin{bmatrix} 0 \\ 0 \\ 3 \end{bmatrix}.$$

With the aid of the Fig. 5.23 show that the answers given graphically match those given computationally for the following,

(a) $(dx \wedge dy + dy \wedge dz)(u, v)$,
(b) $(dx \wedge dy + dy \wedge dz)(v, w)$,
(c) $(dx \wedge dy + dy \wedge dz)(w, u)$.

Intuitively, the picture associated with $dx \wedge dy + dy \wedge dz$ appears to make sense. There is another way to see what is happening. Recall that the planes used to depict the one-form dx are perpendicular to the x (or ∂_x) axis and the planes used to depict dy are perpendicular to the y (or ∂_y) axis. We then found the image for $dx \wedge dy$ by simply depicting both of these sets of planes simultaneously. Our goal is to manipulate the two-form $dx \wedge dy + dy \wedge dz$ into a form which allows us to do something similar,

$$\begin{aligned} dx \wedge dy + dy \wedge dz &= -dy \wedge dx + dy \wedge dz \\ &= dy \wedge (-dx + dz) \\ &= dy \wedge d(-x + z). \end{aligned}$$

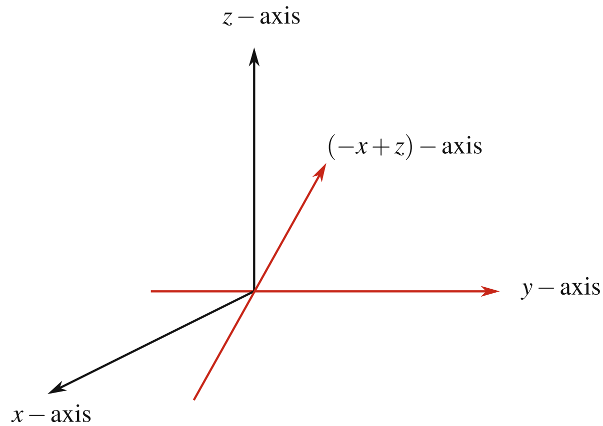


Fig. 5.24 The planes perpendicular to the y-axis and the $(-x+z)$-axis give the two-form $dx \wedge dy + dy \wedge dz$ depicted in Fig. 5.23

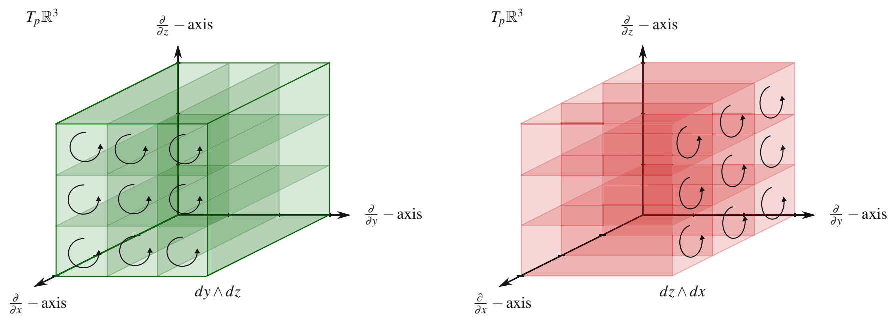


Fig. 5.25 To find the graphical representation of the two-form $dy \wedge dz + dz \wedge dx$ we need to "add" the tubes of $dy \wedge dz$ (left) and $dz \wedge dx$ (right)

So $dx \wedge dy + dy \wedge dz$ can be represented by using the planes perpendicular to the y axis and the $(-x+z)$ "axis." This last bit requires a little explanation. Consider the unit vectors in the x and z directions and add negative the first to the second to give us

$$-\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} = \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix}.$$

The vector $\begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix}$ determines the axis that is used. The planes perpendicular to this axis is the second set of planes we use. In summary, for $dx \wedge dy + dy \wedge dz$ planes perpendicular to the two red axes drawn in Fig. 5.24 are used.

Now we want to find the picture for $dy \wedge dz + dz \wedge dx$, see Fig. 5.25. By inspection of the pictures for $dy \wedge dy$ and $dz \wedge dx$ we can see that the tubes of $dy \wedge dz + dz \wedge dx$ will we parallel with the vector $\begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}$. Or we could rewrite $dy \wedge dz + dz \wedge dx$ as before,

$$\begin{aligned} dy \wedge dz + dz \wedge dx &= -dz \wedge dy + dz \wedge dx \\ &= dz \wedge (-dy + dx) \\ &= dz \wedge d(-y + x), \end{aligned}$$

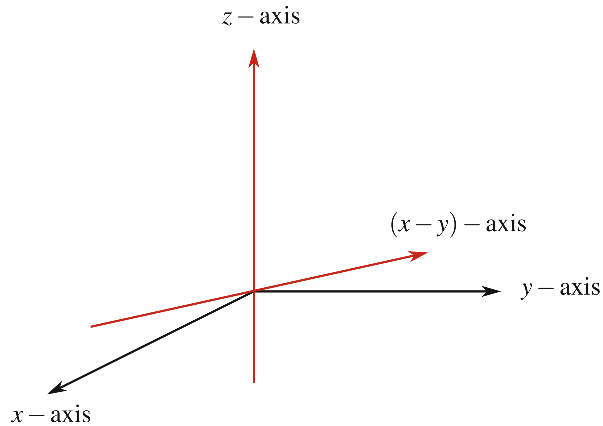


Fig. 5.26 The tubes of $dy \wedge dz + dz \wedge dx$ are made up of planes perpendicular to the x-axis and the $(x-y)$-axis

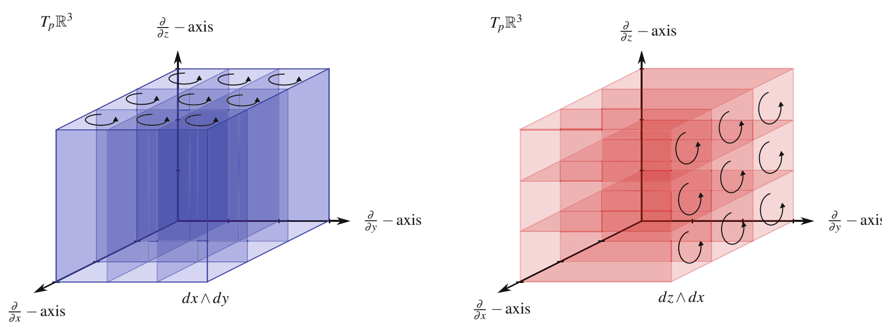


Fig. 5.27 The two-form $dx \wedge dy + dz \wedge dx$ is obtained by adding the tubes of $dx \wedge dy$ (left) with those of $dz \wedge dx$ (right)

and use planes perpendicular to the z axis and the $(-y+x)$ "axis," which is in the direction of the vector $\begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix}$. These two axis are shown in red in Fig. 5.26.

Now we want to find the picture for $dx \wedge dy + dz \wedge dx$. See Fig. 5.27. By inspection we can see that the tubes of $dx \wedge dy + dz \wedge dx$ are parallel to the vector $\begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}$, or we could rewrite the two-form

$$\begin{aligned} dx \wedge dy + dz \wedge dx &= dx \wedge dy - dx \wedge dz \\ &= dx \wedge (dy - dz) \\ &= dx \wedge d(y - z) \end{aligned}$$

to get the two axes used to find the perpendicular planes. The two axes are shown in red in Fig. 5.28.

Finally we would like to find the picture for $dx \wedge dy + dy \wedge dz + dz \wedge dx$. See Fig. 5.29. We can rewrite the two-form $dx \wedge dy + dy \wedge dz + dz \wedge dx$ as follows,

$$\begin{aligned} dx \wedge dy + dy \wedge dz + dz \wedge dx &= dy \wedge d(-x + z) + dz \wedge dx \\ &= dy \wedge d(-x + z) - dz \wedge -dx + dz \wedge dz \\ &= dy \wedge d(-x + z) + dz \wedge (-dx + dz) \\ &= dy \wedge d(-x + z) + dz \wedge d(-x + z) \end{aligned}$$

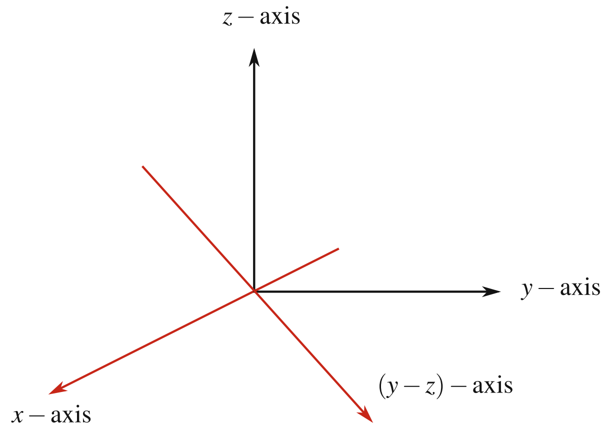


Fig. 5.28 The two-form $dx \wedge dy + dz \wedge dx$ is made up of planes perpendicular to the x-axis and the $(y-z)$-axis

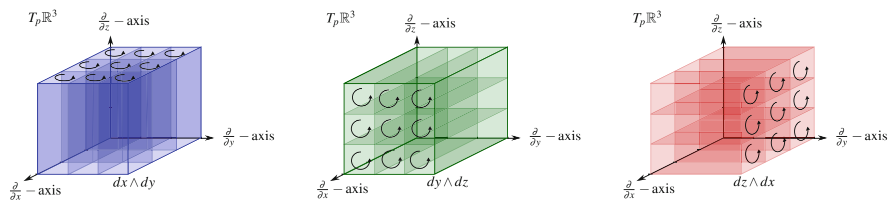


Fig. 5.29 The two-form $dx \wedge dy + dy \wedge dz + dz \wedge dx$ is obtained by "adding" the tubes for $dx + dy$ (left), $dy \wedge dz$ (middle), and $dz \wedge dx$ (right)

$$
\begin{aligned}
&= (dy + dz) \wedge d(-x + z) \\
&= d(y + z) \wedge d(-x + z),
\end{aligned}
$$

which can be used to sketch the two-form. Finally, when moving to two-forms with coefficients all the usual problems occur.

Question 5.14 Sketch the two-form $dy \wedge dz + dz \wedge dx$.

Question 5.15 Sketch the two-form $dx \wedge dy + dz \wedge dx$.

Question 5.16 Sketch the two-form $dx \wedge dy + dy \wedge dz + dz \wedge dx$.

Question 5.17 Sketch a pictures for each of the following two-forms

(a) $2dx \wedge dy + dy \wedge dz$,
(b) $dx \wedge dy + 3dy \wedge dz$,
(c) $2dx \wedge dy + 4dy \wedge dz$.

Question 5.18 Sketch a pictures for each of the following two-forms

(a) $3dy \wedge dz + 2dz \wedge dx$,
(b) $dx \wedge dy + 3dz \wedge dx$,
(c) $2dx \wedge dy + 4dy \wedge dz + 3dz \wedge dx$.

Question 5.19 Sketch the following two-forms

(a) $-dx \wedge dy + dy \wedge dz$,
(b) $dx \wedge dy - dy \wedge dz$,
(c) $-dx \wedge dy - dy \wedge dz$.

What affect do the negative signs have? Use some well chosen vectors to help you figure it out.

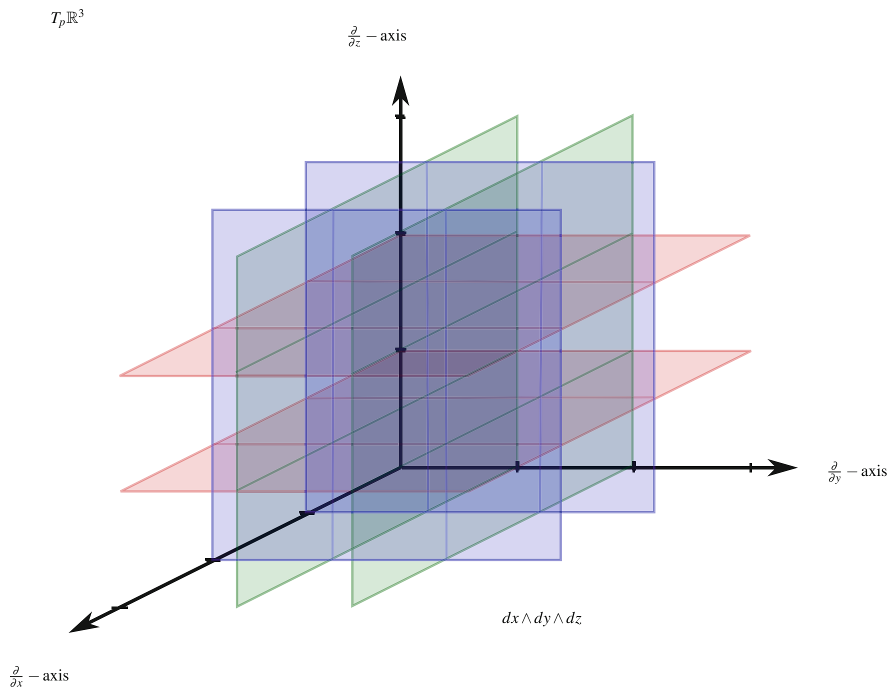


Fig. 5.30 The three-form $dx \wedge dy \wedge dz$ is made up of "cubes" whose sides are the planes perpendicular to the x-axis, the y-axis, and the z-axis

5.4 Three-Forms in $\mathbb{R}^3$

This is probably the shortest section in the entire book. And for a reason. Drawing three-forms in $T_p\mathbb{R}^3$ is in fact exceedingly simple. The space of three-forms on $\mathbb{R}$, denoted $\bigwedge^3(\mathbb{R})$, is actually of dimension one, which means that every three-form on $\mathbb{R}^3$ is of the form $c dx \wedge dy \wedge dz$ for some $c \in \mathbb{R}$. The one-form dx was given by the planes $x = n$, the one-form dy was given by the planes $y = n$, and the one-form dz was given by the planes $z = n$, where we had $n = 0, \pm 1, \pm 2, \ldots$. The three-form is simply all of these planes being drawn simultaneously, thereby filling space with unit cubes. See Fig. 5.30. For three-forms $a dx \wedge dy \wedge dz$, where $a \in \mathbb{R}$, the usual ambiguity arises.

Question 5.20 Based on Fig. 5.30 sketch three different pictures for the three-form $2dx \wedge dy \wedge dz$.

Question 5.21 Let

$$u = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, v = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, w = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}.$$

Evaluate $dx \wedge dy \wedge dz$ on different permutations of u, v, w and use your answers to determine what kind of orientation is $dx \wedge dy \wedge dz$ has. (Note, somehow all three directions have to be taken into account.)

Question 5.22 How might you depict the positive orientation for $dx \wedge dy \wedge dz$ in Fig. 5.30? What about the negative orientation?

5.5 Pictures of Forms on Manifolds

What we have done so far is to attempt to draw graphical representations of one-forms and two-forms in $\mathbb{R}^2$ and one-forms, two-forms, and three-forms in $\mathbb{R}^3$. In each case we looked at some form α_p at the point p and found a representation of α_p in the tangent space at the point p. For a one-form on $\mathbb{R}^2$ our picture was of lines in $T_p\mathbb{R}^2$ (Figs. 5.1–5.7), while for a two-form on $\mathbb{R}^2$ our picture was of boxes in $T_p\mathbb{R}^2$ (Figs. 5.8–5.10). For a one-form in $\mathbb{R}^3$ our picture was of planes in $T_p\mathbb{R}^3$ (Figs. 5.11–5.18), for a two-form in $\mathbb{R}^3$ our picture was of tubes in $T_p\mathbb{R}^3$ (Figs. 5.19–5.29), and for a three-form in $\mathbb{R}^3$ our pictures was of cubes in $T_p\mathbb{R}^3$ (Fig. 5.30).

Sometimes attempts are made to utilize these graphical representations of one-forms and two-forms on $\mathbb{R}^2$ and one-forms, two-forms, and three-forms on $\mathbb{R}^3$ to "draw" forms on the whole manifold, at least for manifolds that are two or three dimensional. There are two ways that this is generally done. The first method at least makes sense mathematically, though it can still end up being incomplete in some sense and can potentially misrepresent the differential form. The second method, unfortunately, has some serious mathematical shortcomings, in addition to potentially misrepresenting the differential form. But when these methods do work they can be an extremely helpful and illuminating way to think about differential forms on two and three dimensional manifolds. It is best to keep in the back of your mind that the pictures these methods give are merely cartoons at best.

First, we will present the first method of visualizing forms on manifolds. Essentially what is done here is that a lattice of points $p_{ij} = (x_i, y_j)$ on the manifold is chosen and the line-stacks that represent the one-form in $T_{p_{ij}}M$ are drawn on the manifold at each of these lattice points. Figure 5.31 shows how this is done for a series of simple one-forms on $\mathbb{R}^2$. The arrow points indicate positive orientation. Of course, we can choose lattices of points on the manifold with differing densities. Figure 5.32 shows the same one-form $xdx + ydy$ on $\mathbb{R}^2$ using two different lattices of points. A two-form would be similar, only with a small grid of boxes drawn at each point. The drawback of this method is the spacing of the lattice of points; it is entirely possible that important details of the differential form can be missed if one chooses an inappropriate lattice. Be that as it may, one generally still obtains a reasonable cartoon that can at least help one imagine the differential form.

For three dimensional manifolds the pictures are analogous. First a three-dimensional lattice of points $p_{ijk} = (x_i, y_j, z_k)$ is chosen on the manifold. A one-form would then be represented as small stacks of sheets at each point p_{ijk}, a two-form would be represented with small bundles of tubes at each point p_{ijk}, and a three-form would be represented with small sets of cubes at each point p_{ijk}.

The second method does something similar, but tries to take it one step further. Consider Fig. 5.32 again. In the second image the lattice of points was chosen close enough so that, particularly in the middle, they start to merge into each other and form concentric circles. In the second method the little pictures at the lattice points are connected up to each other to provide one continuous image on the manifold. When this is actually possible the images are very appealing and simple to think about. First we will look at how method two works when it is appropriate and then we will briefly discuss how method two fails in general.

As an example, a generic one-form α on $\mathbb{R}^3$ is shown in Fig. 5.33. Here the line-stacks at the lattice points (not shown) are connected, as well as is possible, to generate curves that cover the manifold. The value $\alpha(v)$ is (approximately) given by the number of curves the vector v pierces. The case of two-forms on the manifold $\mathbb{R}^2$ is handled identically, only instead of small stacks of lines at each lattice point one would have small patches of boxes. Again, it may be possible to connect up the boxes to give a rough image of the two-form on the manifold. Consider Fig. 5.34. Here a two-form on manifold $\mathbb{R}^2$ is depicted on the manifold $\mathbb{R}^2$. In order to find the value of the two-form α at any given point p two vectors at that point need to be given, v_1 and v_2. The value $\alpha_p(v_1, v_2)$ is then given by the number of boxes that are inside the parallelepiped spanned by v_1 and v_2.

Attempts to picture one-forms, two-forms, and three-forms on the manifold $\mathbb{R}^3$ are very similar. Again, a lattice of points $p_{ijk} = (x_i, y_j, z_k)$ is chosen and at each of these points the picture of the form in $T_{p_{ijk}}\mathbb{R}^3$ is found. For a one-form α at each point p_{ijk} one would have a stack of two-dimensional sheets, similar to Fig. 5.17. Small copies of each of these pictures would then be superimposed on the manifold $\mathbb{R}^3$ at each lattice point. One would then try to connect up these sheets, as best as one could, to arrive at a picture similar to that of Fig. 5.35. How closely the sheets are packed (in other words, how far apart the sheets are from each other) is determined by the particular one-form. To find the (approximate) value of $\alpha(v)$ one would then count the number of sheets the vector v pierces. So, the sheets of the one-form 10α are ten times as close to each other as the sheets of the one-form α.

For a two-form α at each point p_{ijk} one would have a bundle of "tubes" similar to Fig. 5.23. At each lattice point the bundle of tubes would have varying densities and be in different directions. One would then try to connect up these sheets, as best as one could, to arrive at a picture similar to that of Fig. 5.36, where tubes of varying sizes twist through $\mathbb{R}^3$. The density at which the tubes are packed depends on the exact nature of the two-form α. In order to (approximately) find $\alpha(v_p, w_p)$ one finds the parallelepiped spanned by v_p and w_p and counts the number of tubes that go through this parallelepiped.

For a three-form α at each point p_{ijk} one would have a little set of cubes, similar to Fig. 5.30. One could imagine Rubik's cubes of varying sizes and orientations at each lattice point. Then these cubes are all connected up as well as possible to fill $\mathbb{R}^3$ with little cubes. We will not attempted to show that here. Finding the (approximate) value of $\alpha(u_p, v_p, w_p)$ amounts to counting the number of cubes that are inside the parallelepiped spanned by u_p, v_p, w_p.

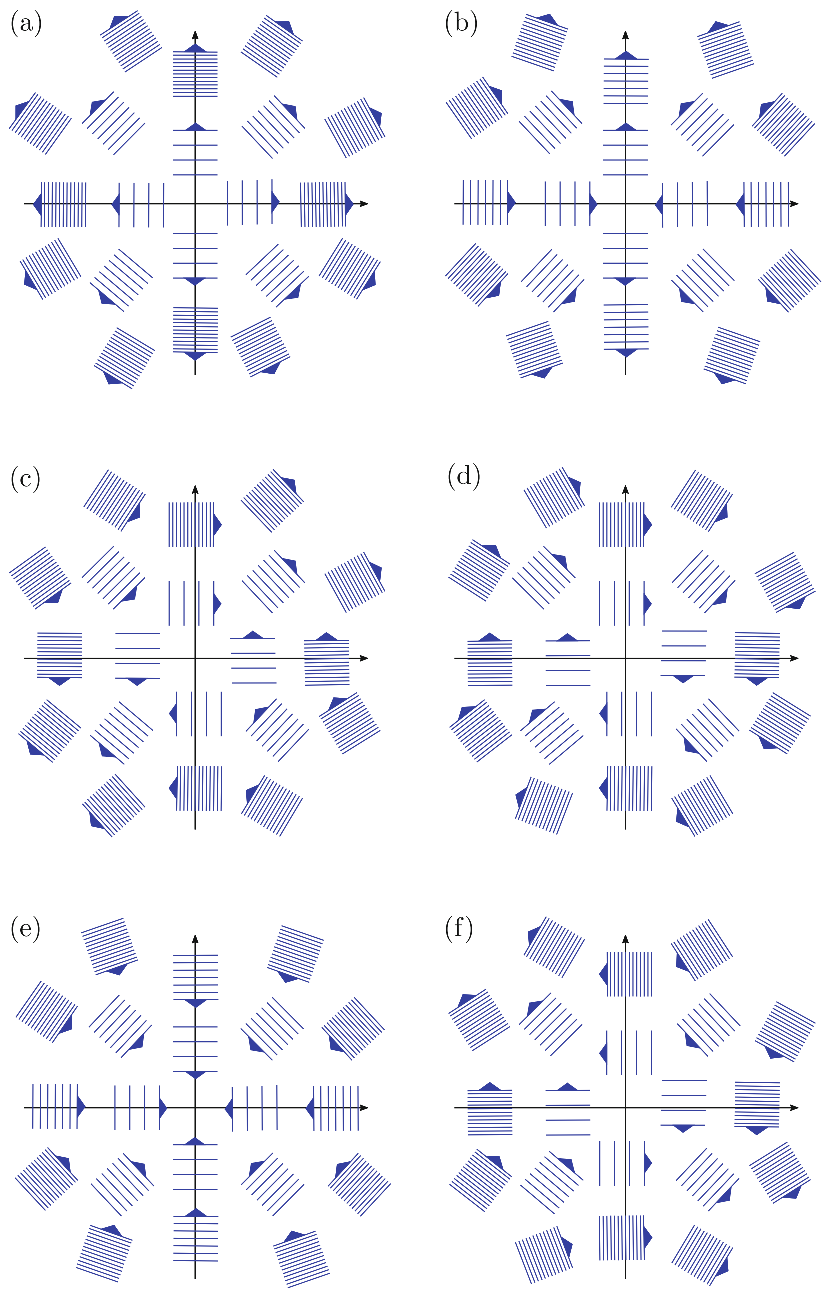


Fig. 5.31 One-forms on the manifold $\mathbb{R}^2$ represented as a grid of line-stacks on the manifold $\mathbb{R}^2$. The arrow points indicate positive direction (Images generated with Vector Field Analyzer II, Version 2, Kawski, 2001) . (**a**) $xdx + ydy$. (**b**) $-xdx + ydy$. (**c**) $ydx + xdy$. (**d**) $ydx - xdy$. (**e**) $-xdx - ydy$. (**f**) $-ydx - xdy$

There is something very appealing about this way of looking at differential forms, and in certain instances it helps provide an additional way of thinking about differential forms on two-dimensional and three-dimensional manifolds. Unfortunately, in general method two fails. In fact, one does not have to look very far to find examples of forms where it is actually impossible to visualize them using method two. One of the real issues is the "connecting" bit. Mathematically this is a complete slight-of-hand with no reasonable justification. In fact, each of the line-stacks for a one-form are in a different tangent space so "connecting" the line-stacks simply doesn't make sense. And even when we draw the pictures on the manifold as we have done, there is no reason to suppose that the line-stacks at points close to each other would be oriented in a way that would allow them to be connected, as was the case in Fig. 5.32. For a general one-form they need not be. And of course this applies to the plane-stacks and tube-bundles for three-dimensional manifolds as well.

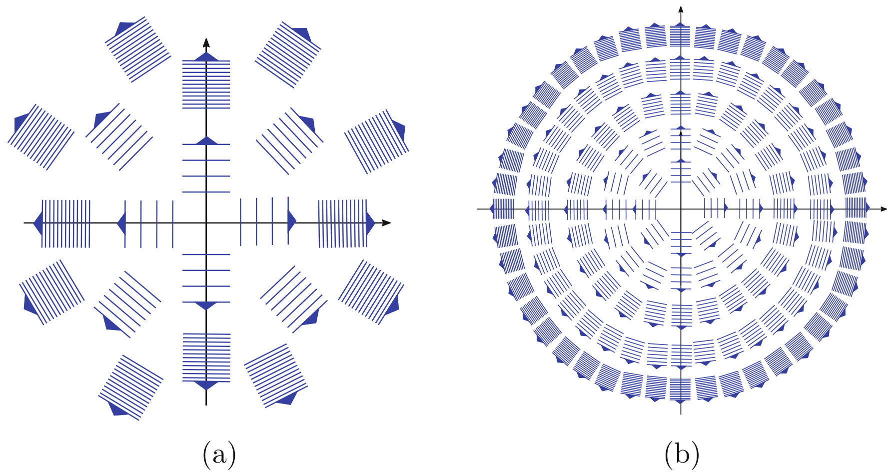


Fig. 5.32 The one-form $xdx + ydy$ on the manifold $\mathbb{R}^2$ represented as a grid of line-stacks using two different arrays of points. The arrow points indicate positive direction. Notice how in the image on the right the line-stacks in the middle start to look like they are merging into concentric circles (Images generated with Vector Field Analyzer II, Version 2, Kawski, 2001)

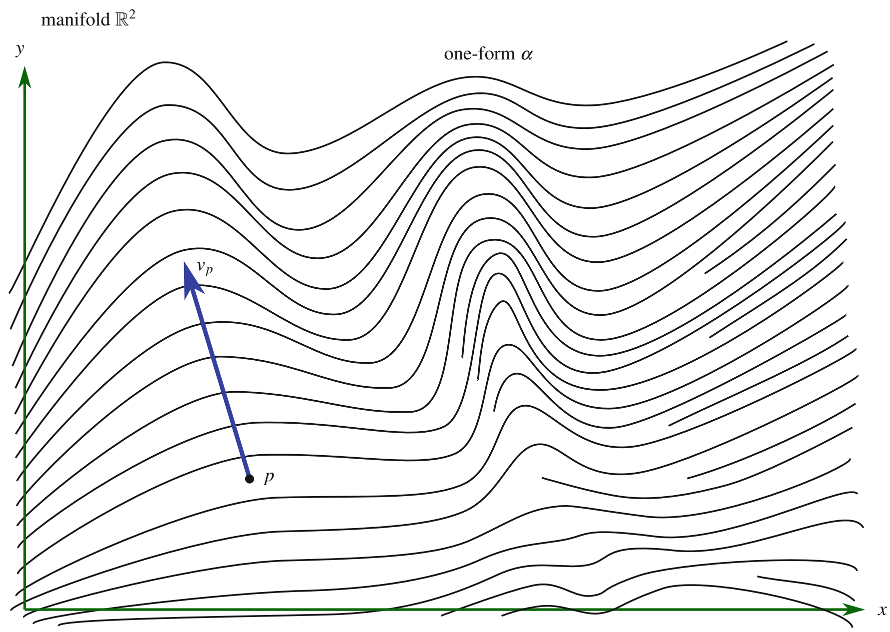


Fig. 5.33 A one-form α is shown on the manifold $\mathbb{R}^2$ with all the "line stacks" connected up smoothly as curves. The one-form applied to a vector is (approximately) equal to the number of curves "pierced" by the vector. Here, $\alpha(v_p) \approx 6$

A necessary, but not sufficient, requirement is that the distribution given by the kernel of the differential form be integrable. This would give what is called a foliation of the manifold which would allow the line-stacks, plane-stacks, or tube-bundles to line up. Exploring and explaining this is beyond the scope of this book. It is simply sufficient to realize that even a one-form on $\mathbb{R}^3$ as simple as $xdy + dz$ can not be visualized using method two.

However, this way of visualizing differential forms does appear in physics sometimes. Physicists are not so concerned about a completely general way of visualizing forms, they are interested in a way of visualizing forms that works for the physical problems and situations they are dealing with. Thus this method does have some utility in certain situations. When it is appropriate it does in fact provides a very nice way of thinking about forms. In fact, it gives a nice picture to go along with Stokes' theorem that is very appealing and will be presented in Sect. 11.6.

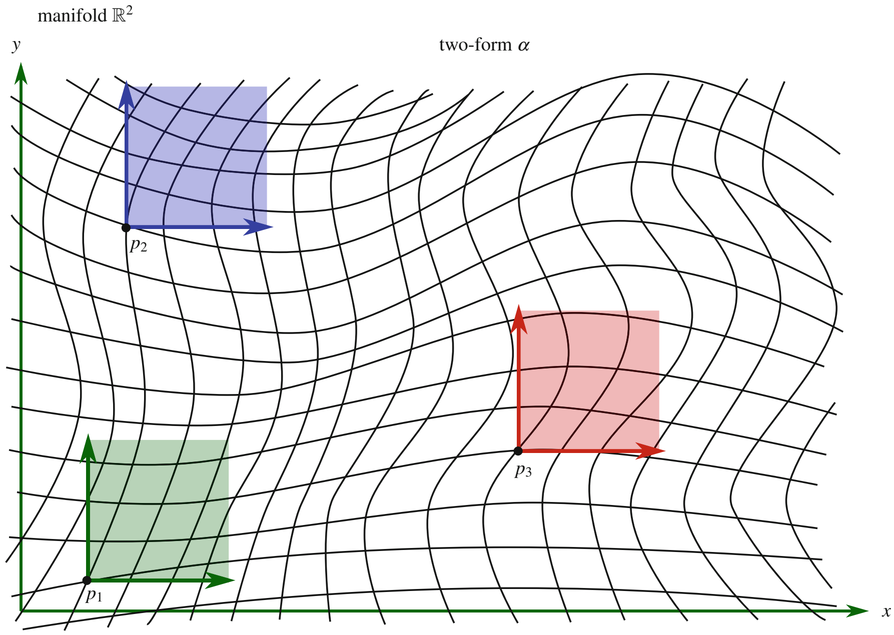


Fig. 5.34 A two-form α is shown on the manifold $\mathbb{R}^2$ with all boxes connected up smoothly. The unit vectors e_1 and e_2 are depicted at points p_1 (green), p_2 (blue), and p_3 (red). One (approximately) finds $\alpha_{p_i}(e_1, e_2)$ by counting the number of boxes that fall within the parallelepiped spanned by e_1 and e_2 at each of these points. Thus $\alpha_{p_1}(e_1, e_2) \approx 12$, $\alpha_{p_2}(e_1, e_2) \approx 16$, and $\alpha_{p_3}(e_1, e_2) \approx 9$. It is clear that while this is a nice picture the answers it gives are only very approximate

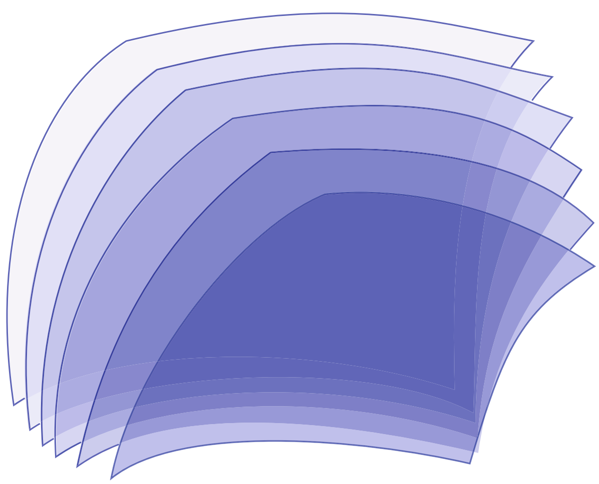


Fig. 5.35 A generic one-form in $\mathbb{R}^3$ is comprised of sheets filling $\mathbb{R}^3$

5.6 A Visual Introduction to the Hodge Star Operator

Introducing the Hodge star operator in a mathematically rigorous way at this point is a little tricky since much of the necessary mathematics has not yet been introduced, yet the geometric pictures of the Hodge star operator that go along with this chapter are really quite nice and easy to understand, so now is a natural time to do it. Thus we will try to strike a balance, providing some of the mathematics behind the pictures but not enough to overwhelm you with theoretical details. For the moment we will stick to the three dimensional case and give plausibility arguments instead of rigorous proofs.

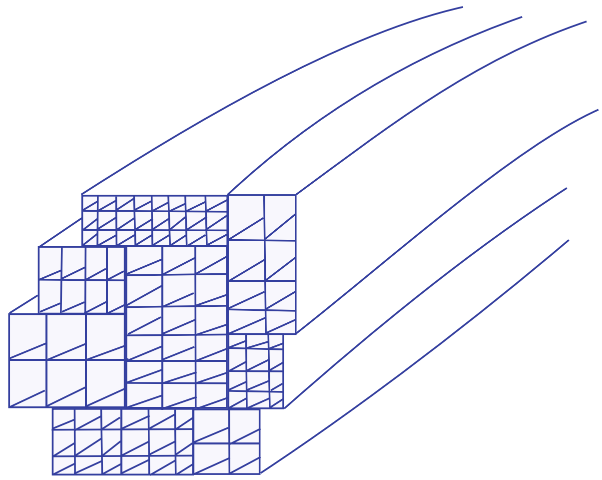


Fig. 5.36 A generic two-form in $\mathbb{R}^3$ is comprised of tubes filling $\mathbb{R}^3$

First we will consider the following vector spaces

$$
\begin{aligned}
\bigwedge\nolimits_p^0(\mathbb{R}^3) &= \text{span}\{1\}, \\
\bigwedge\nolimits_p^1(\mathbb{R}^3) &= \text{span}\{dx, dy, dz\} = T_p^*\mathbb{R}^3, \\
\bigwedge\nolimits_p^2(\mathbb{R}^3) &= \text{span}\{dx \wedge dy, dy \wedge dz, dz \wedge dx\}, \\
\bigwedge\nolimits_p^3(\mathbb{R}^3) &= \text{span}\{dx \wedge dy \wedge dz\},
\end{aligned}
$$

which have dimension one, three, three, and one respectively. Each of these spaces has what is called an *inner product* defined on them. Without getting distracted by a rigorous definition at this point we will simply say that in essence an inner product on a vector space associates with each pair of vectors a real number. With that said, what we will do is explain the inner product on $\bigwedge_p^1(\mathbb{R}^3)$ and $\bigwedge_p^2(\mathbb{R}^3)$, the two spaces we are most interested in. We begin with the space $\bigwedge_p^1(\mathbb{R}^3)$. The inner product on $\bigwedge_p^1(\mathbb{R}^3)$ is given by the matrix

$$
\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}.
$$

But how does this work? Suppose we have two one-forms elements $\alpha, \beta \in \bigwedge_p^1(\mathbb{R}^3)$. We can write these elements in terms of the basis elements as co-vectors

$$
\begin{aligned}
\alpha &= a dx + b dy + c dz = [a, b, c], \\
\beta &= r dx + s dy + t dz = [r, s, t].
\end{aligned}
$$

Then the inner product of α and β, which is generally denoted by $\langle \alpha, \beta \rangle$, is given by

$$
\begin{aligned}
\langle \alpha, \beta \rangle &= \langle [a, b, c], [r, s, t] \rangle \\
&\equiv [a, b, c] \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} [r, s, t]^T
\end{aligned}
$$

$$= [a, b, c] \begin{bmatrix} r \\ s \\ t \end{bmatrix}$$

$$= ar + bs + ct.$$

We take a moment to note that using angled brackets $\langle \cdot, \cdot \rangle$ to denote inner products is standard notation, and is quite a different thing from using angled brackets $\langle \cdot, \cdot \rangle$ to denote the canonical pairing between differential forms and vectors that we used earlier. You must pay careful attention to what is inside the brackets to know what they represent. Also, notice that we are implicitly assuming our basis elements have an order, dx first, dy second, and dz third. Only by knowing that order did we know how to write the matrix representing the inner product.

Similarly, we will define the inner product on the $\bigwedge_p^2(\mathbb{R}^3)$ by the same matrix

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

which works in exactly the same way. Thus if we have

$$\eta = adx \wedge dy + bdy \wedge dz + cdz \wedge dx = [a, b, c],$$

$$\xi = rdx \wedge dy + sdy \wedge dz + tdz \wedge dx = [r, s, t]$$

the inner product of η and ξ, which is denoted by $\langle \alpha, \beta \rangle$, is given by

$$\begin{aligned} \langle \eta, \xi \rangle &= \langle [a, b, c], [r, s, t] \rangle \\ &\equiv [a, b, c] \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} [r, s, t]^T \\ &= [a, b, c] \begin{bmatrix} r \\ s \\ t \end{bmatrix} \\ &= ar + bs + ct. \end{aligned}$$

Again, we implicitly assume that our basis had an order so we would know how to write the matrix.

Now we have the pieces necessary pieces to actually define the **Hodge star operator**, which is also called the Hodge star dual operator. The Hodge star operator is an isomorphism (or mapping) $* : \bigwedge_p^k(\mathbb{R}^n) \rightarrow \bigwedge_p^{n-k}(\mathbb{R}^n)$ that takes k-forms to $(n-k)$-forms. So, how is this isomorphism defined? For each k-form α there is a unique $(n-k)$-form $*\alpha$ such that for any $(n-k)$-form we have

$$\boxed{\text{Hodge Star Operator} \qquad \alpha \wedge \beta = \langle *\alpha, \beta \rangle \sigma \qquad \text{for all } \beta}$$

where the $\langle \cdot, \cdot \rangle$ represents the inner product on $\bigwedge_p^{n-k}(\mathbb{R}^n)$ and σ is the n dimensional volume form. So, $*\alpha$ is defined in terms of this fairly complicated relationship. We would like to unpack this relationship and see what it is trying to say,

$$\underbrace{\underbrace{\alpha}_{k\text{-form}} \wedge \underbrace{\beta}_{(n-k)\text{-form}}}_{\text{an } n\text{-form}} = \underbrace{\langle \underbrace{*\alpha}_{(n-k)\text{-form}}, \underbrace{\beta}_{(n-k)\text{-form}} \rangle}_{\text{a number}} \underbrace{\sigma}_{\substack{\text{volume form} \\ \text{(an } n\text{-form)}}}.$$

So we end up with an n-form on both the left and the right hand side of the equality. Recall, the volume form σ is known, the α is given, and the β can be any $(n-k)$-form at all. We then use this relationship to find out which $(n-k)$-form $*\alpha$ is. Since β can be any $(n-k)$-form at all this means that no matter which $(n-k)$-form is chosen to be β we will always find exactly the same $*\alpha$. In a way that is rather amazing. But the mapping $*$ is unique, a fact which we will not try to prove here, so there is only one $*\alpha$ no matter which $(n-k)$-form is chosen to be β.

Now, let's actually put this relationship to work in our three dimensional case. We will first look at the mapping $* : \bigwedge_p^1(\mathbb{R}^3) \to \bigwedge_p^2(\mathbb{R}^3)$. First we will tackle finding $*dx$. Since we know that $*dx$ is a two-form we know that it must have the form

$$*dx = a\ dx \wedge dy + b\ dy \wedge dz + c\ dz \wedge dx = [a, b, c]$$

for some numbers a, b, c. Our goal is to find out what a, b, and c are using this relationship

$$dx \wedge \beta = \langle *dx, \beta \rangle dx \wedge dy \wedge dz.$$

For the moment we will consider a generic two-form β given by

$$\beta = r\ dx \wedge dy + s\ dy \wedge dz + t\ dz \wedge dx = [r, s, t].$$

So our defining relationship becomes

$$\begin{aligned}
&dx \wedge \big(r\ dz \wedge dy + s\ dy \wedge dz + t\ dz \wedge dx\big) \\
&= \big\langle *dx, r\ dz \wedge dy + s\ dy \wedge dz + t\ dz \wedge dx \big\rangle dx \wedge dy \wedge dz \\
&= \big\langle [a, b, c], [r, s, t] \big\rangle dx \wedge dy \wedge dz \\
&= (ar + bs + ct) dx \wedge dy \wedge dz.
\end{aligned}$$

On the left hand side we have

$$\begin{aligned}
&dx \wedge \big(r\ dx \wedge dy + s\ dy \wedge dz + t\ dz \wedge dx\big) \\
&= r\ \underbrace{dx \wedge dx \wedge dy}_{=0} + s\ dx \wedge dy \wedge dz + t\ \underbrace{dx \wedge dz \wedge dx}_{=0} \\
&= s\ dx \wedge dy \wedge dz.
\end{aligned}$$

Combining we have

$$s\ dx \wedge dy \wedge dz = (ar + bs + ct) dx \wedge dy \wedge dz,$$

which gives us

$$s = ar + bs + ct.$$

Since our defining relationship is true no matter what β we had chosen, if we had chosen β such that $r = 1$, $s = 0$, and $t = 0$ this equality would give us that $a = 0$. If we had chosen β such that $r = 0$, $s = 1$, and $t = 0$ this equality would have given us $b = 1$. If we had chosen β such that $r = 0$, $s = 0$, and $t = 1$ this equality would have given us $c = 0$. Thus we have that

$$*dx = dy \wedge dz.$$

The relationship between dx and $*dx$ is shown pictorially in Fig. 5.37. The tubes of $*dx = dy \wedge dz$ are perpendicular to the planes of dx.

Question 5.23 Show that $dx \wedge \beta = \langle *dx, \beta \rangle dx \wedge dy \wedge dx$ for the following β,

(a) $\beta = 7dx \wedge dy - 3dy \wedge dz + 2dz \wedge dx$,
(b) $\beta = 20dx \wedge dy + 15dy \wedge dz - 10dz \wedge dx$,
(c) $\beta = -4dx \wedge dy - 6dy \wedge dz - 8dz \wedge dx$.

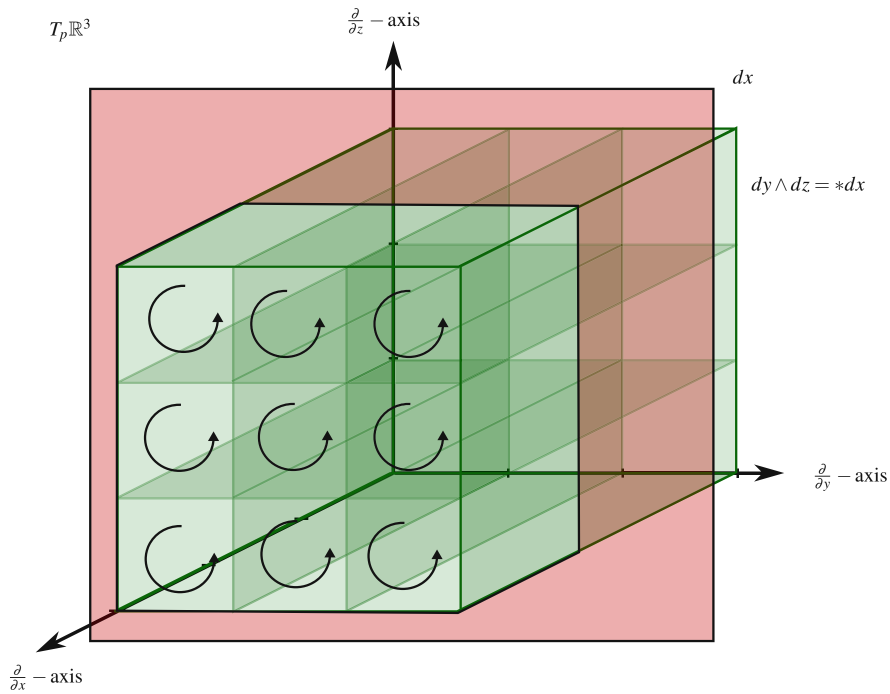


Fig. 5.37 The tubes of $*dx$ are perpendicular to the planes of dx. That is, $*dx = dy \wedge dz$

We will redo the calculation to find $*dy$. As before, since we know that $*dy$ is a two-form we know that it must have the form

$$*dy = a\ dx \wedge dy + b\ dy \wedge dz + c\ dz \wedge dx = [a, b, c]$$

for some numbers a, b, c and we wish to find out what a, b, and c are using this relationship

$$dy \wedge \beta = \langle *dy, \beta\rangle dx \wedge dy \wedge dz.$$

Again, we assume

$$\beta = r\ dx \wedge dy + s\ dy \wedge dz + t\ dz \wedge dx = [r, s, t]$$

resulting in

$$\begin{aligned}
&dy \wedge \big(r\ dz \wedge dy + s\ dy \wedge dz + t\ dz \wedge dx\big)\\
&= \big\langle *dy, r\ dz \wedge dy + s\ dy \wedge dz + t\ dz \wedge dx\big\rangle dx \wedge dy \wedge dz\\
&= (ar + bs + ct)dx \wedge dy \wedge dz.
\end{aligned}$$

The left hand side becomes

$$\begin{aligned}
&dy \wedge \big(r\ dx \wedge dy + s\ dy \wedge dz + t\ dz \wedge dx\big)\\
&= r\ \underbrace{dy \wedge dx \wedge dy}_{=0} + s\ \underbrace{dy \wedge dy \wedge dz}_{=0} + t\ dy \wedge dz \wedge dx\\
&= -t\ dy \wedge dx \wedge dz\\
&= t\ dx \wedge dy \wedge dz.
\end{aligned}$$

Combining we have

$$t\,dx \wedge dy \wedge dz = (ar + bs + ct)dx \wedge dy \wedge dz,$$

which gives us

$$t = ar + bs + ct.$$

Since our defining relationship is true no matter what β we chose, if we had chosen β such that $r = 1$, $s = 0$, and $t = 0$ we would have $a = 0$. If we had chosen it such that $r = 0$, $s = 1$, and $t = 0$ we would have $b = 0$. And if we had chosen it such that $r = 0$, $s = 0$, and $t = 1$ we would have $c = 1$. Thus we would have

$$*dy = dz \wedge dx.$$

Again, notice that the tubes of $*dy$ are perpendicular to the planes of dy. See Fig. 5.38.

Question 5.24 Show that $dy \wedge \beta = \langle *dy, \beta \rangle dx \wedge dy \wedge dx$ for the following β,

(a) $\beta = 7dx \wedge dy - 3dy \wedge dz + 2dz \wedge dx$,
(b) $\beta = 20dx \wedge dy + 15dy \wedge dz - 10dz \wedge dx$,
(c) $\beta = -4dx \wedge dy - 6dy \wedge dz - 8dz \wedge dx$.

Question 5.25 Find $*dz$ using the defining identity $dz \wedge \beta = \langle *dz, \beta \rangle dx \wedge dy \wedge dz$ for all $\beta \in \bigwedge_p^2(\mathbb{R}^3)$.

The image associated with $*dz$ shows the tubes of $*dz$ being perpendicular to the planes of dz, see Fig. 5.39.

Question 5.26 Show that $dz \wedge \beta = \langle *dz, \beta \rangle dx \wedge dy \wedge dx$ for the following β,

(a) $\beta = 7dx \wedge dy - 3dy \wedge dz + 2dz \wedge dx$,
(b) $\beta = 20dx \wedge dy + 15dy \wedge dz - 10dz \wedge dx$,
(c) $\beta = -4dx \wedge dy - 6dy \wedge dz - 8dz \wedge dx$.

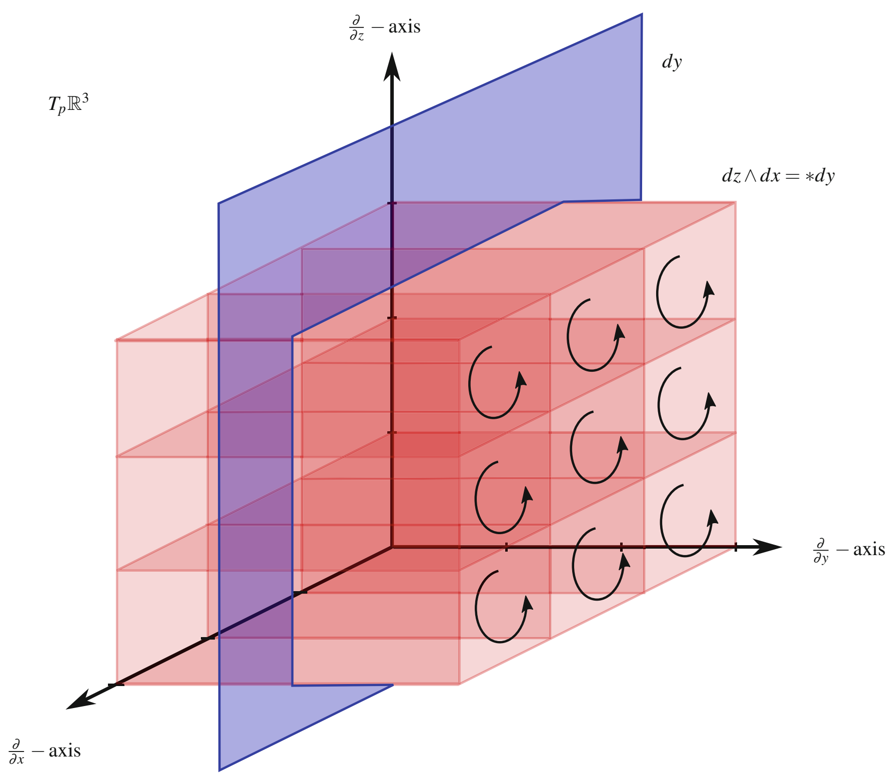


Fig. 5.38 The tubes of $*dy$ are perpendicular to the planes of dy. That is, $*dy = dz \wedge dx$

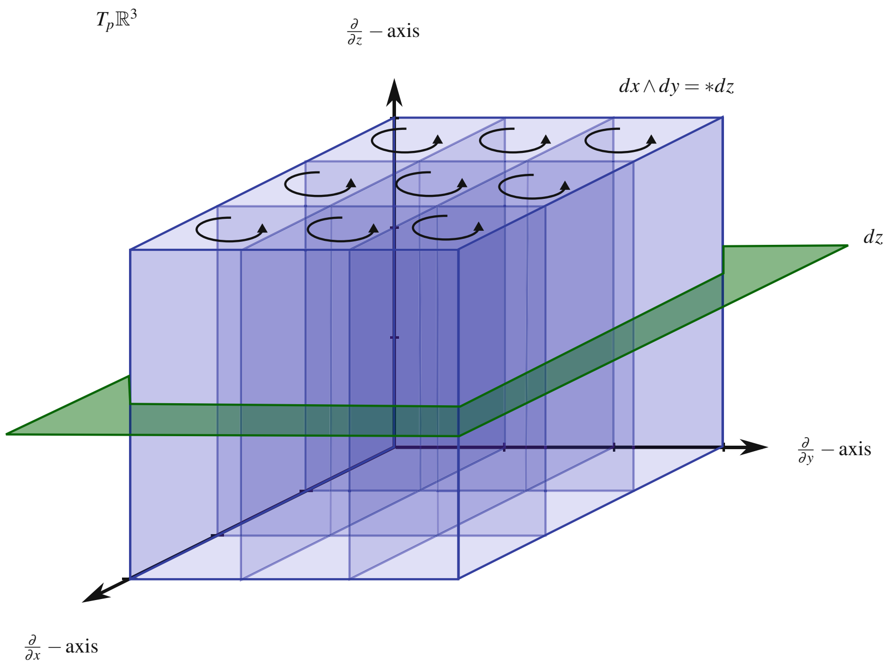


Fig. 5.39 The tubes of $*dz$ are perpendicular to the planes of dz. That is, $*dz = dx \wedge dy$

Question 5.27 Using the definition of the Hodge star operator find $*dx \wedge dy$, $*dy \wedge dz$, and $*dz \wedge dx$.

Question 5.28 You will sometimes see an alternative definition of the Hodge star operator. Since this is a somewhat easier definition to deal with we have left it for an exercise. Show that the alternative definition for the Hodge star operator

$$\alpha \wedge *\beta = \langle \alpha, \beta \rangle \sigma,$$

for all α, gives the identical values for $*dx$, $*dy$, $*dz$, $*dx \wedge dy$, $*dy \wedge dz$, $*dz \wedge dx$.

We will now look at the Hodge star dual operator on forms on $\mathbb{R}^4$. In four dimensions is it much more difficult to try to draw pictures, but the mathematical computations are quite straight-forward. First we will define notation and our spaces. We will use x_1, x_2, x_3, x_4 as our Cartesian coordinate functions. Writing things cyclicly to keep things tidy worked out nicely in $\mathbb{R}^3$ but it is not so helpful in $\mathbb{R}^4$ so we will no longer do that. The convention here is to use increasing order. The spaces we will be working with are

$$\begin{aligned}
\bigwedge^0\left(\mathbb{R}^4\right) &= C\left(\mathbb{R}^4\right) = \text{ functions on } \mathbb{R}^4,\\
\bigwedge^1\left(\mathbb{R}^4\right) &= \text{Span}\left\{dx_1, dx_2, dx_3, dx_4\right\},\\
\bigwedge^2\left(\mathbb{R}^4\right) &= \text{Span}\left\{dx_1 \wedge dx_2, dx_1 \wedge dx_3, dx_1 \wedge dx_4, dx_2 \wedge dx_3, dx_2 \wedge dx_4, dx_3 \wedge dx_4\right\},\\
\bigwedge^3\left(\mathbb{R}^4\right) &= \text{Span}\left\{dx_1 \wedge dx_2 \wedge dx_3, dx_1 \wedge dx_2 \wedge dx_4, dx_1 \wedge dx_3 \wedge dx_4, dx_2 \wedge dx_3 \wedge dx_4\right\},\\
\bigwedge^4\left(\mathbb{R}^4\right) &= \text{Span}\left\{dx_1 \wedge dx_2 \wedge dx_3 \wedge dx_4\right\}.
\end{aligned}$$

The inner products on these spaces, necessary for the right hand side of the definition of the Hodge star dual, are given by the identity matrix. We will remain a bit ambiguous now but we discuss the metric of a manifold, from which the inner product is derived, in greater detail in Sect. A.6.

Now we find the Hodge star dual of a one-form basis element using the definition of the Hodge star dual given in Question 5.28,

$$
\begin{aligned}
dx_1 \wedge *dx_1 &= \underbrace{\langle dx_1, dx_1 \rangle}_{=1} dx_1 \wedge dx_2 \wedge dx_3 \wedge dx_4 \\
\Longrightarrow *dx_1 &= dx_2 \wedge dx_3 \wedge dx_4.
\end{aligned}
$$

Notice that by wedging dx_1 with $*dx_1$ we were able to use the nice property of the inner product on $\bigwedge^1 \left(\mathbb{R}^4\right)$ to our advantage on the right hand side of the equation where we have $\langle dx_1, dx_1 \rangle = 1$. Finding the hodge star duals of two-form basis elements is similar,

$$
\begin{aligned}
(dx_1 \wedge dx_2) \wedge *(dx_1 \wedge dx_2) &= \underbrace{\langle dx_1 \wedge dx_2, dx_1 \wedge dx_2 \rangle}_{=1} dx_1 \wedge dx_2 \wedge dx_3 \wedge dx_4 \\
\Longrightarrow *(dx_1 \wedge dx_2) &= dx_3 \wedge dx_4.
\end{aligned}
$$

Question 5.29 Find $*dx_2, *dx_3, *dx_4$.

Question 5.30 Find $*(dx_1 \wedge dx_3), *(dx_1 \wedge dx_4), *(dx_2 \wedge dx_3), *(dx_2 \wedge dx_4), *(dx_3 \wedge dx_4)$.

Question 5.31 Find $*(dx_1 \wedge dx_2 \wedge dx_3), *(dx_1 \wedge dx_2 \wedge dx_4), *(dx_1 \wedge dx_3 \wedge dx_4), *(dx_2 \wedge dx_3 \wedge dx_4)$.

Question 5.32 Find $*(dx_1 \wedge dx_2 \wedge dx_3 \wedge dx_4)$.

5.7 Summary, References, and Problems

5.7.1 Summary

We now take a moment to conclude this chapter with some remarks. If you are a mathematics major you will probably never see this way of picturing or representing differential forms in any of your textbooks; I certainly never did. If, however, you are a physics major there is a good chance that may see differential forms discussed or presented in this way, either in an electromagnetics course or in a course in relativity. These images are very useful in trying to understand certain equations and picture the physical phenomenon behind those equations. We will be exploring some of these physics applications in future chapters in detail. Though these pictures can be a great aid to picturing what equations are trying to say, using them in actual computations is generally difficult and not worth the effort since they only give approximate answers. Use the computational rules of differential forms to calculate exact numerical answers and use the pictures to aid conceptual understanding.

Also, one can develop higher dimensional analogues to these images. For example, taking into account time, and viewing space-time from the pre-relativity Euclidian perspective, results in a four dimensional version of this. Including time in both the special relativity and general relativity cases also results in four dimensional versions, but in these cases since the metric (think inner product) on the manifolds is different, the Hodge star operator, which is based on the inner product, changes. However, the higher the dimension the more difficult it is to employ these pictures and the less they are used, though it is still generally possible to imagine the 1 and the $n-1$ dimensional cases well enough.

The two equivalent definitions of the Hodge star operator are given below,

Hodge Star Operator Definition One	$\alpha \wedge \beta = \langle *\alpha, \beta \rangle \sigma$	for all β,

Hodge Star Operator Definition Two	$\alpha \wedge *\beta = \langle \alpha, \beta \rangle \sigma$	for all α.

5.7.2 References and Further Reading

The material in this chapter is encountered almost exclusively in physics. There is a reason for this - as a rigorous way of thinking about differential forms it falls short on several fronts. See Bachman [4] for this critique of visualizing forms on manifolds and where this technique falls short. But as a fairly loose way of thinking about differential forms in various physical situations it is quite useful and can be quite illuminating from the standpoint of physical systems. The primary source for much of this material is, as mentioned at the start of the chapter, Misner, Thorne, and Wheeler [34], though Dray [16], Warnick, Selfridge, and Arnold [49], Warnick and Russer [48], and DesChapes [13] were also used. All of these sources do fall within the domain of physics. In fact, the last three references strongly argue the use of these visualization techniques for differential forms as a pedagogical tool for teaching electromagnetism, though we know of no reasonably complete attempt to actually do so. The only other genuinely mathematical source that attempts to give a geometrical explanation along the lines of the material given in this chapter is Burke [8], and his intended audience is again working physicists. The software used for generating the images of one-forms as line-stacks in $\mathbb{R}^2$ was the Vector Field Analyzer II written by Kawski [28] and available online.

5.7.3 Problems

Question 5.33 Sketch the coordinate function z on $\mathbb{R}^2$ and find the "rise" of the coordinate function y along the vectors

a) $\begin{bmatrix} 0 \\ 1 \end{bmatrix}$, that is, from the point $(0, 0)$ to the point $(0, 1)$,

b) $\begin{bmatrix} 1 \\ 1 \end{bmatrix}$, that is, from the point $(0, 0)$ to the point $(1, 1)$,

c) $\begin{bmatrix} -1 \\ 2 \end{bmatrix}$, that is, from the point $(0, 0)$ to the point $(-1, 2)$,

d) $\begin{bmatrix} 2 \\ -3 \end{bmatrix}$, that is, from the point $(0, 0)$ to the point $(2, -3)$.

Question 5.34 Draw a picture representing the one-forms $7dx$, $-3dy$, $\frac{-2}{3}dx$, $\frac{-1}{3}dy$, and $\frac{2}{5}dy$ on $\mathbb{R}^2$.

Question 5.35 Draw the graphical representations of the one-forms $\frac{1}{2}dx + 2dy$, $\frac{-1}{3}dx - \frac{1}{2}dy$, and $-dx - \frac{2}{3}dy$ on $\mathbb{R}^2$.

Question 5.36 Draw three graphical representations for each of the two-forms $5\, dx \wedge dy$, $\frac{1}{2}\, dx - \frac{1}{2}\, dy$, and $2dx + \frac{2}{3}dy$ on $\mathbb{R}^2$.

Question 5.37 Sketch the one-forms $dx - dy$, $-dx + dy$, $dy - dz$, and $-dx - dz$ on $\mathbb{R}^3$.

Question 5.38 Find two graphical representations for each of the two-forms $5dx \wedge dy$, $6dy \wedge dz$, and $\frac{1}{4}dz \wedge dx$ on $\mathbb{R}^3$.

Question 5.39 Sketch a picture for each of the two-forms $3dy \wedge dz + 2dz \wedge dx$, $dx \wedge dy + 3dz \wedge dx$, and $2dx \wedge dy + 4dy \wedge dz + 3dz \wedge dx$ on $\mathbb{R}^3$.

Question 5.40 Sketch the two-forms $-dx \wedge dy + dy \wedge dz$, $dx \wedge dy - dy \wedge dz$, and $-dx \wedge dy - dy \wedge dz$ on $\mathbb{R}^3$. What affect do the negative signs have? Use some well chosen vectors to help you figure it out.

Question 5.41 Match the below one-forms with the images depicted in Fig. 5.40

a) $\sin(x)dx + ydy$,
b) $dx - dy$,
c) $xdx + \cos(y)dy$,
d) $\sin(x)dx + xdy$.

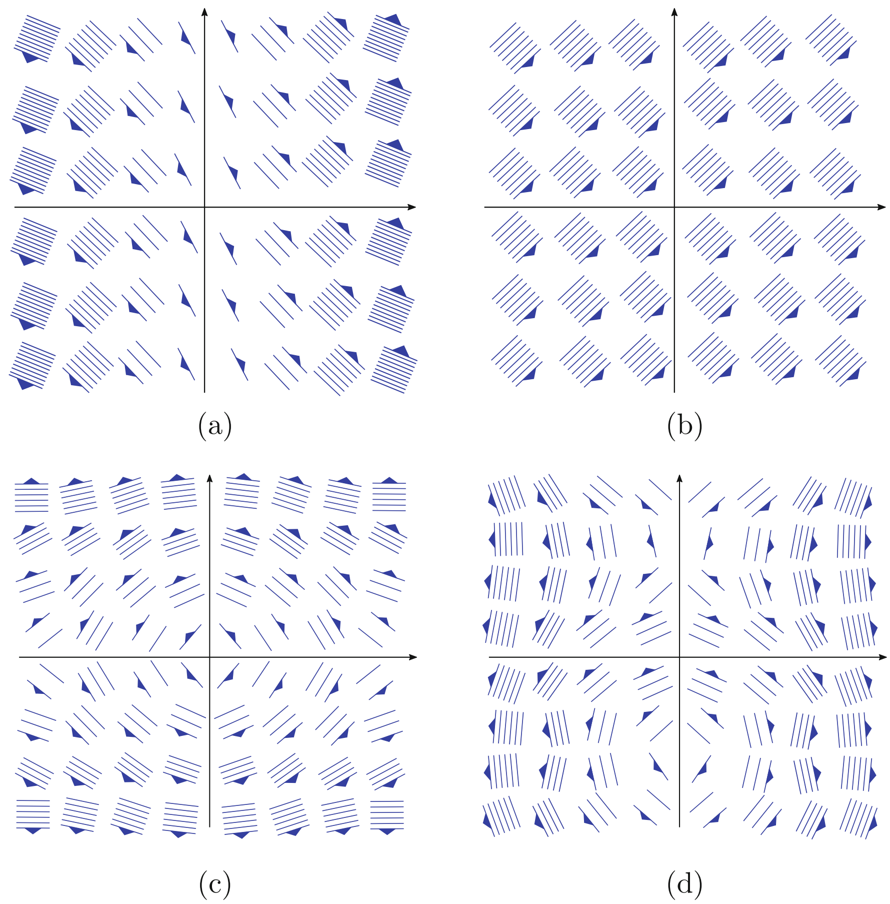


Fig. 5.40 Images for Question 5.41 (Images generated with Vector field analyzer II, version 2, Kawski, 2001)

Question 5.42 Assume the manifold is $\mathbb{R}^4$. Find $*dx_2$, $*dx_3$, $*dx_4$.

Question 5.43 Assume the manifold is $\mathbb{R}^4$. Find $*(dx_1 \wedge dx_3)$, $*(dx_1 \wedge dx_4)$, $*(dx_2 \wedge dx_3)$, $*(dx_2 \wedge dx_4)$, $*(dx_3 \wedge dx_4)$.

Question 5.44 Assume the manifold is $\mathbb{R}^4$. Find $*(dx_1 \wedge dx_2 \wedge dx_3)$, $*(dx_1 \wedge dx_2 \wedge dx_4)$, $*(dx_1 \wedge dx_3 \wedge dx_4)$, $*(dx_2 \wedge dx_3 \wedge dx_4)$.

Question 5.45 Assume the manifold is $\mathbb{R}^4$. Find $*(dx_1 \wedge dx_2 \wedge dx_3 \wedge dx_4)$.

Chapter 6
Push-Forwards and Pull-Backs

In this chapter we introduce two extremely important concepts, the push-forward of a vector and the pull-back of a differential form. In section one we take a close look at a simple change of coordinates and see what affect this change of coordinates has on the volume of the unit square. This allows us to motivate the push-forward of a vector in section two. Push-forwards of vectors allow us to move, or "push-forward," a vector from one manifold to another. In the case of coordinate changes the two manifolds are actually the same manifold, only equipped with different coordinate systems.

Using the fact that differential forms eat vectors, in section three we use the push-forwards of vectors to define the pull-back of a differential form, where we "pull-back" the differential form on one manifold to another manifold. Again, in the case of coordinate changes the two manifolds are really the same manifold, only equipped with different coordinate systems. We then look more closely at the pull-back of a volume form, which plays an essential role in the integration problems in the section three. We are able to derive a nice formula for the pull-back of a volume form that involves the Jacobian matrix of the coordinate change.

We then use push-forwards and pull-backs to look closely at some familiar examples. Polar coordinates are treated in section four and both cylindrical and spherical coordinates are considered in section five. In section six we consider the pull-back of differential forms that are not volume forms, and finally in section seven we prove three identities that are crucial for us to do computations involving pull-backs.

6.1 Coordinate Change: A Linear Example

This section will serve as our first look at coordinate changes. We aren't going to try to be at all rigorous in this section, this is really an attempt to just start to understand the big picture a little better. Before we start getting fancy with polar and spherical coordinates, let's consider something simple. A change in coordinates is nothing more than a mapping between $\mathbb{R}^n$ and $\mathbb{R}^n$ that is both one-to-one and onto. Since $\mathbb{R}^n$ is the same as $\mathbb{R}^n$ we often think of it as a mapping from $\mathbb{R}^n$ to itself. Consider the two mappings $u, v : \mathbb{R}^2 \to \mathbb{R}^2$ given by

$$\begin{aligned} u(x, y) &= x + y, \\ v(x, y) &= x - y. \end{aligned}$$

Figure 6.1 shows this change of coordinates. On the left is the $\mathbb{R}^2$ in xy-coordinates, which we will also call the xy-plane or $\mathbb{R}^2_{xy}$, and on the right is $\mathbb{R}^2$ in uv-coordinates, which we will also call the uv-plane or $\mathbb{R}^2_{uv}$. So, even though $\mathbb{R}^2$ is $\mathbb{R}^2$, we are making a distinction between the two copies based on what coordinates we are using. On the top left the x and y grid lines from the xy-plane are mapped to lines on the uv-plane on the right, and on the bottom right the u and v grid lines are mapped to lines on the xy-plane on the left.

Question 6.1 Find the image of the following lines under the mapping $u = x + y$ and $v = x - y$ and compare what you find with the mapping shown at the top of Fig. 6.1.

$a)\ x = 0$	$b)\ x = 1$	$c)\ x = 4$
$d)\ y = 0$	$e)\ y = 1$	$f)\ y = -3$

J. P. Fortney, *A Visual Introduction to Differential Forms and Calculus on Manifolds*,
https://doi.org/10.1007/978-3-319-96992-3_6

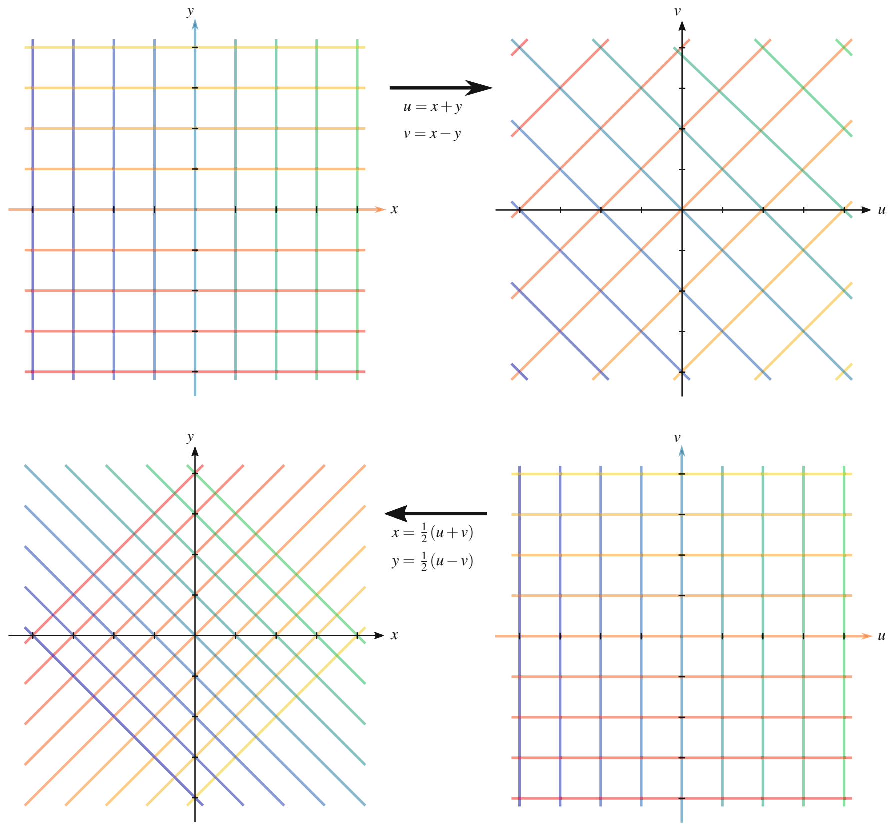


Fig. 6.1 The planes $\mathbb{R}^2_{xy}$ (left) and $\mathbb{R}^2_{uv}$ (right). The change of coordinates given by $u(x, y) = x + y$ and $v(x, y) = x - y$ is on the top, left to right. The inverse change of coordinates $x(u, v) = 0.5(u + v)$ and $y(u, v) = 0.5(u - v)$ is on the bottom, right to left. The color coding shows how various lines are mapped

Question 6.2 Find the image of the following lines under the inverse mapping $x = \frac{1}{2}(u + v)$ and $y = \frac{1}{2}(u - v)$ and compare what you find with the mapping shown at the bottom of Fig. 6.1.

$a)\ u = 0$	$b)\ u = 1$	$c)\ u = 4$
$d)\ v = 0$	$e)\ v = 1$	$f)\ v = -3$

We use the coordinate functions $u = x + y$ and $v = x - y$ to map points in the plane $\mathbb{R}^2_{xy}$ to points in $\mathbb{R}^2_{uv}$ as follows:

$$\begin{aligned}
\underline{(x, y)} &\longrightarrow \underline{(x + y, x - y) = (u, v)}\\
(0, 0) &\longrightarrow (0 + 0, 0 - 0) = (0, 0)\\
(1, 0) &\longrightarrow (1 + 0, 1 - 0) = (1, 1)\\
(1, 1) &\longrightarrow (1 + 1, 1 - 1) = (2, 0)\\
(0, 1) &\longrightarrow (0 + 1, 0 - 1) = (1, -1)
\end{aligned}$$

As we move around the unit square in $\mathbb{R}^2_{xy}$ we can follow what happen in $\mathbb{R}^2_{uv}$.

$$\begin{aligned}
(x, y):&\quad (0, 0) \longrightarrow (1, 0) \longrightarrow (1, 1) \longrightarrow (0, 1) \longrightarrow (0, 0)\\
(u, v):&\quad (0, 0) \longrightarrow (1, 1) \longrightarrow (2, 0) \longrightarrow (1, -1) \longrightarrow (0, 0)
\end{aligned}$$

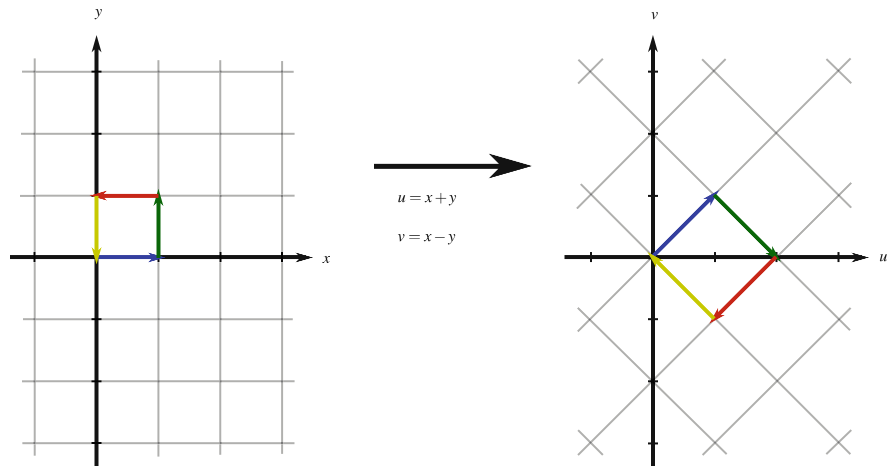


Fig. 6.2 A unit square in $\mathbb{R}^2_{xy}$ (left) is mapped to the diamond in $\mathbb{R}^2_{uv}$ (right). Notice the orientation switches from counter-clockwise to clockwise and the area increases from one to two

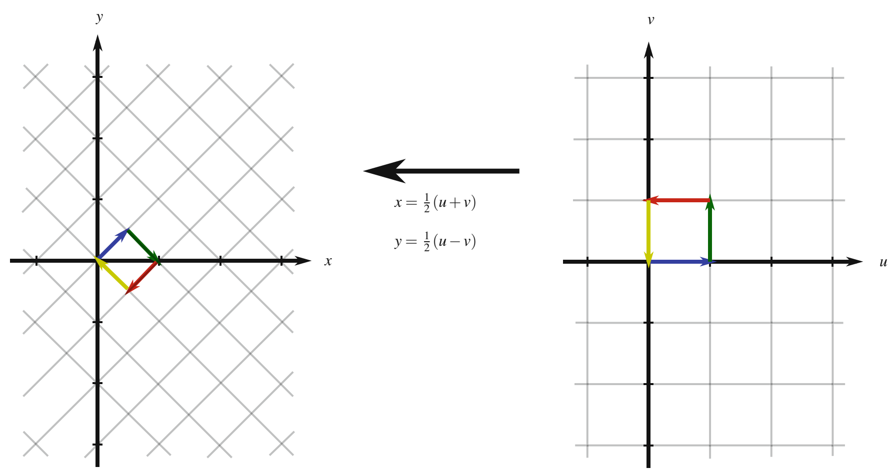


Fig. 6.3 A unit square in $\mathbb{R}^2_{uv}$ (right) is mapped to the diamond in $\mathbb{R}^2_{xy}$ (left). Notice the orientation switches from counter-clockwise to clockwise and the area decreases from one to a half

With this we can see, as shown in Fig. 6.2, that the unit square from the xy-coordinates gets mapped to a diamond in uv-coordinates. Notice that as you move around the unit square in xy-coordinates in a counter-clockwise direction, shown on the left of Fig. 6.2, you move around the diamond in the uv-coordinates in a clockwise direction, shown on the right of Fig. 6.2. There is also another interesting difference. Consider the area of the unit square in the xy-plane; it has an area of one, while the area of the diamond in the uv-plane, which is the image of the unit square, has an area of two.

Question 6.3 Show what happens to the points $(0, 0)$, $(1, 0)$, $(1, 1)$, and $(0, 1)$ under the inverse mapping $x = \frac{1}{2}(u + v)$ and $y = \frac{1}{2}(u - v)$.

Similarly, if we used the inverse mapping to take a unit square from the uv-coordinate plane, shown on the right of Fig. 6.3, to the xy-coordinate plane we get a diamond in the xy-coordinate plane, shown on the left of Fig. 6.3, with half the area and the direction reversed. Next notice that

$$dx \wedge dy \underbrace{\left(\begin{bmatrix} 1 \\ 0 \end{bmatrix}_{(0,0)}, \begin{bmatrix} 0 \\ 1 \end{bmatrix}_{(0,0)} \right)}_{\text{vectors in } xy\text{-plane}} = 1,$$

$$du \wedge dv \underbrace{\left(\begin{bmatrix} 1 \\ 0 \end{bmatrix}_{(0,0)}, \begin{bmatrix} 0 \\ 1 \end{bmatrix}_{(0,0)} \right)}_{\text{vectors in } uv\text{-plane}} = 1$$

so $dx \wedge dy$ is the area form of the xy-plane and $du \wedge dv$ is the area form of the uv-plane.

First we will write $du \wedge dv$ in terms of $dx \wedge dy$. Part of this has already been done. Since $u = x + y$ and $v = x - y$ we have

$$\begin{aligned} du &= d(x+y) = dx + dy, \\ dv &= d(x-y) = dx - dy \end{aligned}$$

so we have

$$\begin{aligned} du \wedge dv &= (dx + dy) \wedge (dx - dy) \\ &= dx \wedge (dx - dy) + dy \wedge (dx - dy) \\ &= dx \wedge dx - dx \wedge dy + dy \wedge dx - dy \wedge dy \\ &= -dx \wedge dy - dx \wedge dy \\ &= -2dx \wedge dy. \end{aligned}$$

Next we will write $dx \wedge dy$ in terms of $du \wedge dv$. In the above question you showed that solving for x and y in terms of u and v gives us

$$\begin{aligned} x &= \frac{1}{2}(u+v), \\ y &= \frac{1}{2}(u-v) \end{aligned}$$

which in turn gives us

$$\begin{aligned} dx &= d\left(\frac{1}{2}(u+v)\right) = \frac{1}{2}du + \frac{1}{2}dv, \\ dy &= d\left(\frac{1}{2}(u-v)\right) = \frac{1}{2}du - \frac{1}{2}dv \end{aligned}$$

so we have

$$\begin{aligned} dx \wedge dy &= \left(\frac{1}{2}du + \frac{1}{2}dv\right) \wedge \left(\frac{1}{2}du - \frac{1}{2}dv\right) \\ &= \frac{1}{4}dv \wedge du - \frac{1}{4}du \wedge dv \\ &= -\frac{1}{2}du \wedge dv. \end{aligned}$$

In summary, what we have just done is compute the following relations between the xy and the uv area forms

$$\begin{aligned} du \wedge dv &= -2dx \wedge dy, \\ dx \wedge dy &= -\frac{1}{2}du \wedge dv. \end{aligned}$$

Compare what is happening in Figs. 6.2 and 6.3 with the relationships between the area forms. How do they relate to each other? First, we try to understand what the volume form relations are telling us. First consider Fig. 6.2 where the unit square in the xy-plane was mapped to the diamond in the uv-plane. The volume of the diamond in the uv-plane is -2 times the volume of the unit square in the xy-plane. We can see where the 2 comes from by noticing that the unit square in the xy-plane is mapped to the diamond in the uv-plane with twice the area. And where does the negative sign come from? From noticing that the counter-clockwise rotation in the xy-plane becomes a clockwise rotation in the uv-plane. This is our signed volume. So, what we are showing must have some relation to the identity

$$du \wedge dv = -2dx \wedge dy.$$

Next consider Fig. 6.3 where the unit square in the uv-plane was mapped to the diamond in the xy-plane. The area of the diamond in the xy-plane is $-\frac{1}{2}$ times the volume of the unit square in the uv-plane. This means that the unit square with (signed) volume 1 in the uv-plane is mapped to the diamond with (signed) volume $-\frac{1}{2}$ in the xy-plane, where the negative sign again indicates that a counter-clockwise rotation around the square in the uv-plane becomes a clockwise rotation around diamond in the xy-plane. Again, this clearly has some relation to the identity

$$dx \wedge dy = -\frac{1}{2} du \wedge dv.$$

Recall that when we were deriving the formula for the determinant in Sect. 1.2 we started out by specifying three fundamental and intuitive properties that we thought volumes should have:

(1) A unit cube should have volume one (in any dimension).
(2) Degenerate parallelepipeds (that is, parallelepipeds of less than n dimensions in $\mathbb{R}^n$) should have an n-dimensional volume of zero.
(3) Scaling an edge by a factor c should change the volume by the factor c.

By assuming our volume function D had only these three properties we discovered the fourth property, that by interchanging (switching) two edges of our parallelepiped we also changed the sign of the volume. In other words, fundamental to our intuitive notion of volume is the notion of orientation. In three dimensions orientation is often disguised as the "right-hand rule." See Fig. 6.4.

Over the last few pages we have been a little imprecise, but at least now you may be starting to see how these volume forms in different coordinate systems relate to each other - they somehow encode changes of volume as you move from one coordinate system to another. This is exactly the sort of thing that would be useful when we change variables during integration.

Now we want to look at volume forms a little more closely. The unit square in the xy-plane is given by (spanned by) the xy-plane vectors

$$\begin{bmatrix} 1 \\ 0 \end{bmatrix}_{\substack{xy \\ (0,0)}} \quad \text{and} \quad \begin{bmatrix} 0 \\ 1 \end{bmatrix}_{\substack{xy \\ (0,0)}}.$$

In other words, the parallelepiped spanned by these two vectors is the unit square. Here we have used the xy in the subscript to indicate the coordinate plane that the vectors are in. The projection of these vectors under the change of coordinates $u = x + y$ and $v = x - y$ is given by

$$\begin{bmatrix} 1 \\ 0 \end{bmatrix}_{\substack{xy \\ (0,0)}} \longmapsto \begin{bmatrix} 1 \\ 1 \end{bmatrix}_{\substack{uv \\ (0,0)}},$$

$$\begin{bmatrix} 0 \\ 1 \end{bmatrix}_{\substack{xy \\ (0,0)}} \longmapsto \begin{bmatrix} 1 \\ -1 \end{bmatrix}_{\substack{uv \\ (0,0)}}.$$

See Fig. 6.5. It is easy to compute that

$$dx \wedge dy \left(\begin{bmatrix} 1 \\ 0 \end{bmatrix}_{\substack{xy \\ (0,0)}}, \begin{bmatrix} 0 \\ 1 \end{bmatrix}_{\substack{xy \\ (0,0)}} \right) = \begin{vmatrix} 1 & 0 \\ 0 & 1 \end{vmatrix} = 1,$$

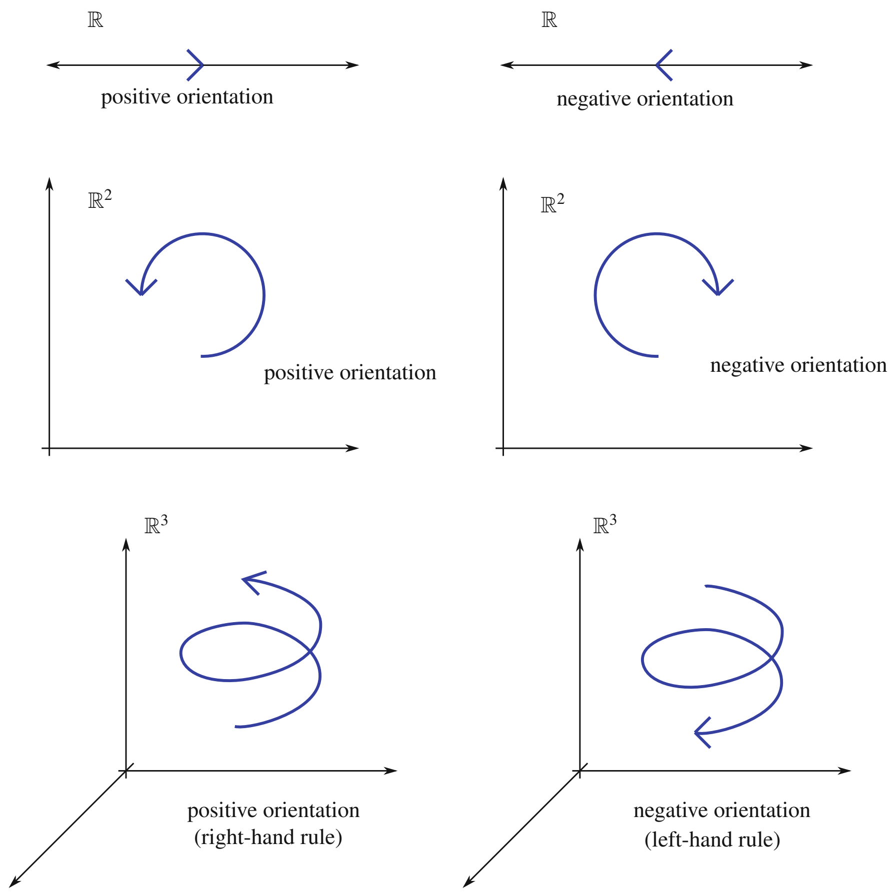


Fig. 6.4 The two orientations for $\mathbb{R}$ (top), for $\mathbb{R}^2$ (middle), and for $\mathbb{R}^3$ (bottom)

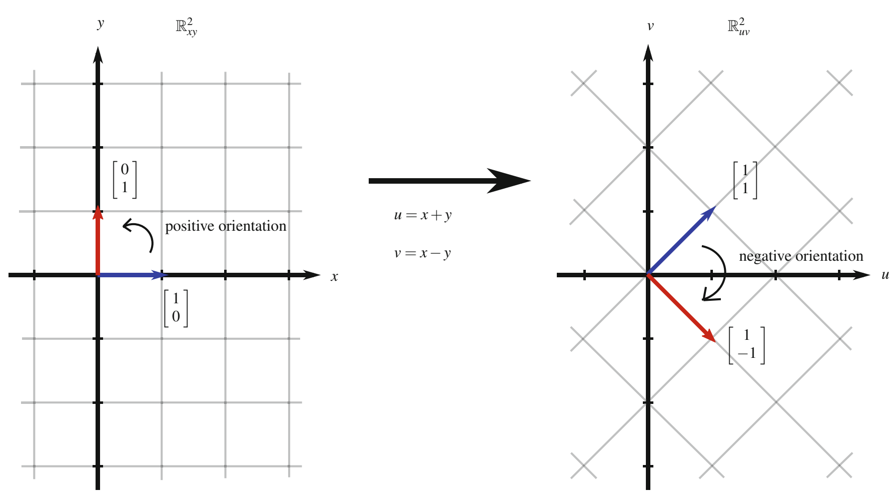


Fig. 6.5 The basis vectors in the xy-plane mapped to two vectors in the uv-plane. Notice the orientation changes

$$du \wedge dv \left(\begin{bmatrix} 1 \\ 1 \end{bmatrix}_{\substack{uv \\ (0,0)}}, \begin{bmatrix} 1 \\ -1 \end{bmatrix}_{\substack{uv \\ (0,0)}} \right) = \begin{vmatrix} 1 & 1 \\ 1 & -1 \end{vmatrix} = -2.$$

Recalling that the area form finds the area of the parallelepiped spanned by the input vectors, this coincides exactly with what we would expect from our picture. We also start to get a clearer picture of what is meant by the identity

$$dx \wedge dy = -\frac{1}{2} du \wedge dv.$$

Actually we have

$$dx \wedge dy \underbrace{\left(\begin{bmatrix} 1 \\ 0 \end{bmatrix}_{\substack{xy \\ (0,0)}}, \begin{bmatrix} 0 \\ 1 \end{bmatrix}_{\substack{xy \\ (0,0)}} \right)}_{\text{vectors in } xy\text{-plane}} = -\tfrac{1}{2} du \wedge dv \underbrace{\left(\begin{bmatrix} 1 \\ 1 \end{bmatrix}_{\substack{uv \\ (0,0)}}, \begin{bmatrix} 1 \\ -1 \end{bmatrix}_{\substack{uv \\ (0,0)}} \right)}_{\text{projected vectors in } uv\text{-plane}}.$$

The area form $dx \wedge dy$ has to eat vectors from the xy-plane and the area form $du \wedge dv$ has to eat vectors from the uv-plane. Furthermore, the vectors that the area form $du \wedge dv$ is eating are the projections under the u and v mappings ($u = x + y$ and $v = x - y$) of the vectors that the area form $dx \wedge dy$ is eating.

Similarly, the unit square in the uv-plane is spanned by the vectors

$$\begin{bmatrix} 1 \\ 0 \end{bmatrix}_{\substack{uv \\ (0,0)}} \text{ and } \begin{bmatrix} 0 \\ 1 \end{bmatrix}_{\substack{uv \\ (0,0)}}.$$

Under the change of coordinates given by the inverse of the above, $x = \frac{1}{2}(u+v)$ and $y = \frac{1}{2}(u-v)$, these vector's projections are given by

$$\begin{bmatrix} 1 \\ 0 \end{bmatrix}_{\substack{uv \\ (0,0)}} \longmapsto \begin{bmatrix} 1/2 \\ 1/2 \end{bmatrix}_{\substack{xy \\ (0,0)}},$$

$$\begin{bmatrix} 0 \\ 1 \end{bmatrix}_{\substack{uv \\ (0,0)}} \longmapsto \begin{bmatrix} 1/2 \\ -1/2 \end{bmatrix}_{\substack{xy \\ (0,0)}}.$$

See Fig. 6.6. Again, it is easy to compute

$$du \wedge dv \left(\begin{bmatrix} 1 \\ 0 \end{bmatrix}_{\substack{uv \\ (0,0)}}, \begin{bmatrix} 0 \\ 1 \end{bmatrix}_{\substack{uv \\ (0,0)}} \right) = \begin{vmatrix} 1 & 0 \\ 0 & 1 \end{vmatrix} = 1,$$

$$dx \wedge dy \left(\begin{bmatrix} 1/2 \\ 1/2 \end{bmatrix}_{\substack{xy \\ (0,0)}}, \begin{bmatrix} 1/2 \\ -1/2 \end{bmatrix}_{\substack{xy \\ (0,0)}} \right) = \begin{vmatrix} 1/2 & 1/2 \\ 1/2 & -1/2 \end{vmatrix} = -\frac{1}{2}.$$

Again this coincides with what we would expect from our picture. And again, we start to get a better idea of how the identity

$$du \wedge dv = -2 dx \wedge dy$$

actually works,

$$du \wedge dv \underbrace{\left(\begin{bmatrix} 1 \\ 0 \end{bmatrix}_{\substack{uv \\ (0,0)}}, \begin{bmatrix} 0 \\ 1 \end{bmatrix}_{\substack{uv \\ (0,0)}} \right)}_{\text{vectors in } uv\text{-plane}} = -2 dx \wedge dy \underbrace{\left(\begin{bmatrix} 1/2 \\ 1/2 \end{bmatrix}_{\substack{xy \\ (0,0)}}, \begin{bmatrix} 1/2 \\ -1/2 \end{bmatrix}_{\substack{xy \\ (0,0)}} \right)}_{\text{projected vectors in } xy\text{-plane}}.$$

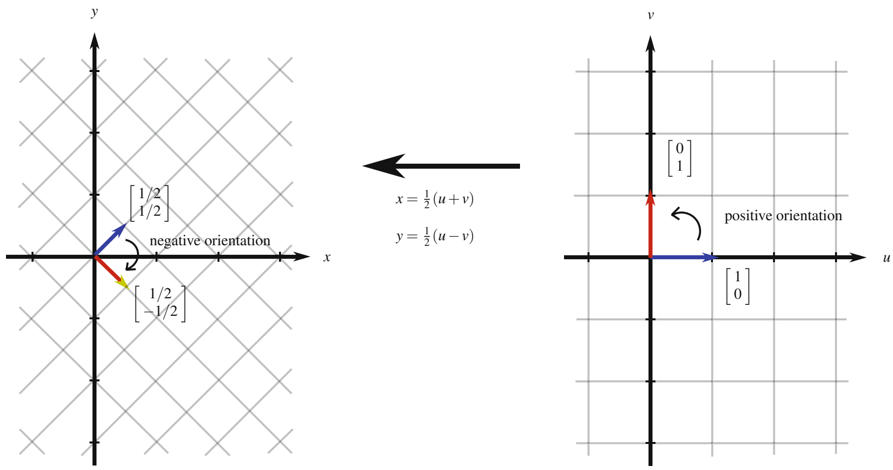


Fig. 6.6 The basis vectors in the uv-plane mapped to two vectors in the xy-plane. Notice the orientation changes

The area form $du \wedge dv$ eats vectors from the uv-plane and the area form $dx \wedge dy$ eats vectors from the xy-plane. Furthermore, the vectors that $dx \wedge dy$ is eating are projections under the x and y mappings ($x = \frac{1}{2}(u+v)$ and $y = \frac{1}{2}(u-v)$) of the vectors that $du \wedge dv$ is eating. So our identities

$$du \wedge dv = -2dx \wedge dy,$$

$$dx \wedge dy = -\frac{1}{2}du \wedge dv$$

hold, but only when there is a specific relationship between the vectors that they are eating, that is, when the vectors they are eating are related by the mappings between the two spaces.

Of course, in this section we have committed the cardinal vector calculus sin and have viewed the various vectors being eaten by the volume forms to be part of the manifold $\mathbb{R}^2$ and not elements of $T_{(0,0)}\mathbb{R}^2$ like they really are. We will correct this in the next section. Our main point in this section was to try to understand the subtleties involved with the identities between the volume forms a little better.

6.2 Push-Forwards of Vectors

In the last section we looked at the (linear change of coordinate) mapping $f(x, y) = \big(u(x, y), v(x, y)\big) : \mathbb{R}^2 \to \mathbb{R}^2$ given by $u(x, y) = x+y$ and $v(x, y) = x-y$. We then used these to find the identities $du \wedge dv = -2dx \wedge dy$ and $dx \wedge dy = -\frac{1}{2}du \wedge dv$ and we started to explore what these identities actually mean. They have to do with how the areas in the domain manifold $\mathbb{R}^2$ with xy-coordinates, which we will denote $\mathbb{R}^2_{xy}$, relate to the areas in the range manifold $\mathbb{R}^2$ with uv-coordinates, which we will denote $\mathbb{R}^2_{uv}$. We found that the identity $du \wedge dv = -2dx \wedge dy$ worked when the vectors that $du \wedge dv$ ate were the projections under the mapping $f(x, y) = \big(u(x, y), v(x, y)\big) = (x + y, x - y)$ of the vectors that $dx \wedge dy$ ate. Similarly, when we were considering the inverse mapping we found the identity $dx \wedge dy = -\frac{1}{2}du \wedge dv$ worked when the vectors that $dx \wedge dy$ ate were projections under the inverse mapping of the vectors that $du \wedge dv$ ate. However, when doing this we were rather imprecise with the vectors that we used to determine the parallelepipeds whose volumes we were finding. As we know vectors really live in tangent spaces, not in the manifold. As we will see, we were able to gloss over that fact here because the mapping was linear.

Not that we want to bore you to death with this example, but in this section we will continue to use this example while being more mathematically precise. This will serve as motivation and an introduction to the idea of **push-forwards** of vectors. Later in the chapter this example will also serve as motivation and an introduction to the idea of **pull-backs** of differential forms. Also, we want to work with a simple example before introducing more complicated examples like polar or spherical or cylindrical coordinates.

Next we recall that both the manifold $\mathbb{R}^2_{xy}$ (with xy-coordinates) and the manifold $\mathbb{R}^2_{uv}$ (with uv-coordinates), see Fig. 6.1, have a tangent space associated to each point similar to what was shown in Fig. 2.15. That is, if (x, y) is a point in $\mathbb{R}^2_{xy}$ then $T_{(x,y)}\mathbb{R}^2_{xy}$ is the tangent space of $\mathbb{R}^2_{xy}$ at the point (x, y). The point $f(x, y) = (x + y, x - y)$ is the projection of the point (x, y) to $\mathbb{R}^2_{uv}$ and $T_{(x+y,x-y)}\mathbb{R}^2_{uv}$ is the tangent space of $\mathbb{R}^2_{uv}$ at the projected point $(x + y, x - y)$.

Also, for the mapping $f(x, y) = \big(u(x, y), v(x, y)\big) = (x + y, x - y)$ we have the derivative of f at the point (x, y) given by the Jacobian matrix Df,

$$D_{(x,y)}f = \begin{bmatrix} \frac{\partial u}{\partial x} & \frac{\partial u}{\partial y} \\ \frac{\partial v}{\partial x} & \frac{\partial v}{\partial y} \end{bmatrix}_{(x,y)}.$$

Let us think about what the Jacobian matrix is and does. It is the closest linear approximation to the function $f(x, y)$ at the point (x, y). Suppose we have some vector $v = \begin{bmatrix} a \\ b \end{bmatrix}$ based at the point (x, y). This vector is really in the tangent space of $\mathbb{R}^2_{xy}$ at the point (x, y) but for now we still think of it as being in the manifold $\mathbb{R}^2_{xy}$. If we wanted to know how much f changed as we moved from (x, y) along the vector v we could find $f(x, y)$ and $f(x+a, y+b)$ to see the change. In fact, we could think of the change as being a vector from $f(x, y)$ to $f(x + a, y + b)$. Here both the points (x, y) and $(x + a, y + b)$ are in the manifold $\mathbb{R}^2_{xy}$ and the points $f(x, y)$ and $f(x + a, y + b)$ are in the manifold $\mathbb{R}^2_{uv}$. Or we could estimate the change using the closest linear approximation to f at the point (x, y), which is the derivative of f at the point (x, y), and which is given by the Jacobian matrix Df at the point (x, y). To do this we multiply the Jacobian matrix by the vector v, or find $Df \cdot v$. This gives us another vector. But where is this vector at? Since we are estimating a change in f from the point $f(x, y)$, this point is the base point for our new vector, so our new vector is in the tangent space of $\mathbb{R}^2_{uv}$ at the point $f(x, y)$. In summary, the Jacobian matrix is actually a mapping

$$D_{(x,y)}f : T_{(x,y)}\mathbb{R}^2_{xy} \longrightarrow T_{f(x,y)}\mathbb{R}^2_{uv}.$$

For reasons that we will explain a little bit later, we will generally use the notation $T_{(x,y)}f$ to denote the Jacobian matrix at the point (x, y) instead of the more common $D_{(x,y)}f$. For our particular mapping $f(x, y) = \big(u(x, y), v(x, y)\big) = (x+y, x-y)$ we have

$$T_{(x,y)}f = \begin{bmatrix} \frac{\partial u}{\partial x} & \frac{\partial u}{\partial y} \\ \frac{\partial v}{\partial x} & \frac{\partial v}{\partial y} \end{bmatrix}_{(x,y)} = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}_{(x,y)}.$$

For example, at the point $(0, 0)$ the vector $\begin{bmatrix} 1 \\ 0 \end{bmatrix}_{(0,0)}$ is sent to

$$T_{(0,0)}f \cdot \begin{bmatrix} 1 \\ 0 \end{bmatrix}_{(0,0)} = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}_{(0,0)} \cdot \begin{bmatrix} 1 \\ 0 \end{bmatrix}_{(0,0)} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}_{f(0,0)=(0,0)}$$

while the vector $\begin{bmatrix} 0 \\ 1 \end{bmatrix}_{(0,0)}$ is sent to

$$T_{(0,0)}f \cdot \begin{bmatrix} 0 \\ 1 \end{bmatrix}_{(0,0)} = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}_{(0,0)} \cdot \begin{bmatrix} 0 \\ 1 \end{bmatrix}_{(0,0)} = \begin{bmatrix} 1 \\ -1 \end{bmatrix}_{f(0,0)=(0,0)}.$$

So, from the function

$$\begin{aligned} f : \mathbb{R}^2_{xy} &\longrightarrow \mathbb{R}^2_{uv} \\ p &\longmapsto f(p) \end{aligned}$$

we have *induced* a mapping

$$\begin{aligned} T_p f : T_p\mathbb{R}^2_{xy} &\longrightarrow T_{f(p)}\mathbb{R}^2_{uv} \\ v_p &\longmapsto T_p f \cdot v_p \end{aligned}$$

from the tangent bundle of the domain manifold $\mathbb{R}^2_{xy}$ to the tangent bundle of the range manifold $\mathbb{R}^2_{uv}$, which is nothing more than the Jacobian matrix. By *induced* we mean that we used f to derive another map Tf, or $T_p f$ when we take into account the base point p. We call $T_p f \cdot v_p$ the **push-forward** of v_p by $T_p f$ and $T_p f$ is sometimes called the **push-forward mapping**.

Let us do one more example, suppose we wanted to find $T_p f \cdot v_p$ when

$$v_p = \begin{bmatrix} 1 \\ 2 \end{bmatrix}_{(1,1)}.$$

First we find $f(1,1) = (1+1, 1-1) = (2,0)$. Next we have

$$T_{(1,1)} f \cdot \begin{bmatrix} 1 \\ 2 \end{bmatrix}_{(1,1)} = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}_{(1,1)} \cdot \begin{bmatrix} 1 \\ 2 \end{bmatrix}_{(1,1)} = \begin{bmatrix} 3 \\ -1 \end{bmatrix}_{f(1,1)} = \begin{bmatrix} 3 \\ -1 \end{bmatrix}_{(2,0)}.$$

Notice that on the left the base point is $(1,1)$ but on the right the base point is $f(1,1) = (2,0)$.

Question 6.4 Using the same function f, find $T_p f \cdot v_p$ of the following vectors.

(a) $v_p = \begin{bmatrix} -1 \\ -1 \end{bmatrix}_{(5,4)}$

(b) $v_p = \begin{bmatrix} -2 \\ 3 \end{bmatrix}_{(2,-3)}$

Since f is already a linear function then the Jacobian of f, that is the mapping $T_p f$, is the same at each point. We will soon meet examples where this is not true. But there is something else to notice as well. Consider a generic vector given by $v_p = \begin{bmatrix} x \\ y \end{bmatrix}_p$. Then

$$T_p f \cdot \begin{bmatrix} x \\ y \end{bmatrix}_p = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}_p \cdot \begin{bmatrix} x \\ y \end{bmatrix}_p = \begin{bmatrix} x+y \\ x-y \end{bmatrix}_{f(p)}.$$

We have that $T_p f$ acting on $\begin{bmatrix} x \\ y \end{bmatrix}_p \in T_p\mathbb{R}^2$, where $p \in$ manifold $\mathbb{R}^2$, looks exactly the same as the function f acting on $(x,y) \in$ manifold $\mathbb{R}^2$,

$$\begin{aligned} f(x,y) &= (x+y, x-y) \in \mathbb{R}^2 \\ T_p f \cdot \begin{bmatrix} x \\ y \end{bmatrix}_p &= \begin{bmatrix} x+y \\ x-y \end{bmatrix}_{f(p)} \in T_{f(p)}\mathbb{R}^2, \;\; p \in \mathbb{R}^2. \end{aligned}$$

Of course the domains and ranges of Tf and f are different spaces, but the way the functions Tf and f act on the input points *looks* the same. Very often *for linear functions* f one will see it written that

$$Tf = f.$$

Clearly, since the domains and ranges are different this can not be technically true, but this is one of those convenient so-called "abuses of notation." (Actually, abuses of notation are exceedingly common in mathematics. When you are first learning something these abuses of notation can often imped understanding, but once you understand what is going on they make doing computations and writing things down so much simpler and faster. To be absolutely precise all the time becomes overwhelming and difficult, and unnecessary to someone who really understands what is going on.)

Putting what we have just learned into use we will revisit what we did in the last section. Sticking with our same function $f(x, y) = \big(u(x, y), v(x, y)\big) = (x + y, x - y)$, consider the point $(5, 7) \in \mathbb{R}^2_{xy}$. We find that $f(5, 7) = (5 + 7, 5 - 7) = (12, -2)$. Next, we find the area of the parallelepiped spanned by the Euclidian unit vectors

$$\begin{bmatrix} 1 \\ 0 \end{bmatrix}_{(5,7)}, \begin{bmatrix} 0 \\ 1 \end{bmatrix}_{(5,7)} \in T_{(5,7)}\mathbb{R}^2_{xy}$$

at that point. Technically, this parallelepiped lives in the tangent space $T_{(5,7)}\mathbb{R}^2_{xy}$, but given that we can naturally identify the tangent space with the underlying manifold $\mathbb{R}^2_{xy}$ we often also think of the parallelepiped as being in the manifold. Using our area form $dx \wedge dy$ we get the area as

$$(dx \wedge dy)_{(5,7)} \left(\begin{bmatrix} 1 \\ 0 \end{bmatrix}_{(5,7)}, \begin{bmatrix} 0 \\ 1 \end{bmatrix}_{(5,7)} \right) = \begin{vmatrix} 1 & 0 \\ 0 & 1 \end{vmatrix} = 1.$$

Next we use $T_{(5,7)}f$ to find the push-forwards of the vectors to the space $T_{(12,-2)}\mathbb{R}^2_{uv}$;

$$T_{(5,7)}f \cdot \begin{bmatrix} 1 \\ 0 \end{bmatrix}_{(5,7)} = \begin{vmatrix} 1 & 1 \\ 1 & -1 \end{vmatrix} \cdot \begin{bmatrix} 1 \\ 0 \end{bmatrix}_{(5,7)} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}_{(12,-2)} \in T_{(12,-2)}\mathbb{R}^2_{uv},$$
$$T_{(5,7)}f \cdot \begin{bmatrix} 0 \\ 1 \end{bmatrix}_{(5,7)} = \begin{vmatrix} 1 & 1 \\ 1 & -1 \end{vmatrix} \cdot \begin{bmatrix} 0 \\ 1 \end{bmatrix}_{(5,7)} = \begin{bmatrix} 1 \\ -1 \end{bmatrix}_{(12,-2)} \in T_{(12,-2)}\mathbb{R}^2_{uv}.$$

Now, we use the area form $du \wedge dv$ to find the area of the parallelepiped spanned by these pushed-forward vectors,

$$(du \wedge dv)_{(12,-2)} \left(T_{(5,7)}f \cdot \begin{bmatrix} 1 \\ 0 \end{bmatrix}_{(5,7)}, T_{(5,7)}f \cdot \begin{bmatrix} 0 \\ 1 \end{bmatrix}_{(5,7)} \right) = \begin{vmatrix} 1 & 1 \\ 1 & -1 \end{vmatrix} = -1 - 1 = -2.$$

In summary, we have found for $p = (5, 7)$

$$(dx \wedge dy)_p \underbrace{\left(\begin{bmatrix} 1 \\ 0 \end{bmatrix}_p, \begin{bmatrix} 0 \\ 1 \end{bmatrix}_p \right)}_{\text{vectors in } T_p\mathbb{R}^2_{xy}} = 1$$

and

$$(du \wedge dv)_{f(p)} \underbrace{\left(T_pf \cdot \begin{bmatrix} 1 \\ 0 \end{bmatrix}_p, T_pf \cdot \begin{bmatrix} 0 \\ 1 \end{bmatrix}_p \right)}_{\substack{\text{push-forward of vectors in } T_p\mathbb{R}^2_{xy}, \\ \text{which are in } T_{f(p)}\mathbb{R}^2_{uv}}} = -2.$$

Combining everything we get the following equality

$$\underbrace{-2 \cdot \underbrace{(dx \wedge dy)_p \left(\begin{bmatrix} 1 \\ 0 \end{bmatrix}_p, \begin{bmatrix} 0 \\ 1 \end{bmatrix}_p \right)}_{=1}}_{=-2} = \underbrace{(du \wedge dv)_{f(p)} \left(T_pf \cdot \begin{bmatrix} 1 \\ 0 \end{bmatrix}_p, T_pf \cdot \begin{bmatrix} 0 \\ 1 \end{bmatrix}_p \right)}_{=-2}$$

which is exactly what we had from the previous section, only recognizing that the vectors that $du \wedge dv$ ate are the push-forwards of the vectors that $dx \wedge dy$ ate. So, in order to make rigorous sense of the identity $-2dx \wedge dy = du \wedge dv$ the concept that we need to work with is the push-forwards of vectors.

Question 6.5 Consider the same change in coordinates $f(x, y) = \big(u(x, y), v(x, y)\big) = (x + y, x - y)$.

(a) Find $f(3. - 2)$.

(b) Find $(dx \wedge dy)_{(3,-2)}\left(\begin{bmatrix} 1 \\ 1 \end{bmatrix}_{(3,-2)}, \begin{bmatrix} 5 \\ -1 \end{bmatrix}_{(3,-2)}\right)$.

(c) Find $T_{(3,-2)}f \cdot \begin{bmatrix} 1 \\ 1 \end{bmatrix}_{(3,-2)}$ and $T_{(3,-2)}f \cdot \begin{bmatrix} 5 \\ -1 \end{bmatrix}_{(3,-2)}$.

(d) Find $(du \wedge dv)_{f(3,-2)}\left(T_{(3,-2)}f \cdot \begin{bmatrix} 1 \\ 1 \end{bmatrix}_{(3,-2)}, T_{(3,-2)}f \cdot \begin{bmatrix} 5 \\ -1 \end{bmatrix}_{(3,-2)}\right)$.

(e) Do the answers match what you would expect from the last example and the relationship $dx \wedge dy = \frac{-1}{2} du \wedge dv$?

Now we briefly look at push-forwards of vectors in a somewhat more general setting instead of in the context of our particular problem. Consider a map between two manifolds, $f : M \to N$. For the moment we will restrict our attention to the case where $M = N = \mathbb{R}^n$ for some n. If the map $f = (f_1, f_2, \ldots, f_n)$ is differentiable at a point $p \in M$ (that means each f_i is differentiable at p) and the coordinates of manifold M are given by $x_1, x_2, \ldots, x_n$, then the derivative of the map at the point $p \in M$ is given by the traditional Jacobian matrix

$$T_p f = \text{Jacobian of } f \text{ at } p = \begin{bmatrix} \frac{\partial f_1}{\partial x_1}\Big|_p & \frac{\partial f_1}{\partial x_2}\Big|_p & \cdots & \frac{\partial f_1}{\partial x_n}\Big|_p \\ \frac{\partial f_2}{\partial x_1}\Big|_p & \frac{\partial f_2}{\partial x_2}\Big|_p & \cdots & \frac{\partial f_2}{\partial x_n}\Big|_p \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f_n}{\partial x_1}\Big|_p & \frac{\partial f_n}{\partial x_2}\Big|_p & \cdots & \frac{\partial f_n}{\partial x_n}\Big|_p \end{bmatrix}.$$

The mapping $T_p f$ given by the Jacobian matrix is simply the closest linear approximation of the map $f : M \to N$. We have already spend a lot of time considering real-valued functions $g : \mathbb{R}^n \to \mathbb{R}$ and the differential dg. We know that the differential of g at the point p is in fact the closest linear approximation of g at the point p. In fact, we have said that dg_p in a sense "encodes" the tangent plane of g at the point p, and of course the tangent plane of g at p is the closest linear approximation to the graph of g. With Cartesian coordinates we can write the differential of g as

$$dg = \frac{\partial g}{\partial x_1} dx_1 + \frac{\partial g}{\partial x_2} dx_2 + \cdots + \frac{\partial g}{\partial x_n} dx_n = \left[\frac{\partial g}{\partial x_1}, \frac{\partial g}{\partial x_2}, \cdots, \frac{\partial g}{\partial x_n}\right].$$

In the case of our function $f = (f_1, f_2, \ldots, f_n)$ each component function f_i is itself a real-valued function $f_i : \mathbb{R}^n \to \mathbb{R}$ and hence we could consider the differentials of each f_i separately,

$$\begin{aligned}
df_1 &= \frac{\partial f_1}{\partial x_1} dx_1 + \frac{\partial f_1}{\partial x_2} dx_2 + \cdots + \frac{\partial f_1}{\partial x_n} dx_n = \left[\frac{\partial f_1}{\partial x_1}, \frac{\partial f_1}{\partial x_2}, \cdots, \frac{\partial f_1}{\partial x_n}\right], \\
df_2 &= \frac{\partial f_2}{\partial x_1} dx_1 + \frac{\partial f_2}{\partial x_2} dx_2 + \cdots + \frac{\partial f_2}{\partial x_n} dx_n = \left[\frac{\partial f_2}{\partial x_1}, \frac{\partial f_2}{\partial x_2}, \cdots, \frac{\partial f_2}{\partial x_n}\right], \\
&\vdots \\
df_n &= \frac{\partial f_n}{\partial x_1} dx_1 + \frac{\partial f_n}{\partial x_2} dx_2 + \cdots + \frac{\partial f_n}{\partial x_n} dx_n = \left[\frac{\partial f_n}{\partial x_1}, \frac{\partial f_n}{\partial x_2}, \cdots, \frac{\partial f_n}{\partial x_n}\right].
\end{aligned}$$

Compare the co-vectors of the component functions of f with the rows of the Jacobian matrix. They are the same. Thus *the rows of the Jacobian matrix of f are nothing more than the differentials of the component functions of f, df_i, written as row-matrices.* The Jacobian Tf is the closest linear approximation of the mapping $f : \mathbb{R}^n \to \mathbb{R}^n$ because its rows are the closest linear approximations of each $f_i : \mathbb{R}^n \to \mathbb{R}$.

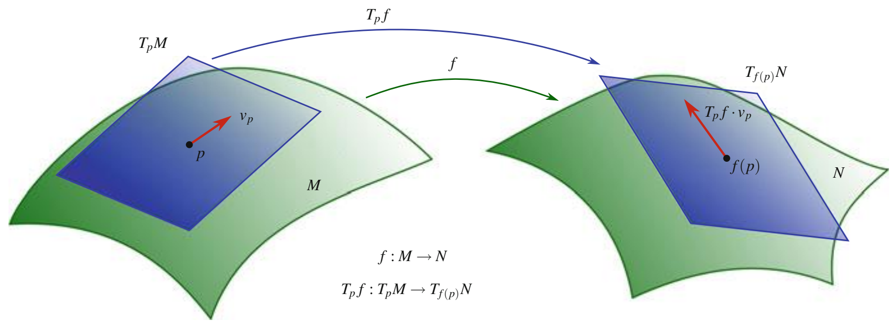


Fig. 6.7 The map $f : M \to N$ induces the tangent map $Tf : TM \to TN$. At a particular point $p \in M$ we get $T_p f : T_p M \to T_{f(p)} N$. Thus the vector $v_p \in T_p M$ gets pushed-forward to the vector $T_p f \cdot v_p \in T_{f(p)} N$

As before, suppose we have some vector $v_p \in T_p M$. $T_p f \cdot v_p$ gives us the linear approximation of the change in f as we move along v_p, which is another vector in $T_{f(p)} N$. Since we can find $T_p f : T_p M \to T_{f(p)} N$ at each point $p \in M$ where f is differentiable we end up with a mapping Tf from the tangent bundle of M to the tangent bundle of N. We say that the map $f : M \to N$ *induces* the map $Tf : TM \longrightarrow TN$. The mapping Tf **pushes-forward** vectors in $T_p M$ to vectors in $T_{f(p)} N$. This mapping is often called the tangent mapping induced by f. See Fig. 6.7. Often we would show this by writing something like this,

$$\begin{aligned}
TM &\xrightarrow{\substack{Tf\\ f_*\\ Df}} TN\\
v &\longmapsto Tf \cdot v\\
M &\xrightarrow{\ f\ } N\\
p &\longmapsto f(p).
\end{aligned}$$

There are three different notations you will encounter for this induced map, $D_p f$, $T_p f$, and $f_*(p)$. $D_p f$ is a pretty standard notation which you will frequently encounter, especially in calculus classes. Probably the most common notation, however, is f_*. If we want to include the base point it is often written as $f_*(p)$, which is a little cumbersome. This notation is most frequently encountered in differential geometry books. However, in general we will use the notation Tf or $T_p f$. Frankly, this is a fairly non-standard notation but we like it because as a notation it packs in a lot of information. The T in Tf tells us that we are dealing with the tangent map going from one tangent bundle to another, induced by the mapping f. The notation also can include the base point p nicely when we want it to in the same manner the base point is included in the tangent space notation. Thus $T_p f$ is used to specify a mapping from the tangent space $T_p M$ to the tangent space $T_{f(p)} N$.

6.3 Pull-Backs of Volume Forms

In this section we will define the **pull-backs** of volume forms. While most differential forms are not volume forms, the volume forms play a very important role in integration so we will concentrate on them in this section. We will explore the pull-backs of more general differential forms later. For the moment we will continue with our linear example. Our basic goal is to hone in on a more precise understanding of the relationship

$$\underbrace{-2 \cdot \underbrace{(dx \wedge dy)_p \left(\underbrace{\begin{bmatrix} 1 \\ 0 \end{bmatrix}_p, \begin{bmatrix} 0 \\ 1 \end{bmatrix}_p}_{\text{vectors in } T_p\mathbb{R}^2_{xy}} \right)}_{=1}}_{=-2} = \underbrace{(du \wedge dv)_{f(p)} \left(\underbrace{T_p f \cdot \begin{bmatrix} 1 \\ 0 \end{bmatrix}_p, T_p f \cdot \begin{bmatrix} 0 \\ 1 \end{bmatrix}_p}_{\substack{\text{push-forward of vectors in } T_p\mathbb{R}^2_{xy},\\ \text{which are in } T_{f(p)}\mathbb{R}^2_{uv}}} \right)}_{=-2}.$$

Recall, for the mapping $f(x, y) = (x + y, x - y)$ we had

$$\begin{aligned} du \wedge dv &= d(x+y) \wedge d(x-y) \\ &= (dx + dy) \wedge (dx - dy) \\ &= dx \wedge (dx - dy) + dy \wedge (dx - dy) \\ &= dx \wedge dx - dx \wedge dy + dy \wedge dx - dy \wedge dy \\ &= -dx \wedge dy - dx \wedge dy \\ &= -2dx \wedge dy. \end{aligned}$$

We reperform this computation in generality. That is, consider a change of variables $f : \mathbb{R}^2_{xy} \to \mathbb{R}^2_{uv}$ given by $f(x, y) = \big(u(x, y), v(x, y)\big) = (u, v)$. First we find

$$du = \frac{\partial u}{\partial x}dx + \frac{\partial u}{\partial y}dy \quad \text{and} \quad dv = \frac{\partial v}{\partial x}dx + \frac{\partial v}{\partial y}dy,$$

and then using the properties of the wedgeproduct we get

$$\begin{aligned} du \wedge dv &= \left(\frac{\partial u}{\partial x}dx + \frac{\partial u}{\partial y}dy\right) \wedge \left(\frac{\partial v}{\partial x}dx + \frac{\partial v}{\partial y}dy\right) \\ &= \frac{\partial u}{\partial x}dx \wedge \left(\frac{\partial v}{\partial x}dx + \frac{\partial v}{\partial y}dy\right) + \frac{\partial u}{\partial y}dy \wedge \left(\frac{\partial v}{\partial x}dx + \frac{\partial v}{\partial y}dy\right) \\ &= \frac{\partial u}{\partial x}\frac{\partial v}{\partial x}\underbrace{dx \wedge dx}_{=0} + \frac{\partial u}{\partial x}\frac{\partial v}{\partial y}dx \wedge dy + \frac{\partial u}{\partial y}\frac{\partial v}{\partial x}\underbrace{dy \wedge dx}_{-dx\wedge dy} + \frac{\partial u}{\partial y}\frac{\partial v}{\partial y}\underbrace{dy \wedge dy}_{=0} \\ &= \frac{\partial u}{\partial x}\frac{\partial v}{\partial y}dx \wedge dy - \frac{\partial u}{\partial y}\frac{\partial v}{\partial x}dx \wedge dy \\ &= \left(\frac{\partial u}{\partial x}\frac{\partial v}{\partial y} - \frac{\partial u}{\partial y}\frac{\partial v}{\partial x}\right)dx \wedge dy \\ &= \underbrace{\begin{vmatrix} \frac{\partial u}{\partial x} & \frac{\partial u}{\partial y} \\ \frac{\partial v}{\partial x} & \frac{\partial v}{\partial y} \end{vmatrix}}_{\substack{\text{Determinant of} \\ \text{Jacobian} \\ \text{matrix}}} dx \wedge dy. \end{aligned}$$

In essence it appears that the Jacobian matrix just “falls out” of the wedgeproduct when we are finding the relationship between two volume forms. In the last section we tried to get a better idea what this identity meant by introducing the induced tangent map Tf and the idea of push-forwards of vectors. The induced tangent map Tf operated according to

$$\begin{aligned} T_p\mathbb{R}^2 &\xrightarrow{T_p f} T_{f(p)}\mathbb{R}^2 \\ \mathbb{R}^2 &\xrightarrow{f} \mathbb{R}^2, \end{aligned}$$

where the tangent map at p, $T_p f$, is given by the Jacobian matrix evaluated at p,

$$T_p f = \begin{bmatrix} \frac{\partial u}{\partial x} & \frac{\partial u}{\partial y} \\ \frac{\partial v}{\partial x} & \frac{\partial v}{\partial y} \end{bmatrix}_p .$$

We defined the push-forward of the vector $v_p = \begin{bmatrix} v_1 \\ v_2 \end{bmatrix}_p \in T_p\mathbb{R}^2$ by

$$T_p f \cdot v_p = \begin{bmatrix} \frac{\partial u}{\partial x} & \frac{\partial u}{\partial y} \\ \frac{\partial v}{\partial x} & \frac{\partial v}{\partial y} \end{bmatrix}_p \cdot \begin{bmatrix} v_1 \\ v_2 \end{bmatrix}_p = \begin{bmatrix} \frac{\partial u}{\partial x} v_1 + \frac{\partial u}{\partial y} v_2 \\ \frac{\partial v}{\partial x} v_1 + \frac{\partial v}{\partial y} v_2 \end{bmatrix}_{f(p)} \in T_{f(p)}\mathbb{R}^2.$$

Omitting the base point p from our calculations and using det to represent the determinant where convenient, we have

$$\begin{aligned} du \wedge dv(Tf \cdot v, Tf \cdot w) &= \det \begin{bmatrix} du(Tf \cdot v) & du(Tf \cdot w) \\ dv(Tf \cdot v) & dv(Tf \cdot w) \end{bmatrix} \\ &= \det \begin{bmatrix} \frac{\partial u}{\partial x} v_1 + \frac{\partial u}{\partial y} v_2 & \frac{\partial u}{\partial x} w_1 + \frac{\partial u}{\partial y} w_2 \\ \frac{\partial v}{\partial x} v_1 + \frac{\partial v}{\partial y} v_2 & \frac{\partial v}{\partial x} w_1 + \frac{\partial v}{\partial y} w_2 \end{bmatrix} \\ &= \det \left(\begin{bmatrix} \frac{\partial u}{\partial x} & \frac{\partial u}{\partial y} \\ \frac{\partial v}{\partial x} & \frac{\partial v}{\partial y} \end{bmatrix} \begin{bmatrix} v_1 & w_1 \\ v_2 & w_2 \end{bmatrix} \right) \\ &= \det \begin{bmatrix} \frac{\partial u}{\partial x} & \frac{\partial u}{\partial y} \\ \frac{\partial v}{\partial x} & \frac{\partial v}{\partial y} \end{bmatrix} \cdot \det \begin{bmatrix} v_1 & w_1 \\ v_2 & w_2 \end{bmatrix} \\ &= \begin{vmatrix} \frac{\partial u}{\partial x} & \frac{\partial u}{\partial y} \\ \frac{\partial v}{\partial x} & \frac{\partial v}{\partial y} \end{vmatrix} \cdot \begin{vmatrix} dx(v) & dx(w) \\ dy(v) & dy(w) \end{vmatrix} \\ &= \begin{vmatrix} \frac{\partial u}{\partial x} & \frac{\partial u}{\partial y} \\ \frac{\partial v}{\partial x} & \frac{\partial v}{\partial y} \end{vmatrix} dx \wedge dy(v, w) \end{aligned}$$

giving us the identity

$$\begin{vmatrix} \frac{\partial u}{\partial x} & \frac{\partial u}{\partial y} \\ \frac{\partial v}{\partial x} & \frac{\partial v}{\partial y} \end{vmatrix} dx \wedge dy(v, w) = du \wedge dv(Tf \cdot v, Tf \cdot w).$$

Again, the Jacobian essentially "fell out" when we computed how the area form changes under a change of basis given by f. But also, the Jacobian is exactly our Tf that pushes vectors forward when we change basis, and vectors (or push-forwards of vectors) are obviously necessary to finding the areas of the parallelepiped spanned by the vectors. It is almost as if the Jacobian matrix "pulls through" in some sense.

Now we will define what the **pull-back** of a differential form is. Keep in mind, in this section the only differential form we have actually looked at is the volume form because of its importance to integration. But this basic definition actually applies to all differential forms. Suppose we have a mapping between manifolds $f : M \to N$. This mapping induces the tangent mapping $Tf : TM \to TN$. Now suppose that we have a differential k-form ω on N. Then we define the differential k-form $T^*f \cdot \omega$ as the pull-back of ω to M where

Definition of Pull-Back of Differential Form	$\big(T^*f \cdot \omega\big)\big(v_1, v_2, \ldots, v_k\big) = \omega\big(Tf \cdot v_1, Tf \cdot v_2, \ldots, Tf \cdot v_k\big).$

What exactly is this trying to say? If we are given a k-form ω on N, the range of f, then we are trying to find a k-form on M that is in some sense the "same" as ω. What do we mean by "the same"? The vectors that these two-forms will eat are from different tangent spaces. But we already know that tangent vectors on M can be pushed forward in a natural way to tangent vectors on N, so we decide to use this idea. When the k-form on M that we are trying to find eats k tangent vectors we want it to give the same value that the k-form ω on N gives when it eats the push-forwards of these same vectors. This k-form on M that we have found is denoted by $T^*f \cdot \omega$ or $f^*\omega$ and called the pull-back of ω. The most common notation in use is f^*,

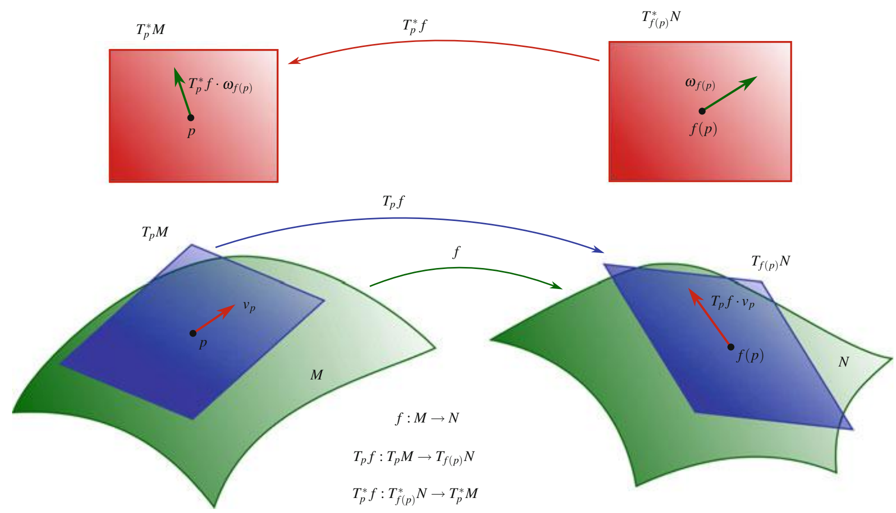


Fig. 6.8 The mappings f and Tf are shown as in Fig. 6.7, but now the pull-back mapping $T^*f : T^*N \to T^*M$ is included. At a particular point $f(p) \in N$ we have $T_p^*f : T^*_{f(p)}N \to T_p^*M$. Notice that the mapping T_p^*f is actually indexed by p, the point of the range. This is quite unusual, but doing so helps to keep the notation consistent

but for the same reasons as explained in the section on push-forwards of tangent vectors, we will use T^*f. The pull-back map is also sometimes called the **cotangent map**.

Here we show all the mappings together with base point notation included. See Fig. 6.8. There is one very important thing to notice, that when it comes to the pull-back map, *it is indexed by the base point in the image and not the domain!* This is quite unusual in mathematics, but in this situation it ends up simply being the best way to keep track of base points. We often write the mapping f and its induced mappings $T_p f$ and $T_p^* f$ as

$$
\begin{aligned}
{\bigwedge}_p^k(M) &\xleftarrow{T_p^* f} {\bigwedge}_{f(p)}^k(N) \\
T_p^* f \cdot \omega_{f(p)} &\longleftarrow\!\shortmid\ \omega_{f(p)} \\
T_p M &\xrightarrow{T_p f} T_{f(p)} N \\
v_p &\longmapsto T_{f(p)} f \cdot v_p \\
M &\xrightarrow{f} N \\
p &\longmapsto f(p).
\end{aligned}
$$

Now we are ready to understand what is going on in our example. Here everything is put together,

$$
\underbrace{\overbrace{\begin{vmatrix} 1 & 1 \\ 1 & -1 \end{vmatrix} \cdot (dx \wedge dy)_p}^{\substack{T^*f\cdot(du\wedge dv),\\ \text{pull-back}\\ \text{of } du\wedge dv}} \underbrace{\left(\begin{bmatrix} 1 \\ 0 \end{bmatrix}_p, \begin{bmatrix} 0 \\ 1 \end{bmatrix}_p \right)}_{\text{vectors in } T_p\mathbb{R}^2_{xy}}}_{=(-2)(1)=-2} = \underbrace{(du \wedge dv)_{f(p)} \underbrace{\left(T_p f \cdot \begin{bmatrix} 1 \\ 0 \end{bmatrix}_p, T_p f \cdot \begin{bmatrix} 0 \\ 1 \end{bmatrix}_p \right)}_{\substack{\text{push-forward of vectors in } T_p\mathbb{R}^2_{xy},\\ \text{which are in } T_{f(p)}\mathbb{R}^2_{uv}}}}_{=-2}.
$$

The *real* way to write the identity $-2dx \wedge dy = du \wedge dv$ is that the two-form $-2dx \wedge dy$ is the pull-back of $du \wedge dv$,

$$
T^*f \cdot \big(du \wedge dv\big) = -2dx \wedge dy.
$$

As we said earlier, in this section we are primarily concentrating on volume forms because of their role in integration. Now we will consider the pull-backs of volume in full generality. We want to find the pull-back of the volume form under a change of coordinate mapping $\phi : \mathbb{R}^n_{(x_1,\ldots,x_n)} \to \mathbb{R}^n_{(\phi_1,\ldots,\phi_n)}$ given by

$$\phi = \Big(\phi_1(x_1,\ldots,x_n), \phi_2(x_1,\ldots,x_n), \ldots, \phi_n(x_1,\ldots,x_n)\Big).$$

Now consider the volume one-form $\omega = d\phi_1 \wedge d\phi_2 \wedge \cdots \wedge d\phi_n$ on $\mathbb{R}^n_{(\phi_1,\ldots,\phi_n)}$. We want to find a general formula for $T^*\phi \cdot \omega$ on $\mathbb{R}^n_{(x_1,\ldots,x_n)}$. Clearly, we know that $T^*\phi \cdot \omega$ is going to be an n-form as well and hence will have to have the form $f\, dx_1 \wedge dx_2 \wedge \cdots \wedge dx_n$ for some as of yet unknown function f. Letting

$$e_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \quad e_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \quad \ldots, \quad e_n = \begin{bmatrix} 0 \\ \vdots \\ 0 \\ 1 \end{bmatrix}$$

we have

$$dx_1 \wedge dx_2 \wedge \cdots \wedge dx_n(e_1, e_2, \ldots, e_n) = 1.$$

Also, recalling that

$$T\phi = \begin{bmatrix} \frac{\partial \phi_1}{\partial x_1} & \frac{\partial \phi_1}{\partial x_2} & \cdots & \frac{\partial \phi_1}{\partial x_n} \\ \frac{\partial \phi_2}{\partial x_1} & \frac{\partial \phi_2}{\partial x_2} & \cdots & \frac{\partial \phi_2}{\partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial \phi_n}{\partial x_1} & \frac{\partial \phi_n}{\partial x_2} & \cdots & \frac{\partial \phi_n}{\partial x_n} \end{bmatrix}$$

we have

$$T\phi \cdot e_1 = \begin{bmatrix} \frac{\partial \phi_1}{\partial x_1} \\ \frac{\partial \phi_2}{\partial x_1} \\ \vdots \\ \frac{\partial \phi_n}{\partial x_1} \end{bmatrix}, \quad T\phi \cdot e_2 = \begin{bmatrix} \frac{\partial \phi_1}{\partial x_2} \\ \frac{\partial \phi_2}{\partial x_2} \\ \vdots \\ \frac{\partial \phi_n}{\partial x_2} \end{bmatrix}, \quad \ldots, \quad T\phi \cdot e_n = \begin{bmatrix} \frac{\partial \phi_1}{\partial x_n} \\ \frac{\partial \phi_2}{\partial x_n} \\ \vdots \\ \frac{\partial \phi_n}{\partial x_n} \end{bmatrix}$$

so we get

$$\begin{aligned} f &= \big(f\, dx_1 \wedge dx_2 \wedge \cdots \wedge dx_n\big)\big(e_1, e_2, \ldots, e_n\big) \\ &= \Big(T^*\phi \cdot \omega\Big)\big(e_1, e_2, \ldots, e_n\big) \\ &= \omega\Big(T\phi \cdot e_1, T\phi \cdot e_2, \ldots, T\phi \cdot e_n\Big) \\ &= d\phi_1 \wedge d\phi_2 \wedge \cdots \wedge d\phi_n\Big(T\phi \cdot e_1, T\phi \cdot e_2, \ldots, T\phi \cdot e_n\Big) \end{aligned}$$

$$= \begin{vmatrix} d\phi_1(T\phi \cdot e_1) & d\phi_1(T\phi \cdot e_2) & \cdots & d\phi_1(T\phi \cdot e_n) \\ d\phi_2(T\phi \cdot e_1) & d\phi_2(T\phi \cdot e_2) & \cdots & d\phi_2(T\phi \cdot e_n) \\ \vdots & \vdots & \ddots & \vdots \\ d\phi_n(T\phi \cdot e_1) & d\phi_n(T\phi \cdot e_2) & \cdots & d\phi_n(T\phi \cdot e_n) \end{vmatrix}$$

$$= \begin{vmatrix} \frac{\partial \phi_1}{\partial x_1} & \frac{\partial \phi_1}{\partial x_2} & \cdots & \frac{\partial \phi_1}{\partial x_n} \\ \frac{\partial \phi_2}{\partial x_1} & \frac{\partial \phi_2}{\partial x_2} & \cdots & \frac{\partial \phi_2}{\partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial \phi_n}{\partial x_1} & \frac{\partial \phi_n}{\partial x_2} & \cdots & \frac{\partial \phi_n}{\partial x_n} \end{vmatrix}.$$

Combining everything we get the general formula for the pull-back of a volume form by a change in basis $\phi : \mathbb{R}^n_{(x_1,\ldots,x_n)} \to \mathbb{R}^n_{(\phi_1,\ldots,\phi_n)}$ to be

Pull-back of Volume Form

$$\boxed{T^*\phi \cdot \Big(d\phi_1 \wedge d\phi_2 \wedge \cdots \wedge d\phi_n\Big) = \begin{vmatrix} \frac{\partial \phi_1}{\partial x_1} & \frac{\partial \phi_1}{\partial x_2} & \cdots & \frac{\partial \phi_1}{\partial x_n} \\ \frac{\partial \phi_2}{\partial x_1} & \frac{\partial \phi_2}{\partial x_2} & \cdots & \frac{\partial \phi_2}{\partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial \phi_n}{\partial x_1} & \frac{\partial \phi_n}{\partial x_2} & \cdots & \frac{\partial \phi_n}{\partial x_n} \end{vmatrix} dx_1 \wedge dx_2 \wedge \cdots \wedge dx_n.}$$

At least when it comes to maps between manifolds of the same dimension, the pull-back of a volume form has a very nice representation in terms of the change of coordinate mapping ϕ. In essence, the Jacobian used in the push-forward of vector fields, $T\phi$, pulls through the volume form computation. However, we caution you that for more general k-forms there is no nice formula for finding pull-backs. This exceptionally wonderful formula involving the Jacobian matrix for the pull-back of volume forms *only applies to volume forms and to mappings between manifolds with the same dimension.*

6.4 Polar Coordinates

First we will explore the polar coordinate transformations $f : \mathbb{R}^2_{r\theta} \to \mathbb{R}^2_{xy}$ given by

$$f(r,\theta) = \big(x(r,\theta), y(r,\theta)\big) = \big(r\cos\theta, r\sin\theta\big)$$

which is often written in calculus textbooks as

$$x = r\cos\theta,$$
$$y = r\sin\theta.$$

The inverse transformation $f^{-1} : \mathbb{R}^2_{xy} \to \mathbb{R}^2_{r\theta}$ is given by the rather uglier

$$f^{-1}(x,y) = \big(r(x,y), \theta(x,y)\big) = \left(\pm\sqrt{x^2+y^2}, \arctan\left(\frac{y}{x}\right)\right),$$

where the $+$ is chosen if $x \geq 0$ and the $-$ is chosen if $x < 0$, and θ is restricted to $-\frac{\pi}{2} < \theta < \frac{\pi}{2}$, the interval where arctan is defined. This restriction also makes the original function one-to-one. This can also be written as

$$r = \pm\sqrt{x^2+y^2},$$
$$\theta = \arctan\left(\frac{y}{x}\right)$$

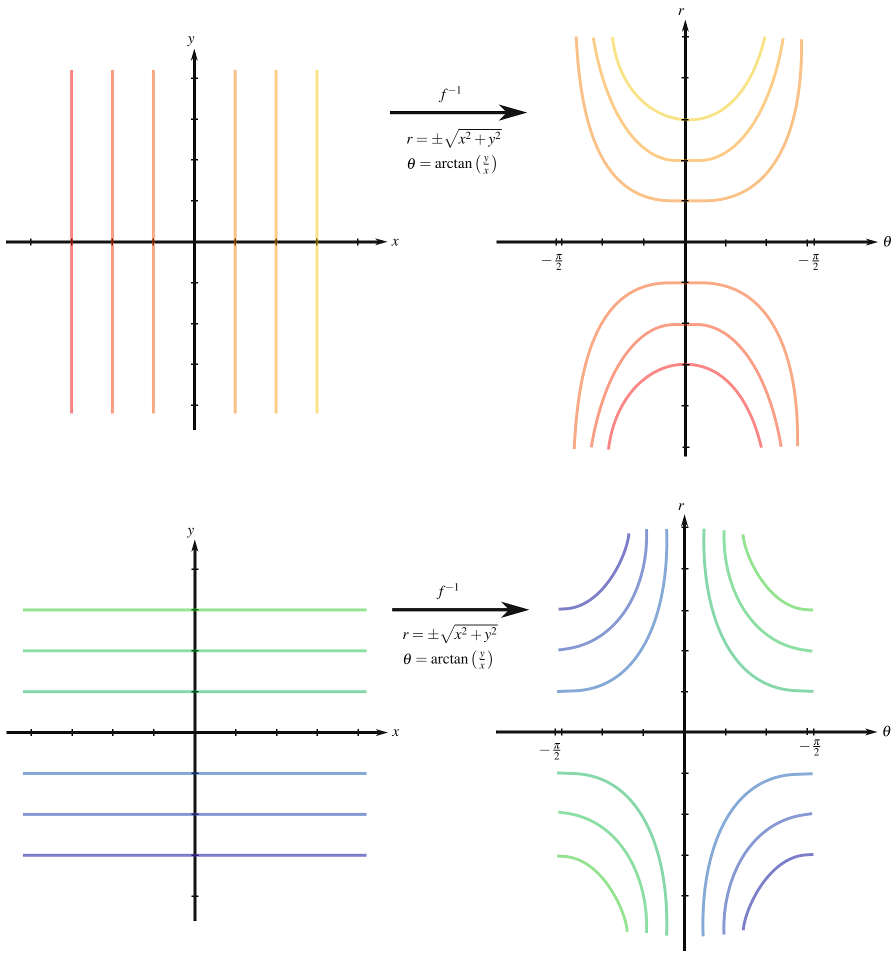


Fig. 6.9 The polar coordinate transformation $f^{-1} : \mathbb{R}^2_{xy} \to \mathbb{R}^2_{r\theta}$ restricted to $-\pi/2 < \theta \leq \pi/2$. This restriction is necessary to make f^{-1} one-to-one. The image of several vertical lines is shown on the top and the image of several horizontal lines is shown on the bottom

and is shown in Fig. 6.9.

Question 6.6 Show that $f^{-1}(x, y)$ is indeed given by $f^{-1}(x, y) = \left(\pm\sqrt{x^2 + y^2}, \arctan\left(\frac{y}{x}\right)\right)$ where the $+$ is chosen if $x \geq 0$ and the $-$ is chosen if $x < 0$, and θ is restricted to $-\frac{\pi}{2} < \theta < \frac{\pi}{2}$.

Question 6.7 Similar to Fig. 6.9, sketch the mapping $f : \mathbb{R}^2_{r\theta} \to \mathbb{R}^2_{xy}$. Where do the lines $r = c$, where c is a constant, map to? Where do the lines $\theta = c$ map to? Compare this with what you learned in your calculus class about polar coordinates.

Actually, it is pretty easy to see how the horizontal and vertical lines on $\mathbb{R}^2_{xy}$ in Fig. 6.9 map to $\mathbb{R}^2_{r\theta}$, we just have to write the equations $x = c$ and $y = c$, where c is a constant, in polar coordinates,

$$x = c \;\Rightarrow\; c = r\cos\theta \;\Rightarrow\; r = \frac{c}{\cos\theta} = c\sec\theta$$

and the line $y = c$ can be written as

$$y = c \;\Rightarrow\; c = r\cos\theta \;\Rightarrow\; r = \frac{c}{\sin\theta} = c\csc\theta,$$

where $\frac{-\pi}{2} < \theta < \frac{\pi}{2}$. For example, the line $x = 1$ can be written as $r = \sec\theta$, the line $x = 2$ is written as $r = 2\sec\theta$, etc.

The first thing that we will do in polar coordinates is "write the area form $dx \wedge dy$ in terms of $d\theta \wedge dr$." Using the mapping $f(r,\theta) = \big(x(r,\theta), y(r,\theta)\big) = \big(r\cos\theta, r\sin\theta\big)$ find dx and dy,

$$
\begin{aligned}
dx &= \frac{\partial x}{\partial r}dr + \frac{\partial x}{\partial \theta}d\theta \\
&= \frac{\partial (r\cos\theta)}{\partial r}dr + \frac{\partial (r\cos\theta)}{\partial \theta}d\theta \\
&= \cos\theta dr - r\sin\theta d\theta
\end{aligned}
$$

and

$$
\begin{aligned}
dy &= \frac{\partial y}{\partial r}dr + \frac{\partial x}{\partial \theta}d\theta \\
&= \frac{\partial (r\sin\theta)}{\partial r}dr + \frac{\partial (r\sin\theta)}{\partial \theta}d\theta \\
&= \sin\theta dr + r\cos\theta d\theta
\end{aligned}
$$

so

$$
\begin{aligned}
dx \wedge dy &= \Big(\cos\theta dr - r\sin\theta d\theta\Big) \wedge \Big(\sin\theta dr + r\cos\theta d\theta\Big) \\
&= \cos\theta\sin\theta \underbrace{dr \wedge dr}_{=0} - r\sin^2\theta d\theta \wedge dr \\
&\quad + r\cos^2\theta \underbrace{dr \wedge d\theta}_{-d\theta\wedge dr} - r^2\sin\theta\cos\theta \underbrace{d\theta \wedge d\theta}_{=0} \\
&= -r(\sin^2\theta + \cos^2\theta) d\theta \wedge dr \\
&= -r d\theta \wedge dr.
\end{aligned}
$$

Thus we get the "identity" $dx \wedge dy = -rd\theta \wedge dr$. Now that we have "written the area form $dx \wedge dy$ in terms of $d\theta \wedge dr$" let us analyze what we have actually done. We were given a mapping $f : \mathbb{R}^2_{r\theta} \to \mathbb{R}^2_{xy}$, which we learned induces the maps Tf and T^*f that behave as follows:

$$
\begin{gathered}
\bigwedge^2\left(\mathbb{R}^2_{r\theta}\right) \xleftarrow{T^*f} \bigwedge^2\left(\mathbb{R}^2_{xy}\right) \\
T^*f \cdot \big(dx \wedge dy\big) \longleftarrow dx \wedge dy \\
T\mathbb{R}^2_{r\theta} \xrightarrow{Tf} T\mathbb{R}^2_{xy} \\
\mathbb{R}^2_{r\theta} \xrightarrow{f} \mathbb{R}^2_{xy}.
\end{gathered}
$$

The induced map T^*f finds the pull-back by f of the volume form on $\mathbb{R}^2_{xy}$. Thus, when we say that we have "written the area form $dx \wedge dy$ in terms of $d\theta \wedge dr$" what we have actually done is find the pull-back. Thus, the identity $dx \wedge dy = -rd\theta \wedge dr$ should actually be written as

$$
\boxed{T^*f \cdot \big(dx \wedge dy\big) = -rd\theta \wedge dr.}
$$

Question 6.8 Referring to Fig. 6.10 explain the formula $T^*f \cdot \big(dx \wedge dy\big) = -rd\theta \wedge dr$ in terms of the shaded regions.

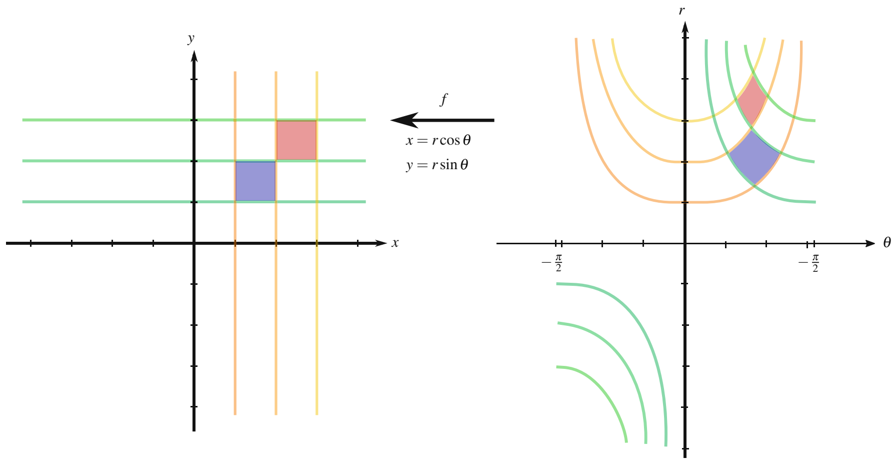


Fig. 6.10 The mapping $f : \mathbb{R}^2_{r\theta} \to \mathbb{R}^2_{xy}$. Notice what happens to the shaded regions

Similarly, we could have worked in the other direction to "write the area form $d\theta \wedge dr$ in terms of $dx \wedge dy$" using $f^{-1}(x, y) = \big(r(x, y), \theta(x, y)\big) = \big(\pm\sqrt{x^2 + y^2}, \arctan\big(\frac{y}{x}\big)\big)$. First we need to find $d\theta$ and dr,

$$\begin{aligned}
d\theta &= \frac{\partial\theta}{\partial x}dx + \frac{\partial\theta}{\partial y}dy \\
&= \frac{\partial \arctan(y/x)}{\partial x}dx + \frac{\partial \arctan(y/x)}{\partial y}dy \\
&\underset{\substack{\text{chain}\\\text{rule}}}{=} \frac{1}{1 + \left(\frac{y}{x}\right)^2}\,\frac{\partial}{\partial x}\left(\frac{y}{x}\right)dx + \frac{1}{1 + \left(\frac{y}{x}\right)^2}\,\frac{\partial}{\partial y}\left(\frac{y}{x}\right)dy \\
&= \frac{-y}{x^2 + y^2}dx + \frac{x}{x^2 + y^2}dy
\end{aligned}$$

and

$$\begin{aligned}
dr &= \frac{\partial r}{\partial x}dx + \frac{\partial r}{\partial y}dy \\
&= \frac{\partial\sqrt{x^2 + y^2}}{\partial x}dx + \frac{\partial\sqrt{x^2 + y^2}}{\partial y}dy \\
&\underset{\substack{\text{chain}\\\text{rule}}}{=} \frac{x}{\sqrt{x^2 + y^2}}dx + \frac{y}{\sqrt{x^2 + y^2}}dy
\end{aligned}$$

so

$$\begin{aligned}
d\theta \wedge dr &= \left(\frac{-y}{x^2 + y^2}dx + \frac{x}{x^2 + y^2}dy\right) \wedge \left(\frac{x}{\sqrt{x^2 + y^2}}dx + \frac{y}{\sqrt{x^2 + y^2}}dy\right) \\
&= \frac{x^2}{(x^2 + y^2)\sqrt{x^2 + y^2}}\,dy \wedge dx - \frac{y^2}{(x^2 + y^2)\sqrt{x^2 + y^2}}\,dx \wedge dy \\
&= \frac{-(x^2 + y^2)}{(x^2 + y^2)\sqrt{x^2 + y^2}}\,dx \wedge dy \\
&= \frac{-1}{\sqrt{x^2 + y^2}}\,dx \wedge dy.
\end{aligned}$$

Thus we have obtained the identity $d\theta \wedge dr = \frac{-1}{\sqrt{x^2+y^2}} dx \wedge dy$. Like before, we had been given a mapping $f^{-1} : \mathbb{R}^2_{xy} \to \mathbb{R}^2_{r\theta}$, which induced mappings Tf^{-1} and T^*f^{-1} such that

$$\begin{aligned} \bigwedge^2 \left(\mathbb{R}^2_{xy}\right) &\xleftarrow{T^*f^{-1}} \bigwedge^2 \left(\mathbb{R}^2_{r\theta}\right) \\ T^*f \cdot \big(d\theta \wedge dr\big) &\longleftarrow\!\shortmid d\theta \wedge dr \\ T\mathbb{R}^2_{xy} &\xrightarrow{Tf^{-1}} T\mathbb{R}^2_{r\theta} \\ \mathbb{R}^2_{xy} &\xrightarrow{f^{-1}} \mathbb{R}^2_{r\theta}. \end{aligned}$$

So when we said "write the area form $d\theta \wedge dr$ in terms of $dx \wedge dy$" we really mean find the pull-back. The identity $d\theta \wedge dr = \frac{-1}{\sqrt{x^2+y^2}} dx \wedge dy$ that we obtained should actually be written as

$$\boxed{T^*f^{-1} \cdot \big(d\theta \wedge dr\big) = \frac{-1}{\sqrt{x^2+y^2}} dx \wedge dy.}$$

Since we are actually dealing with volume forms here, we could have done all of this by a different method, by using the general formula for pull-backs of volume forms derived in the last section. That is, given a mapping $\phi : \mathbb{R}^n \to \mathbb{R}^n$ such that

$$\phi = \Big(\phi_1(x_1, \ldots, x_n), \phi_2(x_1, \ldots, x_n), \ldots, \phi_n(x_1, \ldots, x_n)\Big)$$

the pull-back of the volume form is given by the following equation

$$T^*\phi \cdot \Big(d\phi_1 \wedge d\phi_2 \wedge \cdots \wedge d\phi_n\Big) = \begin{vmatrix} \frac{\partial\phi_1}{\partial x_1} & \frac{\partial\phi_1}{\partial x_2} & \cdots & \frac{\partial\phi_1}{\partial x_n} \\ \frac{\partial\phi_2}{\partial x_1} & \frac{\partial\phi_2}{\partial x_2} & \cdots & \frac{\partial\phi_2}{\partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial\phi_n}{\partial x_1} & \frac{\partial\phi_n}{\partial x_2} & \cdots & \frac{\partial\phi_n}{\partial x_n} \end{vmatrix} dx_1 \wedge dx_2 \wedge \cdots \wedge dx_n.$$

Recall, volume forms are special. While the Jacobian matrix always gives $T\phi$ and so is always used to find the push-forwards of vectors, the Jacobian matrix can only be used to find the pull-back of volume forms, not other differential forms that are not volume forms.

First we find the Jacobian matrix of the mapping $f^{-1}(x, y) = \left(\arctan\left(\frac{y}{x}\right), \pm\sqrt{x^2+y^2}\right)$ to be

$$\begin{bmatrix} \frac{\partial\theta}{\partial x} & \frac{\partial\theta}{\partial y} \\ \frac{\partial r}{\partial x} & \frac{\partial r}{\partial y} \end{bmatrix} = \begin{bmatrix} \frac{-y}{x^2+y^2} & \frac{x}{x^2+y^2} \\ \frac{x}{\sqrt{x^2+y^2}} & \frac{y}{\sqrt{x^2+y^2}} \end{bmatrix}$$

which gives the determinant

$$\begin{vmatrix} \frac{-y}{x^2+y^2} & \frac{x}{x^2+y^2} \\ \frac{x}{\sqrt{x^2+y^2}} & \frac{y}{\sqrt{x^2+y^2}} \end{vmatrix} = \frac{-y^2-x^2}{(x^2+y^2)^{3/2}} = \frac{-1}{\sqrt{x^2+y^2}}$$

which gives

$$T^*f^{-1} \cdot \big(d\theta \wedge dr\big) = \frac{-1}{\sqrt{x^2+y^2}} dx \wedge dy$$

DIRECTLY. The other direction is done similarly,

$$\begin{vmatrix} \frac{\partial x}{\partial \theta} & \frac{\partial x}{\partial r} \\ \frac{\partial y}{\partial \theta} & \frac{\partial y}{\partial r} \end{vmatrix}_{(\theta,r)} = \begin{vmatrix} -r\sin\theta & \cos\theta \\ r\cos\theta & \sin\theta \end{vmatrix}_{(\theta,r)} = -r\sin^2\theta - r\cos^2\theta = -r$$

which gives

$$T^*f \cdot \big(dx \wedge dy\big) = -r d\theta \wedge dr.$$

Now we will work through an example for polar coordinates. Through this example and the following question we will explore one of the ways that changes between Cartesian and polar coordinates are quite different from the linear example we were looking at in earlier sections. Where the polar coordinate change differs from a linear coordinate change is that the relation between the two volume forms depends on what point we are at. That is, the pull-back equation depends on the base point.

For the polar coordinate transformations $f^{-1} : \mathbb{R}^2_{xy} \to \mathbb{R}^2_{r\theta}$ given by $f^{-1}(x, y) = \left(\arctan\left(\frac{y}{x}\right), \pm\sqrt{x^2+y^2}\right)$ we will give θ in radians. Our goal is to relate

$$(dx \wedge dy)_{(1,1)} \left(\begin{bmatrix} 1 \\ 0 \end{bmatrix}_{(1,1)}, \begin{bmatrix} 0 \\ 1 \end{bmatrix}_{(1,1)} \right)$$

and

$$(d\theta \wedge dr)_{f^{-1}(1,1)} \left(T_{(1,1)} f^{-1} \cdot \begin{bmatrix} 1 \\ 0 \end{bmatrix}_{(1,1)}, T_{(1,1)} f^{-1} \cdot \begin{bmatrix} 0 \\ 1 \end{bmatrix}_{(1,1)} \right).$$

The first identity is simple, we get

$$(dx \wedge dy)_{(1,1)} \left(\begin{bmatrix} 1 \\ 0 \end{bmatrix}_{(1,1)}, \begin{bmatrix} 0 \\ 1 \end{bmatrix}_{(1,1)} \right) = \begin{vmatrix} 1 & 0 \\ 0 & 1 \end{vmatrix} = 1.$$

Here we used the xy-plane volume form $dx \wedge dy$ to compute the area of the parallelepiped spanned by the two Euclidian unit vectors. Not surprisingly we got a volume of 1, exactly what we would expect from a unit square. Now we move to the next identity. First we find that $f^{-1}(1, 1) \approx (0.785, 1.414)$. This is the base point of the push-forwards of the two Euclidian unit vectors. Next we find the general tangent mapping

$$T_{(x,y)} f^{-1} = \begin{bmatrix} \frac{\partial \theta}{\partial x} & \frac{\partial \theta}{\partial y} \\ \frac{\partial r}{\partial x} & \frac{\partial r}{\partial y} \end{bmatrix} = \begin{bmatrix} \frac{-y}{x^2+y^2} & \frac{x}{x^2+y^2} \\ \frac{x}{\sqrt{x^2+y^2}} & \frac{y}{\sqrt{x^2+y^2}} \end{bmatrix}_{(x,y)}$$

which means that for the point $(x, y) = (1, 1)$ we have

$$T_{(1,1)} f^{-1} = \begin{bmatrix} \frac{-1}{2} & \frac{1}{2} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix}_{(1,1)}.$$

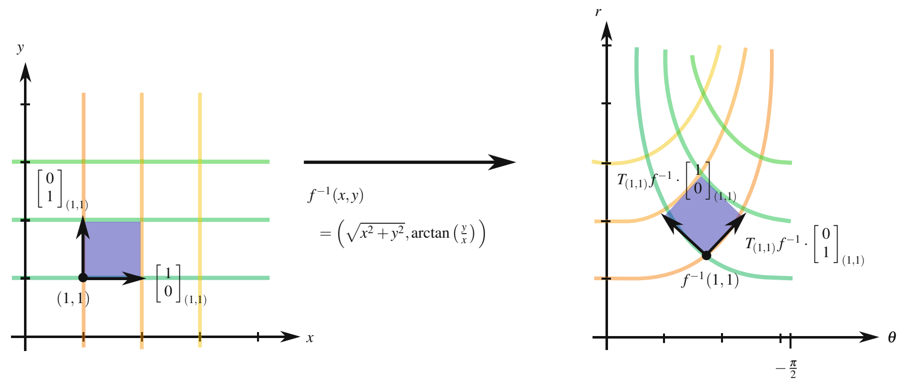


Fig. 6.11 The relationship between the volume forms $dx \wedge dy$ and $d\theta \wedge dr$ at the points $(x, y) = (1, 1)$ and $f^{-1}(1, 1)$

Next we find the push-forwards of the Euclidian unit vectors at $(1, 1)$ by f^{-1},

$$T_{(1,1)}f^{-1} \cdot \begin{bmatrix} 1 \\ 0 \end{bmatrix}_{(1,1)} = \begin{bmatrix} \frac{-1}{2} & \frac{1}{2} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix}_{(1,1)} \begin{bmatrix} 1 \\ 0 \end{bmatrix}_{(1,1)} = \begin{bmatrix} \frac{-1}{2} \\ \frac{1}{\sqrt{2}} \end{bmatrix}_{(0.785,1.414)}$$

$$T_{(1,1)}f^{-1} \cdot \begin{bmatrix} 0 \\ 1 \end{bmatrix}_{(1,1)} = \begin{bmatrix} \frac{-1}{2} & \frac{1}{2} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix}_{(1,1)} \begin{bmatrix} 0 \\ 1 \end{bmatrix}_{(1,1)} = \begin{bmatrix} \frac{1}{2} \\ \frac{1}{\sqrt{2}} \end{bmatrix}_{(0.785,1.414)}$$

which are needed in

$$(d\theta \wedge dr)_{(0.785,1.414)} \left(\begin{bmatrix} \frac{-1}{2} \\ \frac{1}{\sqrt{2}} \end{bmatrix}_{(0.785,1.414)}, \begin{bmatrix} \frac{1}{2} \\ \frac{1}{\sqrt{2}} \end{bmatrix}_{(0.785,1.414)} \right) = \begin{vmatrix} \frac{-1}{2} & \frac{1}{2} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{vmatrix} = \frac{-1}{\sqrt{2}}.$$

Here we used the $r\theta$-plane volume form to find the volume of the parallelepiped spanned by the pushed-forward vectors. The relationship between the areas computed in this way is in exact correspondence with the relationship between the volume forms $dx \wedge dy$ and $d\theta \wedge dr$ given by $T^*f^{-1} \cdot \big(d\theta \wedge dr\big) = \frac{-1}{\sqrt{x^2+y^2}} dx \wedge dy$ at the point $(1, 1) \in \mathbb{R}^2_{xy}$. That is,

$$\underbrace{(d\theta \wedge dr)_{f^{-1}(1,1)} \left(T_{(1,1)}f^{-1} \cdot \begin{bmatrix} 1 \\ 0 \end{bmatrix}_{(1,1)}, T_{(1,1)}f^{-1} \cdot \begin{bmatrix} 0 \\ 1 \end{bmatrix}_{(1,1)} \right)}_{=\frac{-1}{\sqrt{2}}} = \frac{-1}{\sqrt{2}} \underbrace{(dx \wedge dy)_{(1,1)} \left(\begin{bmatrix} 1 \\ 0 \end{bmatrix}_{(1,1)}, \begin{bmatrix} 0 \\ 1 \end{bmatrix}_{(1,1)} \right)}_{=1}.$$

In other words we have just seen, using two vectors that at the point $(1, 1)$, that

$$T^*f^{-1}(d\theta \wedge dr) = \frac{1}{\sqrt{2}}\, dx \wedge dy.$$

We attempt to show what is going on in Fig. 6.11. Now redo the last example, only at a different base point, to see what changes.

Question 6.9 For the polar coordinate transformations $f^{-1} : \mathbb{R}^2_{xy} \to \mathbb{R}^2_{r\theta}$ given by $f^{-1}(x, y) = \left(\arctan\left(\frac{y}{x}\right), \pm\sqrt{x^2 + y^2}\right)$,

(a) Find $f^{-1}(2, 2)$.
(b) Consider $\begin{bmatrix} 1 \\ 0 \end{bmatrix}_{(2,2)}$ and $\begin{bmatrix} 0 \\ 1 \end{bmatrix}_{(2,2)} \in T_{(1,1)}\mathbb{R}^2$. Find

$$(dx \wedge dy)_{(2,2)} \left(\begin{bmatrix} 1 \\ 0 \end{bmatrix}_{(2,2)}, \begin{bmatrix} 0 \\ 1 \end{bmatrix}_{(2,2)} \right).$$

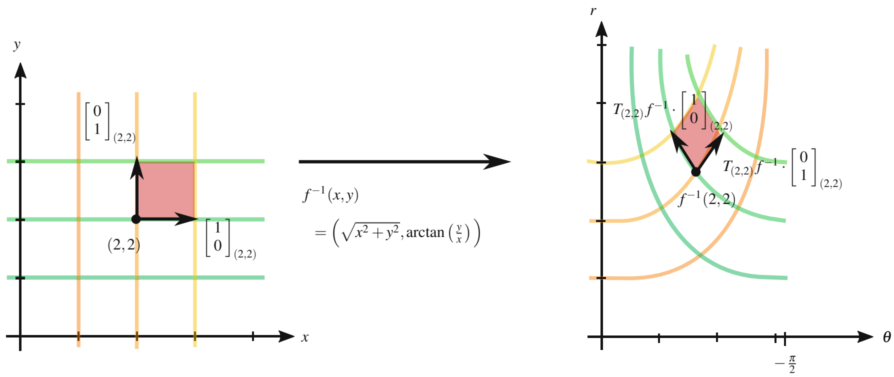


Fig. 6.12 The relationship between the volume forms $dx \wedge dy$ and $d\theta \wedge dr$ at the points $(x, y) = (2, 2)$ and $f^{-1}(2, 2)$

(c) Find $T_{(2,2)} f^{-1}$.

(d) Find $T_{(2,2)} f^{-1} \cdot \begin{bmatrix} 1 \\ 0 \end{bmatrix}_{(2,2)}$ and $T_{(2,2)} f^{-1} \cdot \begin{bmatrix} 0 \\ 1 \end{bmatrix}_{(2,2)}$ (These are the push-forwards of the Euclidian unit vectors at $(2, 2)$ by f^{-1} to $f^{-1}(2, 2)$.)

(e) Find $(d\theta \wedge dr)_{f^{-1}(2,2)} \left(T_{(2,2)} f^{-1} \cdot \begin{bmatrix} 1 \\ 0 \end{bmatrix}_{(2,2)}, T_{(2,2)} f^{-1} \cdot \begin{bmatrix} 0 \\ 1 \end{bmatrix}_{(2,2)} \right)$.

Figure 6.12 shows what is happening in this question. The relationship you should have ended up with in this question is

$$(d\theta \wedge dr)_{f^{-1}(2,2)} \left(T_{(1,1)} f^{-1} \cdot \begin{bmatrix} 1 \\ 0 \end{bmatrix}_{(2,2)}, T_{(1,1)} f^{-1} \cdot \begin{bmatrix} 0 \\ 1 \end{bmatrix}_{(2,2)} \right) = \frac{-1}{\sqrt{8}} (dx \wedge dy)_{(2,2)} \left(\begin{bmatrix} 1 \\ 0 \end{bmatrix}_{(2,2)}, \begin{bmatrix} 0 \\ 1 \end{bmatrix}_{(2,2)} \right).$$

The numerical coefficient has changed. This has happened because the point at which we are evaluating areas has changed. In other words, the relationship between the area forms in the xy-plane and the $r\theta$-plane depend on the point where the area form is being used. So, at the point $(1, 1)$ we have

$$T^*_{(1,1)} f^{-1} \cdot \big(d\theta \wedge dr\big) = \frac{-1}{\sqrt{2}} dx \wedge dy$$

and at the point $(2, 2)$ we have

$$T^*_{(2,2)} f^{-1} \cdot \big(d\theta \wedge dr\big) = \frac{-1}{\sqrt{8}} dx \wedge dy$$

which is exactly what we would expect from the actual identity we found,

$$T^*_{(x,y)} f^{-1} \cdot \big(d\theta \wedge dr\big) = \frac{-1}{\sqrt{x^2 + y^2}} dx \wedge dy.$$

6.5 Cylindrical and Spherical Coordinates

Cylindrical and Spherical coordinates are two coordinate systems that are very common for $\mathbb{R}^3$. Now that we have a pretty good idea of how push-forwards of vectors and pull-backs of volume forms work, we will provide less detail in this section. The cylindrical coordinate transformation $f : \mathbb{R}^3_{r\theta z} \to \mathbb{R}^3_{xyz}$ is given by

$$f(r, \theta, z) = \big(x(r, \theta, z), y(r, \theta, z), z(r, \theta, z)\big) = \big(r \cos\theta, r \sin\theta, z\big),$$

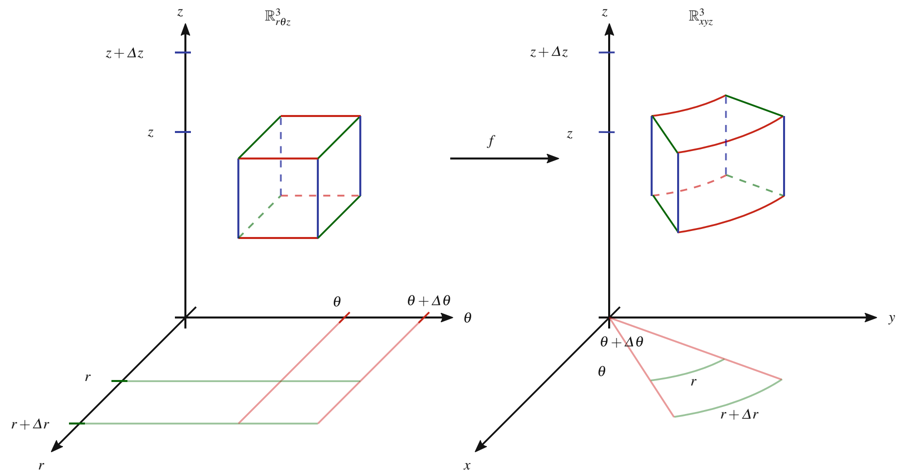


Fig. 6.13 The cylindrical coordinates mapping $f : \mathbb{R}^3_{r\theta z} \to \mathbb{R}^3_{xyz}$ given by $x = r\cos\theta$, $y = r\sin\theta$, and $z = z$. A cube in $\mathbb{R}^3_{r\theta z}$ is mapped to a wedge in $\mathbb{R}^3_{xyz}$

which is generally written in calculus classes as

$$\begin{aligned} x &= r\cos\theta, \\ y &= r\sin\theta, \\ z &= z. \end{aligned}$$

See Fig. 6.13. The cylindrical coordinate transformation f induces the following maps:

$$\begin{array}{ccc} \bigwedge^3\left(\mathbb{R}^2_{r\theta z}\right) & \xleftarrow{T^*f} & \bigwedge^3\left(\mathbb{R}^2_{xyz}\right) \\ T^*_{(r,\theta,z)}f \cdot \big(dx \wedge dy \wedge dz\big) & \longleftarrow & dx \wedge dy \wedge dz \\ \mathbb{R}^3_{r\theta z} & \xrightarrow{Tf} & \mathbb{R}^3_{xyz} \\ v_{(r,\theta,z)} & \longmapsto & T_{(r,\theta,z)}f \cdot v_{(r,\theta,z)} \\ \mathbb{R}^3_{r\theta z} & \xrightarrow{f} & \mathbb{R}^3_{xyz} \\ (\theta, r, z) & \longmapsto & \big(r\cos\theta, r\sin\theta, z\big). \end{array}$$

Proceeding as before and either using the properties of the wedgeproduct or the wedgeproduct formula in terms of the Jacobian from we can write $dx \wedge dy \wedge dz$ in terms of $d\theta \wedge dr \wedge dz$. Here we find

$$\begin{aligned} dx &= \frac{\partial x}{\partial\theta}d\theta + \frac{\partial x}{\partial r}dr + \frac{\partial x}{\partial z}dz \\ &= -r\sin\theta d\theta + \cos\theta dr \\ dy &= \frac{\partial y}{\partial\theta}d\theta + \frac{\partial y}{\partial r}dr + \frac{\partial y}{\partial z}dz \\ &= r\cos\theta d\theta + \sin\theta dr \\ dz &= \frac{\partial z}{\partial\theta}d\theta + \frac{\partial z}{\partial r}dr + \frac{\partial z}{\partial z}dz \\ &= dz \end{aligned}$$

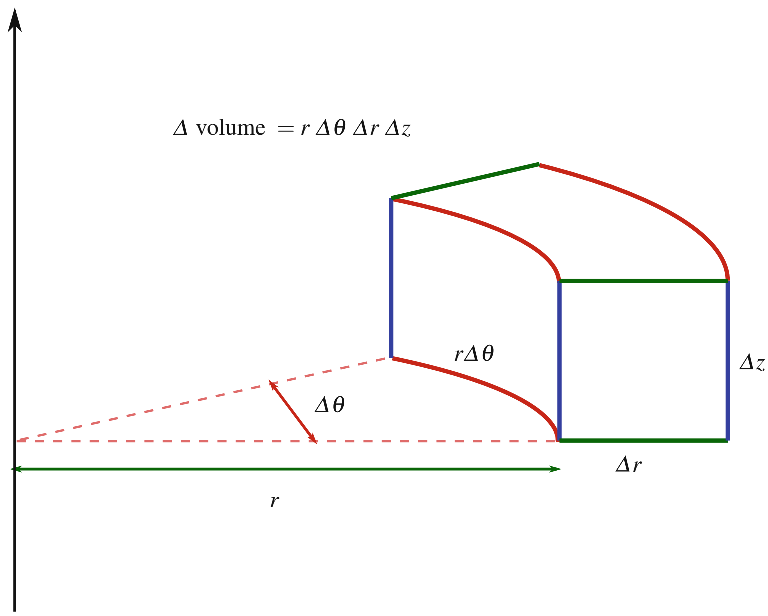


Fig. 6.14 The cylindrical "$\theta r z$ volume element" as shown in most calculus textbooks

which we then use to get

$$\begin{aligned} dx \wedge dy \wedge dz &= \Big(-r\sin\theta d\theta + \cos\theta dr \Big) \wedge \Big(r\cos\theta d\theta + \sin\theta dr \Big) \wedge dz \\ &= \Big(-r\sin^2\theta d\theta \wedge dr + r\cos^2\theta dr \wedge d\theta \Big) \wedge dz \\ &= -r d\theta \wedge dr \wedge dz. \end{aligned}$$

As we know by now, what this identity really means is

$$\boxed{T^*_{(r,\theta,z)} f \cdot \big(dx \wedge dy \wedge dz \big) = -r d\theta \wedge dr \wedge dz.}$$

What we have just done is find the "$\theta r z$ volume element" that is usually presented in calculus. A lot of calculus books give a picture of the "$\theta r z$ volume element" that looks like Fig. 6.14 and then use this picture to find the change in volume formula, though orientation is not taken into account. Thus, even though our presentation has been far more abstract and theoretical then anything you saw in calculus, you are already familiar with this from a different perspective.

Question 6.10 Find $T^*_{(r,\theta,z)} f \cdot \big(dx \wedge dy \wedge dz \big)$ using the formula for the pull-back of a volume form.

Question 6.11 Explain the relationship between the formula $T^*_{(r,\theta,z)} f \cdot \big(dx \wedge dy \wedge dz \big) = -r d\theta \wedge dr \wedge dz$ and the "$\theta r z$ volume element" shown in Fig. 6.14.

We now move to spherical coordinate changes. The spherical coordinate transformation $g : \mathbb{R}^3_{\rho\phi\theta} \to \mathbb{R}^3_{xyz}$ is given by

$$g(\rho,\phi,\theta) = \big(x(\rho,\phi,\theta), y(\rho,\phi,\theta), z(\rho,\phi,\theta) \big) = \big(\rho\sin\phi\cos\theta, \rho\sin\phi\sin\theta, \rho\cos\phi \big),$$

which is generally written in calculus classes as

$$\begin{aligned} x &= \rho\sin\phi\cos\theta, \\ y &= \rho\sin\phi\sin\theta, \\ z &= \rho\cos\phi. \end{aligned}$$

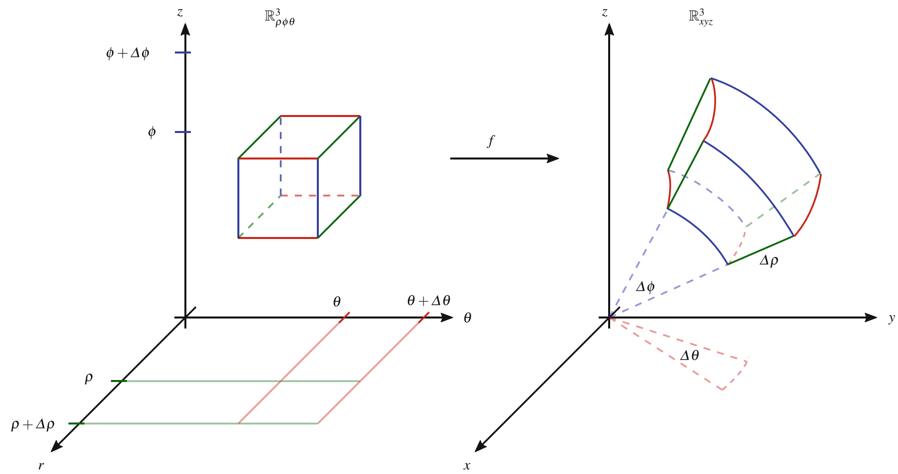


Fig. 6.15 The spherical coordinate transformation $g : \mathbb{R}^3_{\rho\phi\theta} \to \mathbb{R}^3_{xyz}$ is given by $x = \rho \sin\phi \cos\theta$, $y = \rho \sin\phi \sin\theta$, and $z = \rho \cos\phi$

The mapping $g : \mathbb{R}^3_{\rho\phi\theta} \to \mathbb{R}^3_{xyz}$ is illustrated in Fig. 6.15. The spherical coordinate transformation induces the following maps:

$$\begin{array}{rcl}
\bigwedge^3\left(\mathbb{R}^2_{\rho\phi\theta}\right) & \xleftarrow{T^*g} & \bigwedge^3\left(\mathbb{R}^2_{xyz}\right) \\
T^*_{(\rho,\phi,\theta)}g \cdot \big(dx \wedge dy \wedge dz\big) & \longleftarrow & dx \wedge dy \wedge dz \\
\mathbb{R}^3_{\rho\phi\theta} & \xrightarrow{Tg} & \mathbb{R}^3_{xyz} \\
v_{(\rho,\phi,\theta)} & \longmapsto & T_{(\rho,\phi,\theta)}g \cdot v_{(\rho,\phi,\theta)} \\
\mathbb{R}^3_{\rho\phi\theta} & \xrightarrow{g} & \mathbb{R}^3_{xyz} \\
(\rho,\phi,\theta) & \longmapsto & \big(\rho\sin\phi\cos\theta,\ \rho\sin\phi\sin\theta,\ \rho\cos\phi\big).
\end{array}$$

Again, a straightforward computation shows that

$$\begin{aligned}
dx &= \sin\phi\cos\theta d\rho - \rho\sin\phi\sin\theta d\theta + \rho\cos\phi\cos\theta d\phi, \\
dy &= \sin\phi\sin\theta d\rho + \rho\sin\phi\cos\theta d\theta + \rho\cos\phi\sin\theta d\phi, \\
dz &= \cos\phi d\rho - \rho\sin\phi d\phi.
\end{aligned}$$

Another straightforward, though somewhat longer, computation that you should do yourself gives

$$dx \wedge dy \wedge dx = -\rho^2 \sin\phi d\rho \wedge d\theta \wedge d\phi.$$

As we now know, this identity really means

$$\boxed{T^*_{(\rho,\phi,\theta)}g \cdot \big(dx \wedge dy \wedge dz\big) = -\rho^2 \sin\phi d\rho \wedge d\theta \wedge d\phi.}$$

Since $dx \wedge dy \wedge dz$ is a volume form we could have found this using the pull-back of a volume form formula. A lot of calculus books give a picture of the "$\rho\phi\theta$ volume element" that looks like Fig. 6.16. Notice how we can pictorially find the change in volume from the picture, though again orientation is not taken into account. So again, though we have approached this from a much more abstract standpoint, it is something you have already been exposed to in calculus.

Question 6.12 Find $T^*_{(\rho,\phi,\theta)}g \cdot \big(dx \wedge dy \wedge dz\big)$ using the formula for the pull-back of a volume form.

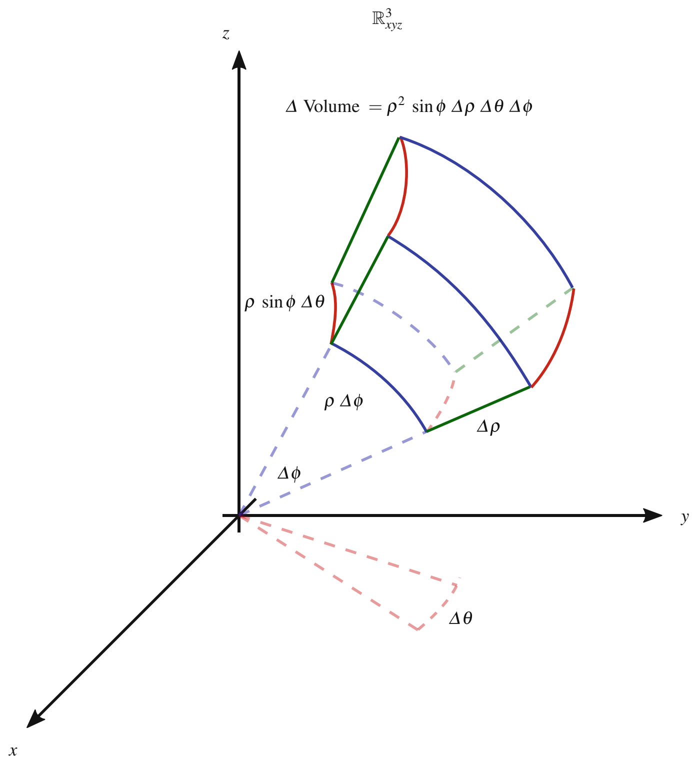


Fig. 6.16 The spherical "$\rho\phi\theta$ volume element" as shown in most calculus textbooks

Question 6.13 Explain the relationship between the formula $T^*_{(\rho,\phi,\theta)}g \cdot \big(dx \wedge dy \wedge dz\big) = -\rho^2 \sin\phi d\rho \wedge d\theta \wedge d\phi$ and the "$\rho\phi\theta$ volume element" shown in Fig. 6.16.

6.6 Pull-Backs of Differential Forms

Consider some of the relationships we have arrived at in the previous few sections

$$T^*f \cdot (dx \wedge dy) = \frac{-1}{2} du \wedge dv \qquad \text{(linear example)}$$

$$T^*f \cdot (dx \wedge dy) = -r d\theta \wedge dr \qquad \text{(polar coordinates)}$$

$$T^*f \cdot (dx \wedge dy \wedge dz) = -r d\theta \wedge dr \wedge dz \qquad \text{(cylindrical coordinates)}$$

$$T^*f \cdot (dx \wedge dy \wedge dz) = -\rho^2 \sin\phi d\rho \wedge d\theta \wedge d\phi \qquad \text{(spherical coordinates)}$$

where f is the appropriate mapping $f : \mathbb{R}^{2/3}_{\text{Other coord}} \to \mathbb{R}^{2/3}_{\text{Euclidian}}$. Based on our definition of pull-backs of volume forms what we really mean by these are

$$\underbrace{\frac{-1}{2}(du \wedge dv)_p}_{= T_p^* f \cdot (dx \wedge dy)_{f(p)}} (v_p, w_p) = (dx \wedge dy)_{f(p)}(T_p f \cdot v_p, T_p f \cdot w_p),$$

$$\underbrace{-r(d\theta \wedge dr)_p}_{= T_p^* f \cdot (dx \wedge dy)_{f(p)}} (v_p, w_p) = (dx \wedge dy)_{f(p)}(T_p f \cdot v_p, T_p f \cdot w_p),$$

$$\underbrace{-r(d\theta \wedge dr \wedge dz)_p}_{=T_p^* f\cdot(dx\wedge dy\wedge dz)_{f(p)}}(u_p, v_p, w_p) = (dx \wedge dy \wedge dz)_{f(p)}(T_p f \cdot u_p, T_p f \cdot v_p, T_p f \cdot w_p),$$

$$\underbrace{-\rho^2 \sin\phi(d\rho \wedge d\theta \wedge d\phi)_p}_{=T_p^* f\cdot(dx\wedge dy\wedge dz)_{f(p)}}(u_p, v_p, w_p) = (dx \wedge dy \wedge dz)_{f(p)}(T_p f \cdot u_p, T_p f \cdot v_p, T_p f \cdot w_p).$$

Now we take a moment to give the definition of the pull-back of a differential form again in some detail. Suppose we have a map $\phi : \mathbb{R}^n \to \mathbb{R}^m$ given by

$$\phi(x_1, x_2, \dots, x_n) = \Big(\phi_1(x_1, x_2, \dots, x_n), \phi_2(x_1, x_2, \dots, x_n), \dots, \phi_m(x_1, x_2, \dots, x_n)\Big).$$

ϕ defines a map, called the push-forward, or tangent map, of ϕ, which is denoted $T\phi : T\mathbb{R}^n \to T\mathbb{R}^m$ from the tangent bundle of $\mathbb{R}^n$ to the tangent bundle of $\mathbb{R}^m$. In coordinates the mapping $T\phi$ is given by the Jacobian matrix

$$T\phi = \begin{bmatrix} \frac{\partial\phi_1}{\partial x_1} & \cdots & \frac{\partial\phi_1}{\partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial\phi_m}{\partial x_1} & \cdots & \frac{\partial\phi_m}{\partial x_n} \end{bmatrix}.$$

As has been mentioned already, other notations for $T\phi$ that one often encounters are $D\phi$ and ϕ_*. We use this push-forward mapping $T\phi$ to help us define another mapping called the pull-back of a differential form by ϕ, or simply the pull-back by ϕ, which is denoted by $T^*\phi$ or ϕ^*. Actually, ϕ^* is the standard notation for the pull-back, but we will use $T^*\phi$.

Before actually defining $T^*\phi$ we will tell what spaces it operates on. Recall that the notation $\bigwedge^k \mathbb{R}^n$ is used to denote the vector space of k-forms on $\mathbb{R}^n$ while $\bigwedge \mathbb{R}^n = \bigoplus_{k=1}^{\infty} \bigwedge^k \mathbb{R}^n$ denotes the set of all differential forms on $\mathbb{R}^n$ of any size. Using this we generally show the maps ϕ and its induced maps $T\phi$ and $T^*\phi$ as

$$\begin{gathered} \bigwedge \mathbb{R}^n \xleftarrow{T^*\phi} \bigwedge \mathbb{R}^m \\ T\mathbb{R}^n \xrightarrow{T\phi} T\mathbb{R}^m \\ \mathbb{R}^n \xrightarrow{\phi} \mathbb{R}^m. \end{gathered}$$

So, the pull-back map operates on the space $\bigwedge \mathbb{R}^m$ by taking a differential form on $\mathbb{R}^m$ and producing an element of $\bigwedge \mathbb{R}^n$, that is, a differential form on $\mathbb{R}^n$. Notice that this mapping is going in the opposite direction from ϕ and $T\phi$. This is the reason $T^*\phi$ is called the *pull-back* while $T\phi$ is called the *push-forward*. Next, even though $T^*\phi$ operates on a k-form of any order, we sometimes abuse notation and write

$$T^*\mathbb{R}^n \xleftarrow{T^*\phi} T^*\mathbb{R}^m,$$

writing $T^*\mathbb{R}^n$ and $T^*\mathbb{R}^m$ instead of $\bigwedge \mathbb{R}^n$ and $\bigwedge \mathbb{R}^m$. We will define the pull-back of the k-form α, which is denoted $T^*\phi \cdot \alpha \in \bigwedge^k \mathbb{R}^n$, to just be $\alpha \in \bigwedge^k \mathbb{R}^m$ acting on the pushed forward vectors, $T\phi \cdot v_i \in T\mathbb{R}^m$,

$$\boxed{\text{Pull-Back of } \alpha : \quad \Big(T^*\phi \cdot \alpha\Big)(v_1, v_2, \dots, v_k) \equiv \alpha\Big(T\phi \cdot v_1, T\phi \cdot v_2, \dots, T\phi \cdot v_k\Big).}$$

If we wanted to be careful and start keeping track of base points the notation starts to get a little more cumbersome, though there are times when you need to do that. The one odd part of this notation is that the map $T^*\phi$ is indexed NOT by the point it is coming from but instead by the point it is going to. This is done to keep the notation for the pull-back and push-forward mappings $T_p^*\phi$ and $T_p\phi$ consistent, that is, “dual” to each other. This is something that does not happen very often and certainly seems odd the first few times you see it.

$$\bigwedge\nolimits_p \mathbb{R}^n \xleftarrow{T_p^*\phi} \bigwedge\nolimits_{\phi(p)} \mathbb{R}^m$$

$$T_p\mathbb{R}^n \xrightarrow{T_p\phi} T_{\phi(p)}\mathbb{R}^m$$
$$\mathbb{R}^n \xrightarrow{\phi} \mathbb{R}^m.$$

With base points added we have the definition

$$\boxed{\text{Pull-Back of } \alpha : \quad \Big(T_p^*\phi \cdot \alpha_{\phi(p)}\Big)(v_{1_p}, v_{2_p}, \ldots, v_{k_p}) \equiv \alpha_{\phi(p)}\Big(T_p\phi \cdot v_{1_p}, T_p\phi \cdot v_{2_p}, \ldots, T_p\phi \cdot v_{k_p}\Big).}$$

Now, to actually see how this works let us consider a few examples. We have worked extensively with volume forms so we now turn our attention to k-forms for arbitrary k. Unfortunately there is no nice formula for finding the pull-backs of arbitrary forms, but there is a fairly standard procedure we can use. Consider the mapping $\phi : \mathbb{R}^2_{xy} \to \mathbb{R}^2_{uv}$ given by $\phi(x, y) = (x + y, x - y)$ and the one-form $\alpha = vdu + udv$ on $\mathbb{R}^2_{uv}$. We want to find $T^*\phi \cdot \alpha$ on $\mathbb{R}^2_{xy}$. Notice, we are now dealing with a one-form, not a volume form.

We will approach this problem in the following way. The pull-back of α, $T^*\phi \cdot \alpha$, will be a one-form on $\mathbb{R}^2_{xy}$ and thus will have the form $f(x, y)dx + g(x, y)dy$ for some, as of yet, unknown functions f and g. It is these functions that we want to find. First we notice that

$$(fdx + gdy)\left(\begin{bmatrix} 1 \\ 0 \end{bmatrix}\right) = fdx\left(\begin{bmatrix} 1 \\ 0 \end{bmatrix}\right) + gdy\left(\begin{bmatrix} 1 \\ 0 \end{bmatrix}\right) = f(1) + g(0) = f$$

and

$$(fdx + gdy)\left(\begin{bmatrix} 0 \\ 1 \end{bmatrix}\right) = fdx\left(\begin{bmatrix} 0 \\ 1 \end{bmatrix}\right) + gdy\left(\begin{bmatrix} 0 \\ 1 \end{bmatrix}\right) = f(0) + g(1) = g.$$

We will use these identities to find out what f and g are. We will also need the identity

$$T\phi = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}.$$

We now proceed to use the above identities to find f and g as follows,

$$\begin{aligned}
f &= \big(fdx + gdy\big)\left(\begin{bmatrix} 1 \\ 0 \end{bmatrix}\right) = \big(T^*\phi \cdot \alpha\big)\left(\begin{bmatrix} 1 \\ 0 \end{bmatrix}\right) \\
&= \alpha\left(T\phi \cdot \begin{bmatrix} 1 \\ 0 \end{bmatrix}\right) = \big(vdu + udv\big)\left(\begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}\begin{bmatrix} 1 \\ 0 \end{bmatrix}\right) \\
&= \big(vdu + udv\big)\left(\begin{bmatrix} 1 \\ 1 \end{bmatrix}\right) = vdu\left(\begin{bmatrix} 1 \\ 1 \end{bmatrix}\right) + udv\left(\begin{bmatrix} 1 \\ 1 \end{bmatrix}\right) \\
&= u + v = (x + y) + (x - y) = 2x
\end{aligned}$$

and

$$\begin{aligned}
g &= \big(fdx + gdy\big)\left(\begin{bmatrix} 0 \\ 1 \end{bmatrix}\right) = \big(T^*\phi \cdot \alpha\big)\left(\begin{bmatrix} 0 \\ 1 \end{bmatrix}\right) \\
&= \alpha\left(T\phi \cdot \begin{bmatrix} 0 \\ 1 \end{bmatrix}\right) = \big(vdu + udv\big)\left(\begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}\begin{bmatrix} 0 \\ 1 \end{bmatrix}\right) \\
&= \big(vdu + udv\big)\left(\begin{bmatrix} 1 \\ -1 \end{bmatrix}\right) = vdu\left(\begin{bmatrix} 1 \\ -1 \end{bmatrix}\right) + udv\left(\begin{bmatrix} 1 \\ -1 \end{bmatrix}\right) \\
&= u - v = (x + y) - (x - y) = 2y.
\end{aligned}$$

Combining everything we have that

$$T^*\phi \cdot \big(v du + u dv\big) = 2x dx + 2y dy.$$

Notice in this example how we went about finding the exact form of $T^*\phi \cdot \alpha$. Knowing the pull-back was a one-form meant that we knew that it had to be a sum of functions multiplied by the basis elements dx and dy. We also knew that $dx\left(\begin{bmatrix}1\\0\end{bmatrix}\right) = 1$ and $dy\left(\begin{bmatrix}0\\1\end{bmatrix}\right) = 1$ so we could use the push-forwards of these vectors to find the desired functions.

Now Consider the same mapping, $\phi : \mathbb{R}^2_{xy} \to \mathbb{R}^2_{uv}$ given by $\phi(x, y) = (x + y, x - y)$, and the two-form $\alpha = du \wedge dv$. We want to find $T^*\phi \cdot \alpha$. I suspect you'll notice we are back to our area forms so you won't be surprised by the answer - we are just going to get it in a different way. Clearly $T^*\phi \cdot \alpha$ will be a two-form on $\mathbb{R}^2_{xy}$ so it will have the form $f(x, y)dx \wedge dy$ for some function f, since $dx \wedge dy$ is the only basis element of the two-forms on $\mathbb{R}^2$. Next we ask ourselves what two vectors, when "eaten" by $dx \wedge dy$ give 1? Clearly $\begin{bmatrix}1\\0\end{bmatrix}$ and $\begin{bmatrix}0\\1\end{bmatrix}$ work since

$$dx \wedge dy\left(\begin{bmatrix}1\\0\end{bmatrix}, \begin{bmatrix}0\\1\end{bmatrix}\right) = \begin{vmatrix}1 & 0\\0 & 1\end{vmatrix} = 1.$$

Using this we have

$$\begin{aligned}
f &= \big(f dx \wedge dy\big)\left(\begin{bmatrix}1\\0\end{bmatrix}, \begin{bmatrix}0\\1\end{bmatrix}\right)\\
&= \Big(T^*\phi \cdot \alpha\Big)\left(\begin{bmatrix}1\\0\end{bmatrix}, \begin{bmatrix}0\\1\end{bmatrix}\right)\\
&= \alpha\Big(T\phi \cdot \begin{bmatrix}1\\0\end{bmatrix}, T\phi \cdot \begin{bmatrix}0\\1\end{bmatrix}\Big)\\
&= du \wedge dv\Big(\begin{bmatrix}1 & 1\\1 & -1\end{bmatrix}\begin{bmatrix}1\\0\end{bmatrix}, \begin{bmatrix}1 & 1\\1 & -1\end{bmatrix}\begin{bmatrix}0\\1\end{bmatrix}\Big)\\
&= du \wedge dv\Big(\begin{bmatrix}1\\1\end{bmatrix}, \begin{bmatrix}1\\-1\end{bmatrix}\Big)\\
&= \begin{vmatrix}1 & 1\\1 & -1\end{vmatrix}\\
&= -2.
\end{aligned}$$

Putting everything together we get

$$-2 dx \wedge dy = T^*\phi \cdot \big(du \wedge dv\big).$$

Question 6.14 Now consider the inverse mapping $\phi^{-1} : \mathbb{R}^2_{uv} \to \mathbb{R}^2_{xy}$ given by $\phi^{-1}(u, v) = \Big(\frac{1}{2}(u + v), \frac{1}{2}(u - v)\Big)$ and suppose $\alpha = dx \wedge dy$. Find $T^*\phi^{-1} \cdot \alpha$ using this method.

Now we will consider a couple more examples of increasing complexity. Suppose $\phi : \mathbb{R}^2_{xy} \longrightarrow \mathbb{R}^3_{uvw}$ is given by $f(x, y) = \big(u(x, y), v(x, y), w(x, y)\big) = (x + y, x - y, xy)$ and consider the one-form $\alpha = v du + w dv + u dw$ on $\mathbb{R}^3$. We want to find $T^*\phi \cdot \alpha$. As before, $T^*\phi \cdot \alpha$ will have the form $f dx + g dy$ for some functions f and g, so as before we will use vectors $\begin{bmatrix}1\\0\end{bmatrix}$ and $\begin{bmatrix}0\\1\end{bmatrix}$. Also,

$$T\phi = \begin{bmatrix}\frac{\partial u}{\partial x} & \frac{\partial u}{\partial y}\\ \frac{\partial v}{\partial x} & \frac{\partial v}{\partial y}\\ \frac{\partial w}{\partial x} & \frac{\partial w}{\partial y}\end{bmatrix} = \begin{bmatrix}1 & 1\\1 & -1\\y & x\end{bmatrix}$$

so

$$
\begin{aligned}
f &= \big(f dx + g dy\big)\left(\begin{bmatrix}1\\0\end{bmatrix}\right)\\
&= \big(T^*\phi \cdot \alpha\big)\left(\begin{bmatrix}1\\0\end{bmatrix}\right)\\
&= \alpha\left(T\phi \cdot \begin{bmatrix}1\\0\end{bmatrix}\right)\\
&= \big(v du + w dv + u dw\big)\left(\begin{bmatrix}1 & 1\\1 & -1\\y & x\end{bmatrix}\begin{bmatrix}1\\0\end{bmatrix}\right)\\
&= \big(v du + w dv + u dw\big)\left(\begin{bmatrix}1\\1\\y\end{bmatrix}\right)\\
&= v(1) + w(1) + u(y)\\
&= (x - y)(1) + (xy)(1) + (x + y)(y)\\
&= x - y + 2xy + y^2
\end{aligned}
$$

and

$$
\begin{aligned}
g &= \big(f dx + g dy\big)\left(\begin{bmatrix}0\\1\end{bmatrix}\right)\\
&= \big(T^*\phi \cdot \alpha\big)\left(\begin{bmatrix}0\\1\end{bmatrix}\right)\\
&= \alpha\left(T\phi \cdot \begin{bmatrix}0\\1\end{bmatrix}\right)\\
&= \big(v du + w dv + u dw\big)\left(\begin{bmatrix}1 & 1\\1 & -1\\y & x\end{bmatrix}\begin{bmatrix}0\\1\end{bmatrix}\right)\\
&= \big(v du + w dv + u dw\big)\left(\begin{bmatrix}1\\-1\\x\end{bmatrix}\right)\\
&= v(1) + w(-1) + u(x)\\
&= (x - y)(1) + (xy)(-1) + (x + y)(x)\\
&= x - y + x^2
\end{aligned}
$$

so we have

$$
T^*\phi \cdot \big(v du + w dv + u dw\big) = \big(x - y + 2xy + y^2\big)dx + \big(x - y + x^2\big)dy.
$$

Next suppose we have the mapping $\phi : \mathbb{R}^3_{abc} \longrightarrow \mathbb{R}^4_{xyzw}$ given by

$$
\begin{aligned}
\phi\big(a, b, c\big) &= \big(x(a, b, c), y(a, b, c), z(a, b, c), w(a, b, c)\big)\\
&= \big(a, b, c, abc\big).
\end{aligned}
$$

If $\omega = x^2 dy \wedge dz + y^2 dz \wedge dw$ is a two-form on $\mathbb{R}$, find $T^*\phi \cdot \omega$. First we find the push-forward map

$$T\phi = \begin{bmatrix} \frac{\partial x}{\partial a} & \frac{\partial x}{\partial b} & \frac{\partial x}{\partial c} \\ \frac{\partial y}{\partial a} & \frac{\partial y}{\partial b} & \frac{\partial y}{\partial c} \\ \frac{\partial z}{\partial a} & \frac{\partial z}{\partial b} & \frac{\partial z}{\partial c} \\ \frac{\partial w}{\partial a} & \frac{\partial w}{\partial b} & \frac{\partial w}{\partial c} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ bc & ac & ab \end{bmatrix}$$

and the push-forwards of the $\mathbb{R}^3$ basis vectors,

$$T\phi \cdot \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \\ 0 \\ bc \end{bmatrix}, \quad T\phi \cdot \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \\ 0 \\ ac \end{bmatrix}, \quad T\phi \cdot \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 1 \\ ab \end{bmatrix}.$$

We know all two-forms on $\mathbb{R}^3$ are in $\text{span}\big\{da \wedge db, db \wedge dc, dc \wedge da\big\}$ and so

$$T^*\phi \cdot \omega = f da \wedge db + g db \wedge dc + h dc \wedge da$$

for some functions f, g, h. Also, we know

$$da \wedge db \left(\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} \right) = \begin{vmatrix} 1 & 0 \\ 0 & 1 \end{vmatrix} = 1,$$

$$db \wedge dc \left(\begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \right) = \begin{vmatrix} 1 & 0 \\ 0 & 1 \end{vmatrix} = 1,$$

$$dc \wedge da \left(\begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \right) = \begin{vmatrix} 1 & 0 \\ 0 & 1 \end{vmatrix} = 1.$$

All other combinations of basis two-forms and vectors equal zero. Following the same strategy as before we have

$$\begin{aligned} f &= \Big(f da \wedge db + g db \wedge dc + h dc \wedge da \Big) \left(\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} \right) \\ &= \Big(T^*\phi \cdot \omega \Big) \left(\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} \right) \\ &= \Big(x^2 dy \wedge dz + y^2 dz \wedge dw \Big) \left(T\Phi \cdot \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, T\Phi \cdot \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} \right) \\ &= \Big(x^2 dy \wedge dz + y^2 dz \wedge dw \Big) \left(\begin{bmatrix} 1 \\ 0 \\ 0 \\ bc \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \\ ac \end{bmatrix} \right) \end{aligned}$$

$$
\begin{aligned}
&= x^2 dy \wedge dz \left(\begin{bmatrix} 1 \\ 0 \\ 0 \\ bc \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \\ ac \end{bmatrix} \right) + y^2 dz \wedge dw \left(\begin{bmatrix} 1 \\ 0 \\ 0 \\ bc \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \\ ac \end{bmatrix} \right) \\
&= x^2 \begin{vmatrix} 0 & 1 \\ 0 & 0 \end{vmatrix} + y^2 \begin{vmatrix} 0 & 0 \\ bc & ac \end{vmatrix} \\
&= 0.
\end{aligned}
$$

Similarly we have

$$
\begin{aligned}
g = \ldots &= x^2 \begin{vmatrix} 1 & 0 \\ 0 & 1 \end{vmatrix} + y^2 \begin{vmatrix} 0 & 1 \\ ac & ab \end{vmatrix} \\
&= x^2(1) + y^2(-ac) \\
&= a^2 - ab^2c
\end{aligned}
$$

and

$$
\begin{aligned}
h = \ldots &= x^2 \begin{vmatrix} 0 & 0 \\ 1 & 0 \end{vmatrix} + y^2 \begin{vmatrix} 1 & 0 \\ ab & bc \end{vmatrix} \\
&= x^2(0) + y^2(bc) \\
&= b^3c.
\end{aligned}
$$

Thus

$$
T^*\phi \cdot \big(x^2 dy \wedge dz + y^2 dz \wedge dw\big) = (a^2 - ab^2c)db \wedge dc + (b^3c)dc \wedge da.
$$

Question 6.15 Fill in the ... for the two computations above.

Question 6.16 What is the pull-back of $dx \wedge dy \wedge dz \wedge dw$ under this same map?

Question 6.17 Suppose we have a map $\phi : \mathbb{R}^n_{x_1\cdots x_n} \longrightarrow \mathbb{R}^m_{\phi_1\cdots\phi_m}$ where $m > n$, $\phi = \big(\phi_1, \ldots, \phi_m\big)$ and ω is an n-form on $\mathbb{R}^m$. Show that

$$
T_p^*\phi \cdot \omega = \omega_{\phi(p)} \left(\frac{\partial \phi}{\partial x_1}, \ldots, \frac{\partial \phi}{\partial x_n} \right) dx_1 \wedge dx_2 \wedge \cdots \wedge dx_n
$$

where

$$
\frac{\partial \phi}{\partial x_i} = \begin{bmatrix} \frac{\partial \phi_1}{\partial x_i} \\ \frac{\partial \phi_2}{\partial x_i} \\ \vdots \\ \frac{\partial \phi_m}{\partial x_i} \end{bmatrix}.
$$

6.7 Some Useful Identities

We have discussed pull-backs in the context of volume forms pertaining to changes of variables, also called changes of coordinates. Suppose we have a change of coordinates $\phi : \mathbb{R}^n_{(x_1,\ldots,x_n)} \to \mathbb{R}^n_{(\phi_1,\ldots,\phi_n)}$ given by

$$
\phi(x_1, \ldots, x_n) = \Big(\phi_1(x_1, \ldots, x_n), \phi_2(x_1, \ldots, x_n), \ldots, \phi_n(x_1, \ldots, x_n)\Big)
$$

with a volume form on $\mathbb{R}^n_{(\phi_1,\dots,\phi_n)}$ given by $d\phi_1 \wedge d\phi_2 \wedge \cdots \wedge d\phi_n$. We found the pull-back of the volume form onto $\mathbb{R}^n_{(x_1,\dots,x_n)}$ by $T^*\phi \cdot \big(d\phi_1 \wedge d\phi_2 \wedge \cdots \wedge d\phi_n\big)$, which is given by the formula

$$T^*\phi \cdot \big(d\phi_1 \wedge d\phi_2 \wedge \cdots \wedge d\phi_n\big) = \begin{vmatrix} \frac{\partial \phi_1}{\partial x_1} & \frac{\partial \phi_1}{\partial x_2} & \cdots & \frac{\partial \phi_1}{\partial x_n} \\ \frac{\partial \phi_2}{\partial x_1} & \frac{\partial \phi_2}{\partial x_2} & \cdots & \frac{\partial \phi_2}{\partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial \phi_n}{\partial x_1} & \frac{\partial \phi_n}{\partial x_m} & \cdots & \frac{\partial \phi_n}{\partial x_n} \end{vmatrix} dx_1 \wedge dx_2 \wedge \cdots \wedge dx_n,$$

where $dx_1 \wedge dx_2 \wedge \cdots \wedge dx_n$ is the volume form on $\mathbb{R}^n_{(x_1,\dots,x_n)}$. If you review the proof of this identity in Sect. 6.3 you will see that the Jacobian matrix $T\phi$ basically "pulls through" the computation because of the nature of the definition of the wedgeproduct as the determinant of a matrix. This gave us the very useful computational formula above. However, this formula only works for the pull-back of a volume form by a mapping ϕ between two manifolds of the same dimension.

As mentioned in Sect. 6.6 there are not any nice formulas like this for the pull-backs of arbitrary k-forms. However, given an arbitrary k-form there is still a lot we can do with it. There are three very important identities that give the relation between the pull-back and sums of forms, the wedgeproduct, and exterior differentiation. Letting α and β be differential forms and c be a constant, these identities are

pull-back identities

1. $T^*\phi \cdot \Big(c\alpha + \beta\Big) = cT^*\phi \cdot \alpha + T^*\phi \cdot \beta,$
2. $T^*\phi \cdot \Big(\alpha \wedge \beta\Big) = T^*\phi \cdot \alpha \wedge T^*\phi \cdot \beta,$
3. $T^*\phi \cdot d\alpha = d\Big(T^*\phi \cdot \alpha\Big).$

Using more traditional notation, these three identities are usually encountered as

pull-back identities

1. $\phi^*\Big(c\alpha + \beta\Big) = c\phi^*\alpha + \phi^*\beta,$
2. $\phi^*\Big(\alpha \wedge \beta\Big) = \phi^*\alpha \wedge \phi^*\beta,$
3. $\phi^*\Big(d\alpha\Big) = d\Big(\phi^*\alpha\Big).$

You will very often see the third identity written without the input argument α as just the composition of functions $\phi^* \circ d = d \circ \phi^*$ or $\phi^* d = d\phi^*$. Notice that the first identity really says that the pull-back is linear, meaning that we have both

$$\phi^*\big(\alpha + \beta\big) = \phi^*(\alpha) + \phi^*(\beta) \qquad \text{and} \qquad \phi^*(c\alpha) = c\phi^*(\alpha).$$

The proof of the first identity is not difficult,

$$\begin{aligned}
&\phi^*\big(c\alpha + \beta\big)\Big(v_1, \cdots, v_k\Big) \\
&= \big(c\alpha + \beta\big)\Big(\phi_* v_1, \cdots, \phi_* v_k\Big) \\
&= (c\alpha)\Big(\phi_* v_1, \cdots, \phi_* v_k\Big) + \beta\Big(\phi_* v_1, \cdots, \phi_* v_k\Big) \\
&= c\Big(\alpha\Big(\phi_* v_1, \cdots, \phi_* v_k\Big)\Big) + \beta\Big(\phi_* v_1, \cdots, \phi_* v_k\Big) \\
&= c\Big(\phi^*\alpha\Big)\Big(v_1, \cdots, v_k\Big) + \phi^*\beta\Big(v_1, \cdots, v_k\Big) \\
&= \Big(c\phi^*\alpha + \phi^*\beta\Big)\Big(v_1, \cdots, v_k\Big).
\end{aligned}$$

Thus we have proved our first identity.

Now we turn our attention to the second identity, which is quite easy when we use one of the definitions for the wedgeproduct from Sect. 3.3.3,

$$\begin{aligned}
&\big(\phi^*\alpha \wedge \phi^*\beta\big)(v_1, \ldots, v_{k+l}) \\
&= \sum_{\substack{\sigma \text{ is a} \\ (k,\ell)-\text{shuffle}}} \operatorname{sgn}(\sigma)\phi^*\alpha\big(v_{\sigma(1)}, \ldots, v_{\sigma(k)}\big)\phi^*\beta\big(v_{\sigma(k+1)}, \ldots, v_{\sigma(k+\ell)}\big) \\
&= \sum_{\substack{\sigma \text{ is a} \\ (k,\ell)-\text{shuffle}}} \operatorname{sgn}(\sigma)\alpha\big(\phi_* v_{\sigma(1)}, \ldots, \phi_* v_{\sigma(k)}\big)\beta\big(\phi_* v_{\sigma(k+1)}, \ldots, \phi_* v_{\sigma(k+\ell)}\big) \\
&= \big(\alpha \wedge \beta\big)(\phi_* v_1, \ldots, \phi_* v_{k+\ell}) \\
&= \phi^*\big(\alpha \wedge \beta\big)(v_1, \ldots, v_{k+\ell}),
\end{aligned}$$

which gives us our second identity $\phi^*(\alpha \wedge \beta) = \phi^*\alpha \wedge \phi^*\beta$.

The third identity is a little more tricky. We first prove it for the special case of a zero-form f on $\mathbb{R}^n_{(\phi_1,\ldots,\phi_n)}$. Recall that a zero-form is simply a function, so df is a one-form on $\mathbb{R}^n_{(\phi_1,\ldots,\phi_n)}$. In this situation we have the following maps

$$\begin{gathered}
\bigwedge\nolimits^0 \Big(\mathbb{R}^n_{(x_1,\ldots,x_n)}\Big) \xleftarrow{\phi^*} \bigwedge\nolimits^0 \Big(\mathbb{R}^n_{(\phi_1,\ldots,\phi_n)}\Big) \\
T\mathbb{R}^n_{(x_1,\ldots,x_n)} \xrightarrow{\phi_*} T\mathbb{R}^n_{(\phi_1,\ldots,\phi_n)} \\
\mathbb{R}^n_{(x_1,\ldots,x_n)} \xrightarrow{\phi} \mathbb{R}^n_{(\phi_1,\ldots,\phi_n)} \xrightarrow{f} \mathbb{R}.
\end{gathered}$$

The way push-forwards of vectors and pull-back of the one-form df work in this situation is a little bit special due to the nature of the zero-form f. Let us first consider the case of $\phi_* v$, the push-forward of a vector $v \in T\mathbb{R}^n_{(x_1,\ldots,x_n)}$ by ϕ. We know $\phi_* v$ is a vector in $T\mathbb{R}^n_{(\phi_1,\ldots,\phi_n)}$ and as such it can be used to find the directional derivative of f, $\phi_* v[f]$. But this is exactly the same as finding the directional derivative of the function $f \circ \phi$ by the vector v, which means we have $v[f \circ \phi] = \phi_* v[f]$.

Question 6.18 Show $v[f \circ \phi] = \phi_* v[f]$ using an explicit calculation using the notation and formulas we have developed in this chapter.

Let us now consider the pull-back by ϕ of the zero-form $f \in \bigwedge^0 \Big(\mathbb{R}^n_{(\phi_1,\ldots,\phi_n)}\Big)$. Our previous definition of the pull-back of a k-form can no longer work since zero-forms do not eat vectors, so we will define the pull-back of a zero-form, that is, a functions, by the only way that makes sense. Pulling the function f back to the manifold $\mathbb{R}^n_{(x_1,\ldots,x_n)}$ is defined to be $f \circ \phi$, which means we have $\phi^* f = f \circ \phi$,

$$\begin{aligned}
\phi^* df(v) &= df\big(\phi_* v\big) && \text{definition of pull-back} \\
&= \phi_* v[f] && \text{definition of differential} \\
&= v[f \circ \phi] && \text{question 6.18} \\
&= v[\phi^* f] && \text{definition of pull-back of zero-form} \\
&= d(\phi^* f)(v) && \text{definition of differential.}
\end{aligned}$$

Writing without the vector v we have, for a zero-form f, $\phi^* df = d\phi^* f$, thereby proving the third identity in the special case of a zero-form.

Now we prove the third identity in full generality. Let $\alpha = \sum_I a_I dx^I = \sum a_I d\phi_{i_1} \wedge \cdots \wedge d\phi_{i_k}$, where $i_1 < \cdots < i_k$. Noting that the ϕ_i are nothing more than zero-forms we have $\phi^*(d\phi_i) = d(\phi^*\phi_i)$,

$$\begin{aligned}
\phi^*\alpha &= \sum (\phi^* a_I)\ \phi^* d\phi_{i_1} \wedge \ldots \wedge \phi^* d\phi_{i_k} \\
&= \sum (\phi^* a_I)\ d(\phi^*\phi_{i_1}) \wedge \ldots \wedge d(\phi^*\phi_{i_k})
\end{aligned}$$

which leads to

$$\begin{aligned}
d(\phi^*\alpha) &= d\left(\sum (\phi^* a_I)\, d(\phi^*\phi_{i_1}) \wedge \ldots \wedge d(\phi^*\phi_{i_k})\right) \\
&= \sum d(\phi^* a_I) \wedge d(\phi^*\phi_{i_1}) \wedge \ldots \wedge d(\phi^*\phi_{i_k}) \\
&= \sum \phi^* da_I \wedge \phi^* d\phi_{i_1} \wedge \ldots \wedge \phi^* d\phi_{i_k} \\
&= \phi^* \left(\sum da_I \wedge d\phi_{i_1} \wedge \cdots \wedge d\phi_{i_k}\right) \\
&= \phi^* \left(d \sum a_I \, d\phi_{i_1} \wedge \cdots \wedge d\phi_{i_k}\right) \\
&= \phi^*(d\alpha).
\end{aligned}$$

Thus we have our third identity, $d\phi^*\alpha = \phi^* d\alpha$.

6.8 Summary, References, and Problems

6.8.1 Summary

Consider a map between two manifolds, $f : M \to N$ where $f = (f_1, f_2, \ldots, f_n)$ and each component function f_i is differentiable at a point $p \in M$. If the coordinates of manifold M are given by $x_1, x_2, \ldots, x_n$, then the derivative of the map f at the point $p \in M$ is given by the traditional Jacobian matrix

$$T_p f = \begin{bmatrix} \left.\frac{\partial f_1}{\partial x_1}\right|_p & \left.\frac{\partial f_1}{\partial x_2}\right|_p & \cdots & \left.\frac{\partial f_1}{\partial x_n}\right|_p \\ \left.\frac{\partial f_2}{\partial x_1}\right|_p & \left.\frac{\partial f_2}{\partial x_2}\right|_p & \cdots & \left.\frac{\partial f_2}{\partial x_n}\right|_p \\ \vdots & \vdots & \ddots & \vdots \\ \left.\frac{\partial f_n}{\partial x_1}\right|_p & \left.\frac{\partial f_n}{\partial x_2}\right|_p & \cdots & \left.\frac{\partial f_n}{\partial x_n}\right|_p \end{bmatrix}.$$

Suppose we have some vector $v_p \in T_p M$. $T_p f \cdot v_p$ gives the linear approximation of the change in f as we move along v_p, which is another vector in $T_{f(p)}N$. Since we can find $T_p f : T_p M \to T_{f(p)}N$ at each point $p \in M$ where f is differentiable we end up with a mapping Tf from the tangent bundle of M to the tangent bundle of N. We say that the map $f : M \to N$ induces the map $Tf : TM \longrightarrow TN$. The mapping Tf pushes-forward vectors in $T_p M$ to vectors in $T_{f(p)}N$ and so is called the push-forward mapping. The push-forward map is also sometimes called the tangent mapping and is also denoted by f_* or Df.

Now suppose that we have a differential k-form ω on N. We define the differential k-form $T^* f \cdot \omega$ as the pull-back of ω to M where

Definition of Pull-Back of Differential Form	$\big(T^* f \cdot \omega\big)\big(v_1, v_2, \ldots, v_k\big) = \omega\big(Tf \cdot v_1, Tf \cdot v_2, \ldots, Tf \cdot v_k\big).$

The pull-back map is also sometimes called the cotangent map and is often denoted by f^*. Notice that when it comes to the pull-back map, it is indexed by the base point in the image and not the domain. This is quite unusual in mathematics, but is simply being the best way to keep track of base points,

$$\begin{gathered}
\bigwedge\nolimits^k_p(M) \xleftarrow{T_p^* f} \bigwedge\nolimits^k_{f(p)}(N) \\
T_p M \xrightarrow{T_p f} T_{f(p)} N \\
M \xrightarrow{f} N.
\end{gathered}$$

In general there is no single formula that one can use to find the pull-back of a differential form. However, in the case of the pull-back of a volume form using a change in basis given by $\phi : \mathbb{R}^n_{(x_1,\ldots,x_n)} \to \mathbb{R}^n_{(\phi_1,\ldots,\phi_n)}$ there is a single formula,

Pull-back of Volume Form

$$T^*\phi \cdot \Big(d\phi_1 \wedge d\phi_2 \wedge \cdots \wedge d\phi_n\Big) = \begin{vmatrix} \frac{\partial \phi_1}{\partial x_1} & \frac{\partial \phi_1}{\partial x_2} & \cdots & \frac{\partial \phi_1}{\partial x_n} \\ \frac{\partial \phi_2}{\partial x_1} & \frac{\partial \phi_2}{\partial x_2} & \cdots & \frac{\partial \phi_2}{\partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial \phi_n}{\partial x_1} & \frac{\partial \phi_n}{\partial x_2} & \cdots & \frac{\partial \phi_n}{\partial x_n} \end{vmatrix} dx_1 \wedge dx_2 \wedge \cdots \wedge dx_n.$$

Finally, some useful pull-back identities are

pull-back identities

1. $T^*\phi \cdot \Big(c\alpha + \beta\Big) = cT^*\phi \cdot \alpha + T^*\phi \cdot \beta,$
2. $T^*\phi \cdot \Big(\alpha \wedge \beta\Big) = T^*\phi \cdot \alpha \wedge T^*\phi \cdot \beta,$
3. $T^*\phi \cdot d\alpha = d\Big(T^*\phi \cdot \alpha\Big).$

6.8.2 *References and Further Reading*

Push-forwards of vector fields and pull-backs of forms are covered in almost any introductory book on manifold theory or differential geometry, see for example Bachman [4], O'Neill [36], Renteln [37], Darling [12], Tu [46], Hubbard and Hubbard [27], or Edwards [18]. Indeed, all of these sources were used. However, we have attempted to present the material from a very down-to-earth point of view via a number of concrete situations. In the process we attempt to make clear the relationship between both push-forwards and pull-backs and material and formulas encountered in vector calculus books such as Stewart [43] or Marsden and Hoffman [31].

6.8.3 *Problems*

Question 6.19 Let the mapping $\varphi : \mathbb{R}^2 \to \mathbb{R}^3$ be defined by $\varphi(u, v) = \Big(x(u, v), y(u, v), z(u, v)\Big) = \Big(u^2+v, 2u-v, u+v^3\Big)$. Find $T_p\varphi \cdot v_p$ for the following vectors,

a) $\begin{bmatrix} 2 \\ 3 \end{bmatrix}_{(2,1)}$, *b)* $\begin{bmatrix} -4 \\ 5 \end{bmatrix}_{(3,-2)}$, *c)* $\begin{bmatrix} 7 \\ -3 \end{bmatrix}_{(-3,-1)}$, *d)* $\begin{bmatrix} -3 \\ -7 \end{bmatrix}_{(-4,8)}$, *e)* $\begin{bmatrix} 3 \\ 1 \end{bmatrix}_{(3,-2)}$.

Question 6.20 Let the mapping $\varphi : \mathbb{R}^2 \to \mathbb{R}^2$ be defined by $\varphi(u, v) = \Big(x(u, v), y(u, v)\Big) = \big(u^2 + 1, uv\big)$. Find $T_p\varphi \cdot v_p$ for the vectors from Question 6.19.

Question 6.21 Let the mapping $\varphi : \mathbb{R}^2 \to \mathbb{R}^2$ be defined by $\varphi(u, v) = \Big(x(u, v), y(u, v)\Big) = \big(u^2 v, u - v\big)$. Find $T_p\varphi \cdot v_p$ for the vectors from Question 6.19.

Question 6.22 Let the mapping $\varphi : \mathbb{R}^2_{xy} \to \mathbb{R}^2_{uv}$ be defined by $\varphi(x, y) = \Big(u(x, y), v(x, y)\Big) = \big(x - y, xy\big)$. Find the pull-back of the volume form $du \wedge dv$. That is, find $T^*\varphi \cdot \big(du \wedge dv\big)$.

Question 6.23 Let the mapping $\varphi : \mathbb{R}^2_{xy} \to \mathbb{R}^2_{uv}$ be defined by $\varphi(x, y) = \Big(u(x, y), v(x, y)\Big) = \big(x + 2y, x - 2y\big)$. Find the pull-back of the volume form $du \wedge dv$. That is, find $T^*\varphi \cdot \big(du \wedge dv\big)$.

Question 6.24 Let the mapping $\varphi : \mathbb{R}^2_{uv} \to \mathbb{R}^2_{x,y}$ be defined by $\varphi(u,v) = \big(x(u,v), y(u,v)\big) = \big(u^2 - v^2, 2uv\big)$. Find the pull-back of the volume form $dx \wedge dy$. That is, find $T^*\varphi \cdot \big(dx \wedge dy\big)$.

Question 6.25 Let the mapping $\varphi : \mathbb{R}^2_{uv} \to \mathbb{R}^2_{x,y}$ be defined by $\varphi(u,v) = \big(x(u,v), y(u,v)\big) = \big(u^2 - v^2, u + v + 5\big)$. Find the pull-back of the volume form $dx \wedge dy$. That is, find $T^*\varphi \cdot \big(dx \wedge dy\big)$.

Question 6.26 Let $f : \mathbb{R}^2 \to \mathbb{R}^3$ be given by $f(u,v) = \big(u^2 + v, 2u - v, u + v^3\big)$ and let $\alpha = (x + y + 1)\, dx + (3z - y)\, dy + x\, dz$ be a one-form on $\mathbb{R}^3$. Find the general formula for $T^* f \cdot \alpha$.

Question 6.27 Let $f : \mathbb{R}^2 \to \mathbb{R}^3$ be given by $f(u,v) = \big(u^2 + v, 2u - v, u + v^3\big)$ and let $\alpha = -dx \wedge dy + z\, dy \wedge dz + (y + 2)\, dz \wedge dx$ be a one-form on $\mathbb{R}^3$. Find the general formula for $T^* f \cdot \alpha$.

Question 6.28 Let $f_1 = y^2 - x^2$ and $f_2 = 3xy$ be functions on $\mathbb{R}^2$ and let $\alpha_1 = x\, dx + xy\, dy$, $\alpha_2 = -3y\, dx + 2x\, dy$, and $\alpha_3 = (x^2 + y^2)\, dy$ be one-forms on $\mathbb{R}^2$. Given the mapping $\varphi : \mathbb{R} \to \mathbb{R}^2$ defined by $\varphi(t) = \big(x(t), y(t)\big) = \big(t, t^2\big)$ find

a) $T^*\varphi \cdot f_1$, *b)* $T^*\varphi \cdot f_2$, *c)* $T^*\varphi \cdot \alpha_1$, *d)* $T^*\varphi \cdot \alpha_2$, *e)* $T^*\varphi \cdot \alpha_3$.

Question 6.29 For the mapping and functions in Question 6.28 show that

a) $T^*\varphi \cdot \big(df_1\big) = d\big(T^*\varphi \cdot f_1\big)$, *b)* $T^*\varphi \cdot \big(df_2\big) = d\big(T^*\varphi \cdot f_2\big)$.

Question 6.30 Let $f_1 = y^2 - x^2$ and $f_2 = 3xy$ be functions on $\mathbb{R}^2$, $\alpha_1 = x\, dx + xy\, dy$, $\alpha_2 = -3y\, dx + 2x\, dy$, and $\alpha_3 = (x^2 + y^2)\, dy$ be one-forms on $\mathbb{R}^2$, and $\beta_1 = (xy - x)\, dx \wedge dy$ and $\beta_2 = (x + y^2)\, dx \wedge dy$ be two-forms on $\mathbb{R}^2$. Given the mapping $\varphi : \mathbb{R}^2 \to \mathbb{R}^2$ defined by $\varphi(u,v) = \big(x(u,v), y(u,v)\big) = \big(u^2 + 1, uv\big)$ find

a) $T^*\varphi \cdot f_1$, *b)* $T^*\varphi \cdot f_2$, *c)* $T^*\varphi \cdot \alpha_1$, *d)* $T^*\varphi \cdot \alpha_2$, *e)* $T^*\varphi \cdot \alpha_3$, *f)* $T^*\varphi \cdot \beta_1$, *g)* $T^*\varphi \cdot \beta_2$.

Question 6.31 For the mapping, functions, and one-forms in Question 6.30 show that

a) $T^*\varphi \cdot \big(df_1\big) = d\big(T^*\varphi \cdot f_1\big)$, *b)* $T^*\varphi \cdot \big(df_2\big) = d\big(T^*\varphi \cdot f_2\big)$, *c)* $T^*\varphi \cdot \big(d\alpha_1\big) = d\big(T^*\varphi \cdot \alpha_1\big)$, *d)* $T^*\varphi \cdot \big(d\alpha_2\big) = d\big(T^*\varphi \cdot \alpha_2\big)$, *e)* $T^*\varphi \cdot \big(d\alpha_3\big) = d\big(T^*\varphi \cdot \alpha_3\big)$.

Question 6.32 For the mapping and one-forms in Question 6.30 show that

a) $T^*\varphi \cdot \big(\alpha_1 \wedge \alpha_2\big) = T^*\varphi \cdot \alpha_1 \wedge T^*\varphi \cdot \alpha_2$, *b)* $T^*\varphi \cdot \big(\alpha_1 \wedge \alpha_3\big) = T^*\varphi \cdot \alpha_1 \wedge T^*\varphi \cdot \alpha_3$, *c)* $T^*\varphi \cdot \big(\alpha_2 \wedge \alpha_3\big) = T^*\varphi \cdot \alpha_2 \wedge T^*\varphi \cdot \alpha_3$.

Question 6.33 Repeat Question 6.30 for the mapping $\varphi : \mathbb{R}^2 \to \mathbb{R}^2$ defined by $\varphi(u,v) = \big(u^2 v, u - v\big)$.

Question 6.34 Repeat Question 6.31 for the mapping $\varphi : \mathbb{R}^2 \to \mathbb{R}^2$ defined by $\varphi(u,v) = \big(u^2 v, u - v\big)$.

Question 6.35 Repeat Question 6.32 for the mapping $\varphi : \mathbb{R}^2 \to \mathbb{R}^2$ defined by $\varphi(u,v) = \big(u^2 v, u - v\big)$.

Chapter 7
Changes of Variables and Integration of Forms

Integration is one of the most a fundamental concepts in mathematics. In calculus you began by learning how to integrate one-variable functions on $\mathbb{R}$. Then, you learned how to integrate two- and three-variable functions on $\mathbb{R}^2$ and $\mathbb{R}^3$. After this you learned how to integrate a function after a change-of-variables, and finally in vector calculus you learned how to integrate vector fields along curves and over surfaces. It turns out that differential forms are actually very nice things to integrate. Indeed, there is an intimate relationship between the integration of differential forms and the change-of-variables formulas you learned in calculus. In section one we define the integral of a two-form on $\mathbb{R}^2$ in terms of Riemann sums. Integrals of n-forms on $\mathbb{R}^n$ can be defined analogously. We then use the ideas from Chap. 6 along with the Riemann sum procedure to derive the change of coordinates formula from first principles. In section two we look carefully at a simple change of coordinates example. Section three continues by looking at changes from Cartesian coordinates to polar, cylindrical, and spherical coordinates. Finally in section four we consider a more general setting where we see how we can integrate arbitrary one- and two-forms on parameterized one- and two-dimensional surfaces.

7.1 Integration of Differential Forms

We begin by looking at the fundamental definition of integration in terms of Riemann sums. We will start out by finding the volume under the graph of a function, basically what you have seen in calculus. For now we assume traditional Cartesian coordinates on the Euclidian manifold $\mathbb{R}^2$. Let R be a region on the plane and let $f(x, y)$ be the function whose graph we are finding the volume under. From calculus class you should recall the following steps:

(1) Choose a lattice of evenly spaced points $\{x_i\}_{i=1}^n$ and $\{y_j\}_{j=1}^m$ that covers the region R. This lattice gives the points $(x_i, y_j) \in \mathbb{R}^2_{xy}$.
(2) Let $\Delta x_i = x_{i+1} - x_i$ and $\Delta y_j = y_{j+1} - y_j$.
(3) For each $i = 1, \ldots, n$ and $j = 1, \ldots, m$ find $f(x_i, y_j)\Delta x_i \Delta y_j$.
(4) Sum over i and j; $\sum_i \sum_j f(x_i, y_j)\Delta x_i \Delta y_j$.
(5) Take the limit of this double sum as $n, m \to \infty$, that is, as $\Delta x_i, \Delta y_j \to 0$, and define this to be $\int\int_R f(x, y)\, dx\, dy$.

In essence, we are approximating the volume under the graph of $f(x, y)$ and above the rectangle $\Delta x_i \Delta y_j$, Fig. 7.1, by the product $f(x_i, y_j)\Delta x_i \Delta y_j$, Fig. 7.2. Of course, this is only the theoretical definition, this is not how we actually compute integrals, we have a whole slew of techniques at our disposal for integration which you learned in calculus.

Now, we are essentially going to copy both the idea and the procedure in a way that is more appropriate for the differential forms setting, thereby giving a theoretical definition for the integral of a differential form $\alpha = f(x, y)dx \wedge dy$ on a region $R \subset \mathbb{R}^2$. However, we caution that the exact procedure we are giving here implicitly relies on the fact that we are still in the Euclidian manifold setting $\mathbb{R}^2$. A more general definition for $\mathbb{R}^n$ is certainly possible, but it does require greater attention to notation, which would obscure the fundamental ideas.

(1) Choose a lattice of evenly spaced points $\{x_i\}_{i=1}^n$ and $\{y_j\}_{j=1}^m$ that covers the region R. This lattice gives the points $(x_i, y_j) \in \mathbb{R}^2_{xy}$.
(2) For each i, j define the vectors

J. P. Fortney, *A Visual Introduction to Differential Forms and Calculus on Manifolds*,
https://doi.org/10.1007/978-3-319-96992-3_7

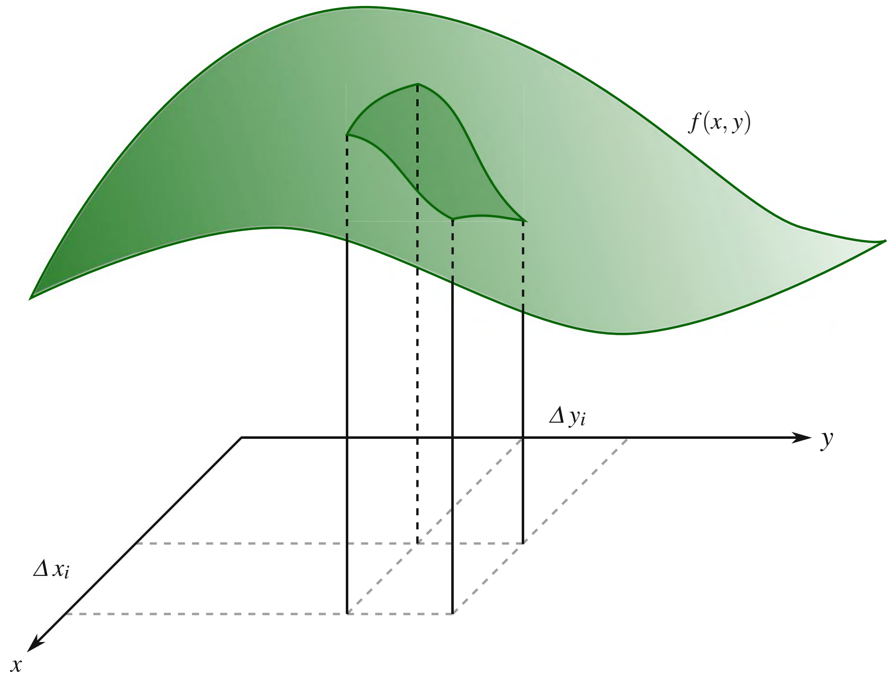


Fig. 7.1 The volume under the graph of $f(x, y)$ and above the rectangle $\Delta x_i \Delta y_j$

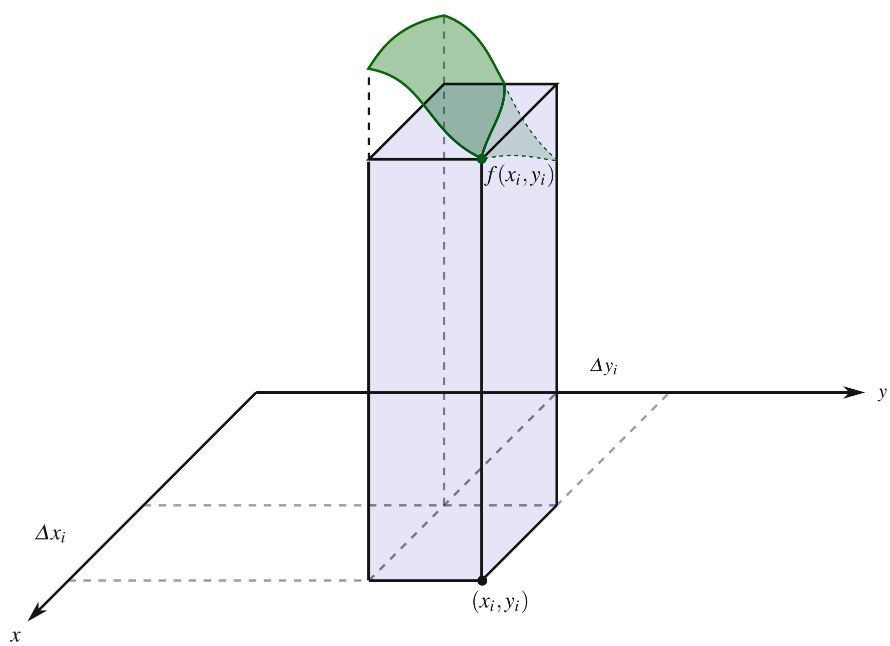


Fig. 7.2 The approximate volume under the graph of $f(x, y)$ and above the rectangle $\Delta x_i \Delta y_j$ given by the product $f(x_i, y_j)\Delta x_i \Delta y_j$, shown in blue

$$V_{i,j}^1 = \begin{bmatrix} x_{i+1} - x_i \\ 0 \end{bmatrix}_{(x_i, y_j)},$$

$$V_{i,j}^2 = \begin{bmatrix} 0 \\ y_{j+1} - y_j \end{bmatrix}_{(x_i, y_j)}.$$

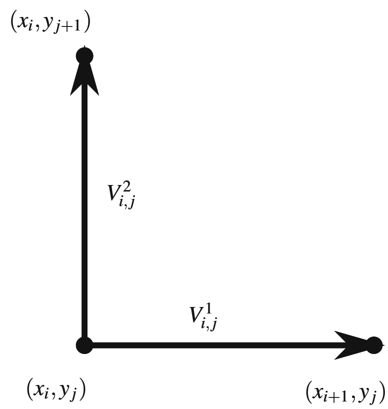


Fig. 7.3 The vectors $V^1_{i,j}$ and $V^2_{i,j}$. Despite the way they were defined we can consider $V^1_{i,j}, V^2_{i,j} \in T_{(x_i,y_j)}\mathbb{R}^2$

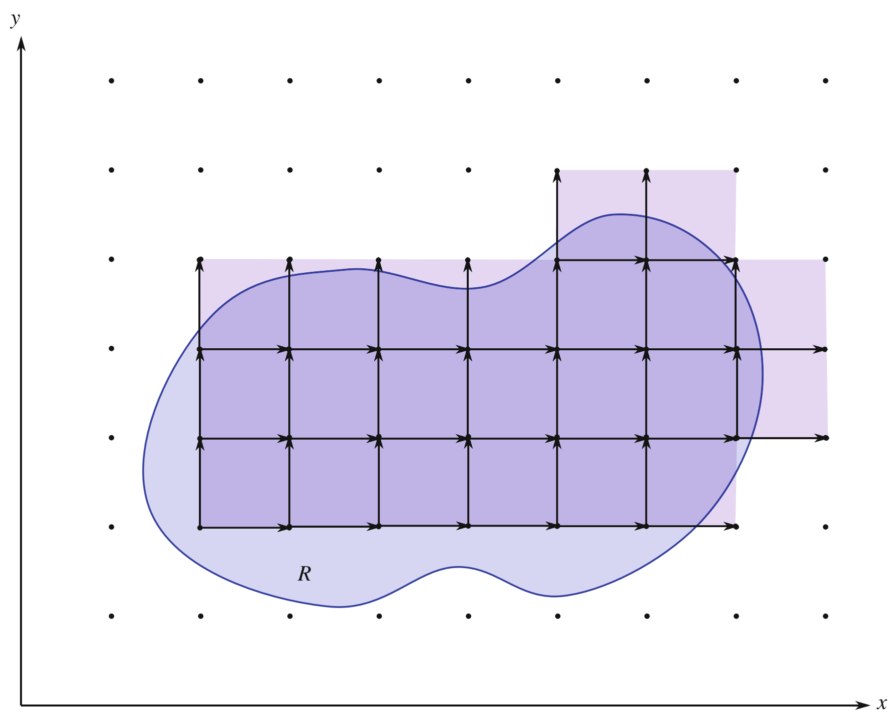


Fig. 7.4 A region $R \subset \mathbb{R}^2$ shown with the lattice of evenly spaced points $\{x_i\}_{i=1}^n$ and $\{y_j\}_{j=1}^m$ along with the vectors $V^1_{i,j}$ and $V^2_{i,j}$ for every point in the region R. As the lattice points get closer together, that is, as $|x_{i+1} - x_i|, |y_{j+1} - y_j| \to 0$, the region R is more accurately approximated

Note that we can view the vectors $V^1_{i,j}$ and $V^2_{i,j}$ as elements of the tangent space $T_{(x_i,y_j)}\mathbb{R}^2$. See Fig. 7.3. This is where we are implicitly assuming that we are in the Euclidian setting and can make this identification. Thus, we have something like Fig. 7.4.

(3) For each i, j compute $f(x_i, y_j)dx \wedge dy(V^1_{i,j}, V^2_{i,j})$.

(4) Sum over i and j; $\sum_i \sum_j f(x_i, y_j)dx \wedge dy(V^1_{i,j}, V^2_{i,j})$.

(5) Take the limit of this double sum as $m, n \to \infty$, that is, as $|x_{i+1} - x_i|, |y_{j+1} - y_j| \to 0$, and define this to be $\int\int_R f dx \wedge dy$. That is,

$$\int\int_R f dx \wedge dy \equiv \lim_{m,n\to\infty} \sum_{i=1}^{n} \sum_{j=1}^{m} f(x_i, y_j)dx \wedge dy(V^1_{i,j}, V^2_{i,j}).$$

Compare and contrast these two procedures. They are essentially the same, except for how areas are found. In step two of the first procedure we simply find the numbers Δx_i and Δy_j and in the second procedure we find the vectors $V^1_{i,j}$ and $V^2_{i,j}$. Then, in step three of the first procedure we found the area of the rectangle in the manifold by simply finding the product $\Delta x_i \Delta y_j$ whereas in the second procedure we found the area by using the area form $dx \wedge dy$ and the vectors $V^1_{i,j}, V^2_{i,j}$, namely, we found $dx \wedge dy(V^1_{i,j}, V^2_{i,j})$. These are, in essence, the only differences between the two procedures. But notice, in the first procedure we defined the integral $\int\int_R f(x,y)dxdy$ whereas in the second procedure we defined the integral $\int\int_R f(x,y)dx \wedge dy$. It is because of this that one will sometimes see the integral of the differential form $f(x,y)dx \wedge dy$ to simply be defined as

$$\int\int_R f(x,y)dx \wedge dy \equiv \int\int_R f(x,y)dxdy$$

instead of being defined in terms of the second procedure using Riemann sums. The only problem here is that the first procedure and its associated integral $\int\int_R f(x,y)dxdy$ does not keep track of orientation whereas the second procedure and its associated integral $\int\int_R f(x,y)dx \wedge dy$ does.

Question 7.1 Using the five steps above as a guide, define the integral of a differential form $\alpha = f(x_1, x_2, \ldots, x_n)dx_1 \wedge dx_2 \wedge \cdots \wedge dx_n$ on a region $R \subset \mathbb{R}^n$ using Riemann sums.

Now, suppose we find that taking the integral $\int_R f(x_1, \ldots, x_n)dx_1 \wedge \cdots \wedge dx_n$, or equivalently the integral $\int_R f(x_1, \ldots, x_n)dx_1 \cdots dx_n$, is difficult but we notice that it would become easier by doing an appropriate change of coordinates

$$\begin{aligned} \mathbb{R}^n_{x_1x_2\cdots x_n} &\xrightarrow{\phi} \mathbb{R}^n_{\phi_1\phi_2\cdots\phi_n} \\ (x_1, x_2, \ldots, x_n) &\longmapsto (\phi_1, \phi_2, \ldots, \phi_n). \end{aligned}$$

We want to develop this change of coordinates formula from first principles, that is, from the Riemann sum definition, in the context of differential forms. In other words, we are going to show that

$$\int_R f\ dx_1 \wedge \cdots \wedge dx_n = \int_{\phi(R)} f \circ \phi^{-1}\ \ T^*\phi^{-1} \cdot \big(dx_1 \wedge \cdots \wedge dx_n\big).$$

By doing the change of coordinates ϕ we map $R \subset \mathbb{R}^n_{x_1x_2\cdots x_n}$ to $\phi(R) \subset \mathbb{R}^n_{\phi_1\phi_2\cdots\phi_n}$. We also need to assume that ϕ has an inverse ϕ^{-1}.

(1) Choose a lattice of points $\big\{(x_{i_1}, \ldots, x_{i_n})\big\}$ on $\mathbb{R}^n_{x_1x_2\cdots x_n}$. This in turn gives a lattice of points

$$\Big\{\big(\phi_1(x_{i_1}, \ldots, x_{i_n}), \ldots, \phi_n(x_{i_1}, \ldots, x_{i_n})\big)\Big\} = \big\{(\phi_1, \ldots, \phi_n)\big\}$$

in the new coordinate system on $\mathbb{R}^n_{\phi_1\phi_2\cdots\phi_n}$, see Fig. 7.5.

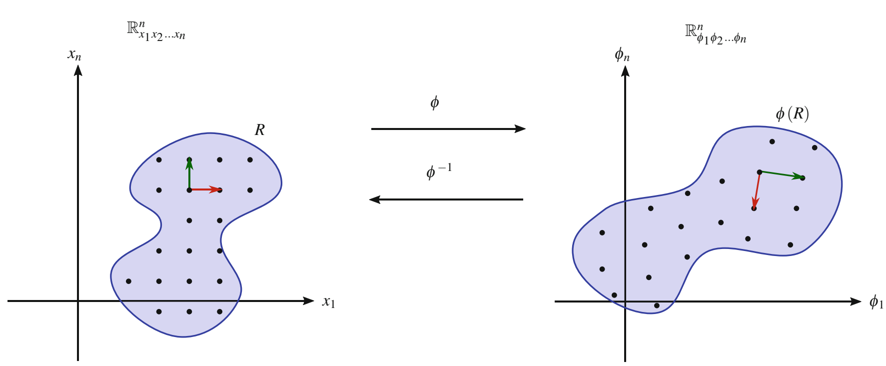


Fig. 7.5 The lattice $\{(x_{i_1}, \ldots, x_{i_n})\}$ in $\mathbb{R}^n_{x_1x_2\cdots x_n}$ produces the lattice $\{(\phi_1, \ldots, \phi_n)\}$ in $\mathbb{R}^n_{\phi_1\phi_2\cdots\phi_n}$. The push-forwards of two vectors are shown

(2) For each $i_1, \ldots, i_n$ define the vectors

$$V^1_{i_1\cdots i_n} = \begin{bmatrix} x_{i_1+1} - x_{i_1} \\ 0 \\ \vdots \\ 0 \end{bmatrix}_{(x_{i_1}, x_{i_2}, \ldots, x_{i_n})},$$

$$\vdots$$

$$V^n_{i_1\cdots i_n} = \begin{bmatrix} 0 \\ \vdots \\ 0 \\ x_{i_n+1} - x_{i_n} \end{bmatrix}_{(x_{i_1}, x_{i_2}, \ldots, x_{i_n})}.$$

Note that since the manifold is a vector space which is isomorphic to the tangent space we can claim that the vectors $V^1_{i_1\cdots i_n}, \ldots, V^n_{i_1\cdots i_n} \in T_{(x_{i_1},\ldots,x_{i_n})}\mathbb{R}^n_{x_1\cdots x_n}$. The volume of the parallelepiped spanned by the vectors $V^1_{i_1\cdots i_n}, \ldots, V^n_{i_1\cdots i_n}$ is given by

$$dx_1 \wedge \cdots \wedge dx_n(V^1_{i_1\cdots i_n}, \ldots, V^n_{i_1\cdots i_n}).$$

(3) The $(n+1)$-dimensional volume over the n-dimensional parallelepiped and under the graph of f is approximated by

$$f(x_{i_1}, \ldots x_{i_n})\, dx_1 \wedge \cdots \wedge dx_n(V^1_{i_1\cdots i_n}, \ldots, V^n_{i_1\cdots i_n}).$$

Next, at each lattice point $p = \{(x_{i_1}, \ldots, x_{i_n})\} \in \mathbb{R}^n_{x_1\cdots x_n}$ the push-forwards of the vectors $V^1_{i_1\cdots i_n}, \ldots, V^n_{i_1\cdots i_n}$ at that point are

$$T_p\phi \cdot V^1_{i_1\cdots i_n}, \ldots, T_p\phi \cdot V^n_{i_1\cdots i_n} \in T_{\phi(p)}\mathbb{R}^n_{\phi_1\phi_2\cdots\phi_n}.$$

Also, note that for each i, $T\phi^{-1} \cdot T\phi \cdot V^i_{i_1\cdots i_n} = V^i_{i_1\cdots i_n}$ so, by definition, we have

$$\begin{aligned} & T^*\phi^{-1} \cdot \big(dx_1 \wedge \cdots \wedge dx_n\big)\Big(T\phi \cdot V^1_{i_1\cdots i_n}, \ldots, T\phi \cdot V^n_{i_1\cdots i_n}\Big) \\ &= dx_1 \wedge \cdots \wedge dx_n\Big(T\phi^{-1} \cdot T\phi \cdot V^1_{i_1\cdots i_n}, \ldots, T\phi^{-1} \cdot T\phi \cdot V^n_{i_1\cdots i_n}\Big) \\ &= dx_1 \wedge \cdots \wedge dx_n\Big(V^1_{i_1\cdots i_n}, \ldots, V^n_{i_1\cdots i_n}\Big). \end{aligned}$$

Next, we notice that $f(x_{i_1}, \ldots x_{i_n}) = f \circ \phi^{-1} \circ \phi(x_{i_1}, \ldots x_{i_n})$. Combining this we have that

$$\begin{aligned} & f(x_{i_1}, \ldots x_{i_n})\, dx_1 \wedge \cdots \wedge dx_n\Big(V^1_{i_1\cdots i_n}, \ldots, V^n_{i_1\cdots i_n}\Big) \\ &= f \circ \phi^{-1} \circ \phi(x_{i_1}, \ldots x_{i_n})\, T^*\phi^{-1} \cdot \big(dx_1 \wedge \cdots \wedge dx_n\big)\Big(T\phi \cdot V^1_{i_1\cdots i_n}, \ldots, T\phi \cdot V^n_{i_1\cdots i_n}\Big) \\ &= f \circ \phi^{-1}(\phi_1, \ldots, \phi_n)\, T^*\phi^{-1} \cdot \big(dx_1 \wedge \cdots \wedge dx_n\big)\Big(T\phi \cdot V^1_{i_1\cdots i_n}, \ldots, T\phi \cdot V^n_{i_1\cdots i_n}\Big). \end{aligned}$$

Notice that in the last equality above we wrote

$$\phi(x_{i_1}, \ldots x_{i_n}) = \big(\phi_1(x_{i_1}, \ldots x_{i_n}), \ldots, \phi_n(x_{i_1}, \ldots x_{i_n})\big) = (\phi_1, \ldots, \phi_n).$$

(4) Now we sum over all $i_1, \ldots, i_n$ to get

$$\sum_{i_1,\ldots,i_n} f(x_{i_1}, \ldots x_{i_n})\, dx_1 \wedge \cdots \wedge dx_n\Big(V^1_{i_1\cdots i_n}, \ldots, V^n_{i_1\cdots i_n}\Big)$$
$$= \sum_{i_1,\ldots,i_n} f \circ \phi^{-1}(\phi_1, \ldots, \phi_n)\; T^*\phi^{-1} \cdot \big(dx_1 \wedge \cdots \wedge dx_n\big)\Big(T\phi \cdot V^1_{i_1\cdots i_n}, \ldots, T\phi \cdot V^n_{i_1\cdots i_n}\Big).$$

(5) Taking the limit as $\left|x_{i_j+1} - x_{i_j}\right| \to 0$ for $j = 1, \ldots, n$ we define

$$\int_R f(x_1, \ldots, x_n)\, dx_1 \wedge \cdots \wedge dx_n$$
$$= \lim \sum_{i_1,\ldots,i_n} f(x_{i_1}, \ldots x_{i_n})\, dx_1 \wedge \cdots \wedge dx_n\Big(V^1_{i_1\cdots i_n}, \ldots, V^n_{i_1\cdots i_n}\Big)$$

and

$$\int_{\phi(R)} f \circ \phi^{-1}(\phi_1, \ldots, \phi_n)\;\; T^*\phi^{-1} \cdot \big(dx_1 \wedge \cdots \wedge dx_n\big)$$
$$= \lim \sum_{i_1,\ldots,i_n} f \circ \phi^{-1}(\phi_1, \ldots, \phi_n)\;\; T^*\phi^{-1} \cdot \big(dx_1 \wedge \cdots \wedge dx_n\big)\Big(T\phi \cdot V^1_{i_1\cdots i_n}, \ldots, T\phi \cdot V^n_{i_1\cdots i_n}\Big).$$

Combining step (4) with these definitions gives us the following equality:

Change of coordinates formula	$\displaystyle\int_R f(x_1, \ldots, x_n)\;\; dx_1 \wedge \ldots \wedge dx_n = \int_{\phi(R)} f \circ \phi^{-1}(\phi_1, \ldots, \phi_n)\;\; T^*\phi^{-1} \cdot (dx_1 \wedge \ldots \wedge dx_n).$

Let us take a closer look at this equality. The left hand side is fairly straight forward; the integral takes place in $x_1 \cdots x_n$-coordinates and we are integrating the function $f(x_1, \ldots, x_n)$ over the region R using the volume form $dx_1 \wedge \cdots \wedge dx_n$ associated with the $x_1 \cdots x_n$-coordinates. Now, let's unpack the right hand side of this equality. First, we have the coordinate transformation $\phi : \mathbb{R}^n \longrightarrow \mathbb{R}^n$ given by

$$\mathbb{R}^n_{x_1x_2\cdots x_n} \xrightarrow{\;\phi\;} \mathbb{R}^n_{\phi_1\phi_2\cdots\phi_n}$$
$$(x_1, x_2, \ldots, x_n) \longmapsto (\phi_1, \phi_2, \ldots, \phi_n)$$

which gives the picture in Fig. 7.6. The region we are integrating over in $\mathbb{R}^n_{x_1x_2\cdots x_n}$ is R and the region we are integrating over in $\mathbb{R}^n_{\phi_1\phi_2\cdots\phi_n}$ is its image $\phi(R)$. The function $f : \mathbb{R}^n \to \mathbb{R}$ is a function of the variables $x_1, x_2, \cdots x_n$ whereas the function $f \circ \phi^{-1} : \mathbb{R}^n \to \mathbb{R}$ is a function of the variables $\phi_1, \phi_2, \cdots \phi_n$. Finally, $dx_1 \wedge \cdots \wedge dx_n$ is the area form on $\mathbb{R}^n_{x_1x_2\cdots x_n}$. But notice that the form on the right hand side is $T^*\phi^{-1} \cdot \big(dx_1 \wedge \cdots \wedge dx_n\big)$, the pull-back of the area form $dx_1 \wedge \cdots \wedge dx_n$ by ϕ^{-1} and NOT the area form $d\phi_1 \wedge \cdots \wedge d\phi_n$. This is an essential point when doing changes of variables.

There is now one final issue we need to mention briefly. Whether or not you have taken an introductory analysis course and have seen the proof of Fubini's theorem, or have even heard of Fubini's theorem before, you are doubtless aware of its consequences. Suppose we have a rectangular region $R = \{a \le x \le b, c \le y \le d\}$ and a continuous function $f : R \to \mathbb{R}$, then Fubini's theorem states

$$\int_a^b \int_c^d f(x, y)\;\; dy\, dx = \int_c^d \int_a^b f(x, y)\;\; dx\, dy.$$

Essentially it says we can change the order of integration. This actually requires our function f be somehow "nice." We do not want to get into the technical details here (we leave that for your analysis course) but suffice it to say that most function you will be dealing with are in fact "nice." However, in the context of differential forms we of course have $dx \wedge dy = -dy \wedge dx$

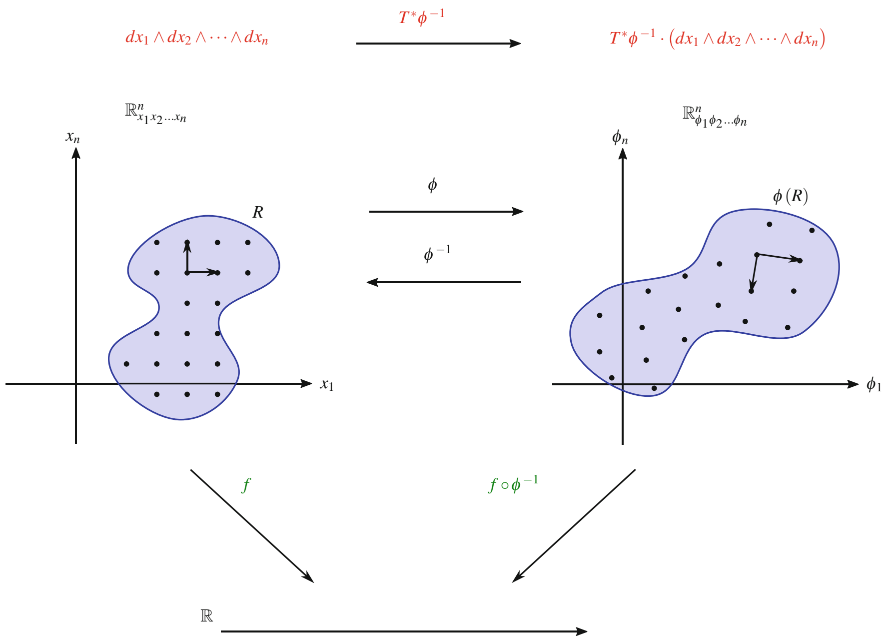


Fig. 7.6 When we change variables we have to change three things in our integral, the region (blue) the function (green), and the volume form (red). The vectors on the left are pushed forward by $T\phi$ to the vectors on the right

so we would have

$$\int_a^b \int_c^d f(x, y) \; dy \wedge dx = -\int_c^d \int_a^b f(x, y) \; dx \wedge dy,$$

unless of course you also change the orientation of the rectangle R as well. This issue will come up again later, but we will not spend a great deal of time on it. In general, integrating with differential forms takes into account orientation, which was not done in calculus. Thus, your answer may be different by a sign.

7.2 A Simple Example

We now apply what we have just done by talking a look at the role that volume forms play in integration during changes of variable by looking at a simple example that uses the same coordinate change that we encountered in Sect. 6.1. We will begin by integrating a function over a given region in $\mathbb{R}^2_{xy}$. Then we will apply the change-of-variables formula and perform the integration on $\mathbb{R}^2_{uv}$. We will also make implicit use of the following identification,

$$\int\int_R f(x, y) \, dx \, dy \equiv \int\int_R f(x, y) \, dx \wedge dy.$$

Suppose we want to integrate the function $f(x, y) = x$ over the region in $\mathbb{R}^2_{xy}$ bounded by the lines $x = 0$, $y = 0$, and $y = x - 2$, shown in Fig. 7.7. We will first integrate with respect to x and then with respect to y,

$$\begin{aligned} \int_{-2}^{0} \int_{y+2}^{0} x \, dx \, dy &= \int_{-2}^{0} \left[\frac{x^2}{2} \right]_{y+2}^{0} dy \\ &= \int_{-2}^{0} \frac{-(y+2)^2}{2} \, dy \end{aligned}$$

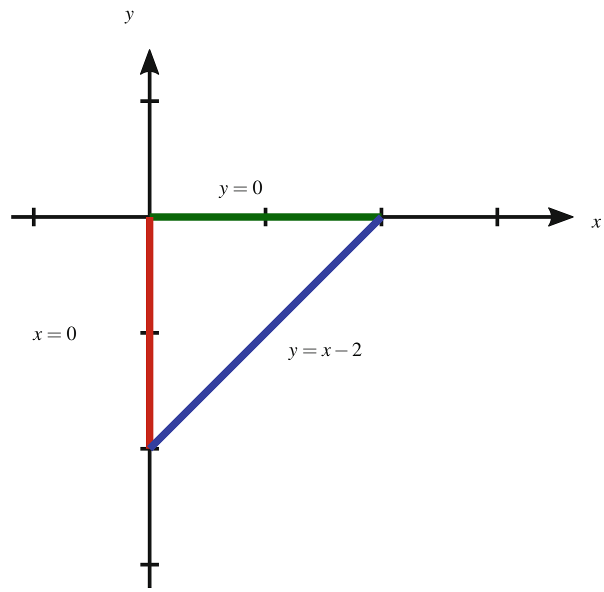


Fig. 7.7 The region in $\mathbb{R}^2$ bounded by the lines $x = 0$, $y = 0$, and $y = x - 2$

$$= \int_{-2}^{0} -\left(\frac{y^2 + 4y + 4}{2}\right) \, dy$$

$$= -\left[\frac{y^3}{6} + y^2 + 2y\right]_{-2}^{0}$$

$$= \frac{-4}{3}.$$

Now we want to do the same integral, only in different coordinates. To do this we must make use of the change of coordinates formula that we derived in Sect. 7.1,

$$\int_R f(x_1, \ldots, x_n) \; dx_1 \wedge \ldots \wedge dx_n = \int_{\phi(R)} f \circ \phi^{-1}(\phi_1, \ldots, \phi_n) \; T^*\phi^{-1} \cdot (dx_1 \wedge \ldots \wedge dx_n).$$

For our particular example the general change of coordinates formula simplifies to

$$\int\int_R f(x, y) \; dx \wedge dy = \int\int_{\phi(R)} f \circ \phi^{-1}(u, v) \; T^*\phi^{-1} \cdot (dx \wedge dy).$$

All of this comes down to the following: when we change variables we have to change three things in our integral, the region (blue), the function (green), and the volume form (red). See Fig. 7.8.

Returning to our example, we will change coordinates to the uv-coordinate systems defined by $u = x + y$ and $v = x - y$. In other words, we will use the coordinate transformation $\phi : \mathbb{R}^2_{xy} \to \mathbb{R}^2_{uv}$ given by $\phi(x, y) = \big(u(x, y), v(x, y)\big) = (x + y, x - y)$. We consider each of the three things we need to change. First we will consider the region we are integrating over in $\mathbb{R}^2_{uv}$. The coordinate transformations $u = x + y$ and $v = x - y$ can be rewritten as $x = \frac{u+v}{2}$ and $y = \frac{u-v}{2}$ which means the under the transformations the equations that define the borders become

$$x = 0 \Longrightarrow u = -v,$$

$$y = 0 \Longrightarrow u = v,$$

$$y = x - 2 \Longrightarrow v = 2.$$

Thus the image of our region in $\mathbb{R}^2_{uv}$ is shown in Fig. 7.9.

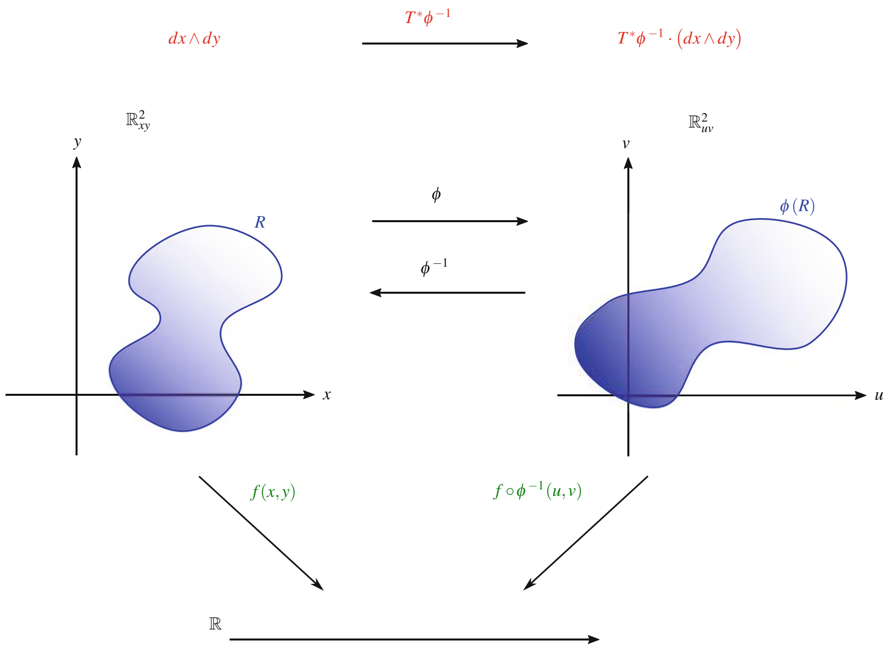


Fig. 7.8 When we change variables we have to change three things in our integral, the region (blue) the function (green), and the volume form (red)

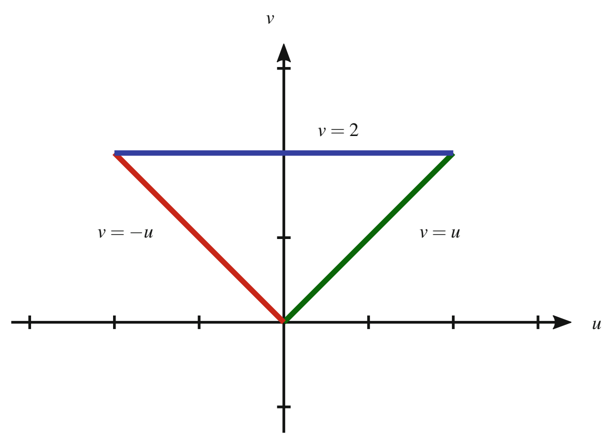


Fig. 7.9 The image in $\mathbb{R}^2_{uv}$ of the region shown in Fig. 7.7 under the transformation $\phi(x, y) = \big(u(x, y), v(x, y)\big) = (x + y, x - y)$

Second, we will consider the function. The function $f(x, y) = x$ that we want to integrate gets transformed into $f \circ \phi^{-1}(u, v) = \frac{u+v}{2}$. Third, we need to consider how the volume form changes. Recall that the mapping ϕ^{-1} induces the pullback mapping $T^*\phi^{-1}$,

$$\begin{gathered}
\mathbb{R}^2_{uv} \xrightarrow{\phi^{-1}} \mathbb{R}^2_{xy} \\
(u, v) \longmapsto (x, y) \\
\bigwedge\left(\mathbb{R}^2_{uv}\right) \xleftarrow{T^*\phi^{-1}} \bigwedge\left(\mathbb{R}^2_{xy}\right) \\
T^*\phi^{-1} \cdot \big(dx \wedge dy\big) \longleftarrow dx \wedge dy.
\end{gathered}$$

In Chap. 6 we found $T^*\phi^{-1} \cdot \big(dx \wedge dy\big) = \frac{-1}{2} du \wedge dv$.

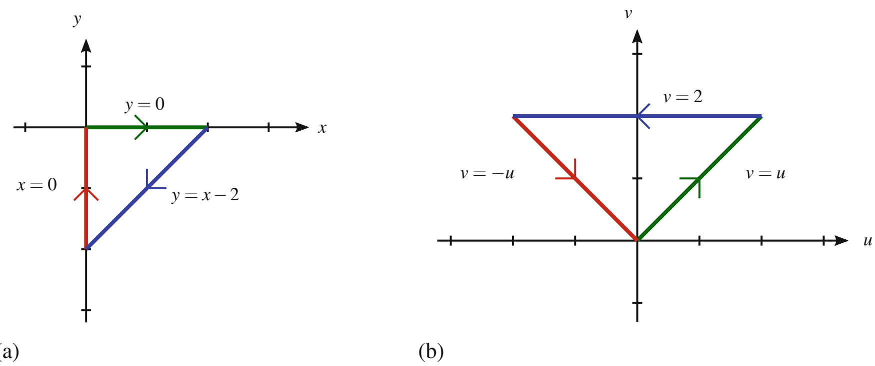


Fig. 7.10 The affect of the mapping $\phi : \mathbb{R}^2_{xy} \longrightarrow \mathbb{R}^2_{uv}$ given by $\phi(x, y) = \big(u(x, y), v(x, y)\big) = (x + y, x - y)$ on the region we are integrating over . (**a**) Enclosed area in $\mathbb{R}^2_{xy}$ is two and orientation is clockwise. (**b**) Enclosed area in $\mathbb{R}^2_{uv}$ is four and orientation is counter-clockwise

Under the transformation ϕ the corners of the region are mapped as follows,

$$
\begin{aligned}
\mathbb{R}^2_{xy} &\xrightarrow{\phi} \mathbb{R}^2_{uv} \\
(x, y) &\longmapsto (x + y, x - y) \equiv (u, v) \\
(0, 0) &\longmapsto (0, 0) \\
(2, 0) &\longmapsto (2, 2) \\
(0, -2) &\longmapsto (-2, 2).
\end{aligned}
$$

Thus we can see that as we trace around the regions the orientation in $\mathbb{R}^2_{uv}$ is different from the orientation in $\mathbb{R}^2_{xy}$. Also, compare the areas of the two regions in Fig. 7.10. This pictures makes it clear why we need to use the pull-back of the area form. Under the change of coordinates both the area and the orientation of the region we are integrating over changes. By using the pull-back of the area form we compensate for these changes. Differential forms, volume forms in particular, carry with them the information we need in order to do changes of coordinates for integration problems.

In summary, $\int\int_R \; x \; dx \; dy$ should really be thought of as $\int\int_R \; x \; dx \wedge dy$. When we change the coordinates the region $R = \big\{(x, y)\big|0 \le x \le 2, x - 2 \le y \le 0\big\}$ becomes $\tilde{R} = \big\{(u, v)\big|0 \le v \le 2, -v \le u \le v\big\}$, $f(x, y) = x$ becomes $\tilde{f}(u, v) = \frac{u+v}{2}$, and $dx \; dy$, which is really $dx \wedge dy$, becomes $\frac{-1}{2} du \wedge dv$, or $\frac{-1}{2} du \; dv$. In other words, we have the following

$$
\int\int_R \; f(x, y) \;\; dx \; dy = \int\int_{\tilde{R}} \; \tilde{f}(u, v) \;\; \frac{-1}{2} du \; dv
$$

which is really

$$
\int\int_R \; f(x, y) \;\; dx \wedge dy = \int\int_{\tilde{R}} \; \tilde{f}(u, v) \;\; \frac{-1}{2} du \wedge dv.
$$

Let us now integrate the transformed function over the transformed area,

$$
\begin{aligned}
\int_0^2 \int_{-v}^{v} \frac{u + v}{2} \, \frac{-1}{2} \, du \; dv &= \frac{-1}{2} \int_0^2 \left[\frac{u^2}{4} + \frac{uv}{2} \right]_{-v}^{v} dv \\
&= \frac{-1}{2} \int_0^2 v^2 \; dv
\end{aligned}
$$

$$
\begin{aligned}
&= \frac{-1}{2} \left[\frac{v^3}{3} \right]_0^2 \\
&= \frac{-1}{2} \cdot \frac{8}{3} \\
&= -\frac{4}{3}.
\end{aligned}
$$

Often changing variables (also called changing basis) makes the function we are we are trying to integrate much simpler thereby making the integration easier. To see an example suppose we want to integrate $f(x, y) = e^{\frac{x+y}{x-y}}$ over the same region as before, $R = \left\{(x, y) \middle| 0 \le x \le 2, x - 2 \le y \le 0\right\}$. Wanting to simplify this function, and thereby simplify the integration, would be the motivation of choosing the new coordinates $u = x + y$ and $v = x - y$,

$$
\begin{aligned}
\int_{-2}^{0} \int_{y+2}^{0} e^{\frac{x+y}{x-y}} \, dx \, dy &= \int_{-2}^{0} \int_{y+2}^{0} e^{\frac{x+y}{x-y}} \, dx \wedge dy \\
&= \int_0^2 \int_{-v}^{v} e^{\frac{u}{v}} \left(\frac{-1}{2} \right) du \wedge dv \\
&= \int_0^2 \int_{-v}^{v} e^{\frac{u}{v}} \left(\frac{-1}{2} \right) du \, dv \\
&= \frac{-1}{2} \int_0^2 \left[v e^{\frac{u}{v}} \right]_{-v}^{v} dv \\
&= \frac{-1}{2} \int_0^2 v \left(e^{\frac{v}{v}} - e^{\frac{-v}{v}} \right) dv \\
&= \frac{-1}{2} \left(e^1 - e^{-1} \right) \int_0^2 v \, dv \\
&= \frac{-1}{2} \left(e^1 - e^{-1} \right) \left[\frac{v^2}{2} \right]_0^2 \\
&= -\left(e^1 - e^{-1} \right) \\
&\approx -2.35.
\end{aligned}
$$

Clearly, $e^{\frac{x+y}{x-y}} > 0$ for all x and y, so you are integrating a function that is positive over an area, which you generally think of as positive, so you would rather naturally expect to get a positive volume instead of a negative one. Of course, in thinking of it this way we are not taking into account the "orientation" of space. But the fact that one would naturally expect volumes to be positive is why in classical calculus classes the absolute value of the extra term, the Jacobian, which is $\frac{-1}{2}$ here, is taken. In essence this amounts to the convention that we always assume that our coordinate axis follows the "right-hand rule" regardless of what happens during the coordinate transformation.

However, as we start to deal with more advanced and abstract mathematics, we want to be able to keep track of the orientations of our manifold, which is something that differential forms do. Recall, the fact that differential forms keep track of orientations all follows from the fact that our wedgeproduct is defined in terms of the determinant, which itself, by the very way we developed it, keeps track of orientation. The negative sign shows up in our example because when we push-forward two vectors from the xy-plane to the uv-plane using our transformations the image vectors change "handedness" so positively oriented areas become negatively oriented areas, and visa-versa.

7.3 Polar, Cylindrical, and Spherical Coordinates

Now we will look at a variety of examples involving polar, cylindrical, and spherical coordinates. Though you have seen these coordinate systems in calculus the presentation here will relate what you already know from the perspective of differential forms.

7.3.1 Polar Coordinates Example

Now, let's get our hands dirty with an actual example involving polar coordinates. Suppose we want to find the volume of the 3-dimensional solid bounded by the $z = 0$ plane and the paraboloid $z = 3 - x^2 - y^2$, Fig. 7.11. First we find the intersection of paraboloid with the $z = 0$ plane

$$z = 0 \;\Rightarrow\; 0 = 3 - x^2 - y^2 \;\Rightarrow\; x^2 + y^2 = 3,$$

which gives the region $R = \{x^2 + y^2 \leq 3\}$ in the xy-plane. Thus the integral we want to compute is

$$\int_R 3 - x^2 - y^2 \, dx \wedge dy$$

or, in more traditional calculus notation,

$$\int_{-\sqrt{3}}^{\sqrt{3}} \int_{-\sqrt{3-y^2}}^{\sqrt{3-y^2}} 3 - x^2 - y^2 \, dxdy.$$

There is no doubt the computation would be messy.

Question 7.2 Attempt to do this integration with the techniques you remember from calculus class without doing a change of coordinates and see how far you can get.

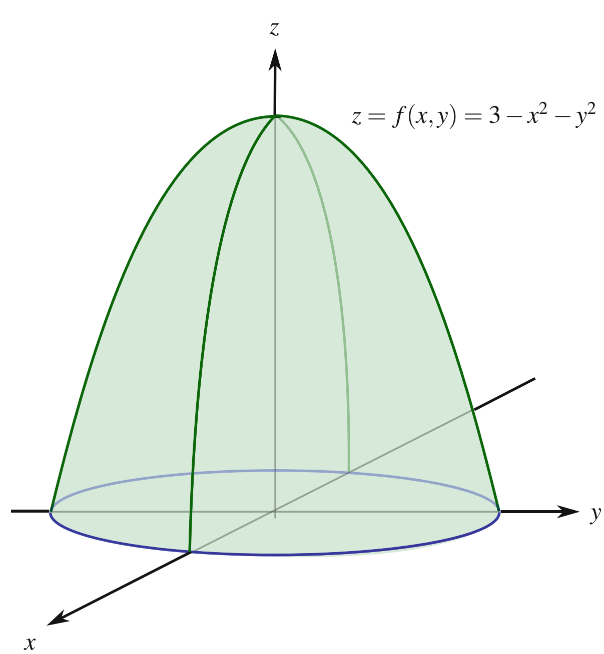


Fig. 7.11 The 3-dimensional solid bounded by the $z = 0$ plane and the paraboloid $z = 3 - x^2 - y^2$

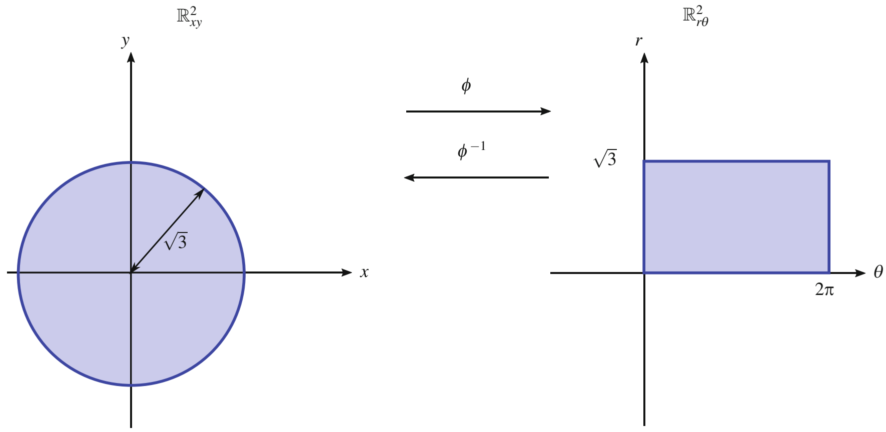


Fig. 7.12 The image of the circular region $\{x^2 + y^2 \le 3\}$ under a polar coordinate change is the rectangular region $\{0 \le r \le \sqrt{3}, 0 \le \theta \le 2\pi\}$

When we actually compute integrals it is clearly much simpler to integrate over rectangular regions then regions of other shapes. Since this region is a disk in the xy-plane using a polar coordinate change results in a rectangular region in the $r\theta$-plane. That is, given the polar coordinate transformation

$$x = r\cos\theta,$$
$$y = r\sin\theta$$

the image of this region under a polar coordinate change is

$$\phi(R) = \{0 \le r \le \sqrt{3}, 0 \le \theta \le 2\pi\}$$

as is shown in Fig. 7.12.

Question 7.3 Show that the region $\phi(R)$ is indeed $\{0 \le r \le \sqrt{3}, 0 \le \theta \le 2\pi\}$.

Thus we want to use the coordinate transformation to simplify our integral by making use of the equality derived from first principles in the last section,

$$\int_R f\, dx_1 \wedge \cdots \wedge dx_n = \int_{\phi(R)} f \circ \phi^{-1}\ \ T^*\phi^{-1} \cdot \big(dx_1 \wedge \cdots \wedge dx_n\big).$$

Here we take a few moments to make some points about both the notation we are using and the comparison between the xy-plane and the $r\theta$-plane. First is that the polar transformation that is usually presented in calculus textbooks, $x = r\cos\theta$ and $y = r\sin\theta$, is the transformation from the $r\theta$-plane to the xy-plane and not visa-versa. The transformation from the xy-plane to the $r\theta$-plane is given by the substantially more complex

$$r = \sqrt{x^2 + y^2}$$
$$\theta = \arctan\left(\frac{y}{x}\right).$$

We would also need to restrict ourselves to $0 < \theta < 2\pi$ in the $r\theta$-plane in order for arctan to be defined. We will play a little fast and loose with this, but for our integration problems doing so will not present a problem. Also, which direction we label ϕ and which we label ϕ^{-1} is of course completely arbitrary, but in order to remain consistent with the notation in the previous section we label the transformation from the xy-plane to the $r\theta$-plane ϕ and the reverse transformation is labeled ϕ^{-1}. It is a little surprising when you think about it, but actually having ϕ^{-1}, that is, x and y in terms of r and θ, allows us to find $\phi(R)$ and $f \circ \phi^{-1}$ easily,

$$\begin{aligned} f \circ \phi^{-1} &= 3 - (r\cos\theta)^2 - (r\sin\theta)^2 \\ &= 3 - r^2(\sin^2\theta + \cos^2\theta) \\ &= 3 - r^2. \end{aligned}$$

Finally, we find $T^*\phi^{-1} \cdot \big(dx \wedge dy\big)$. To do this we first note that $T^*\phi^{-1} \cdot \big(dx \wedge dy\big)$ has the form $g(r,\theta)d\theta \wedge dr$ for some function g. Next we notice that

$$\begin{aligned} g(r,\theta) &= \Big(gd\theta \wedge dr\Big)\left(\begin{bmatrix}1\\0\end{bmatrix}, \begin{bmatrix}0\\1\end{bmatrix}\right) \\ &= \Big(T^*\phi^{-1} \cdot \big(dx \wedge dy\big)\Big)\left(\begin{bmatrix}1\\0\end{bmatrix}, \begin{bmatrix}0\\1\end{bmatrix}\right) \\ &= dx \wedge dy\left(T\phi^{-1} \cdot \begin{bmatrix}1\\0\end{bmatrix}, T\phi^{-1} \cdot \begin{bmatrix}0\\1\end{bmatrix}\right). \end{aligned}$$

In order to proceed further we now need to find $T\phi^{-1}$,

$$T\phi^{-1} = \begin{bmatrix} \frac{\partial x}{\partial\theta} & \frac{\partial x}{\partial r} \\ \frac{\partial y}{\partial\theta} & \frac{\partial y}{\partial r} \end{bmatrix} = \begin{bmatrix} -r\sin\theta & \cos\theta \\ r\cos\theta & \sin\theta \end{bmatrix}$$

so we can then compute

$$\begin{aligned} T\phi^{-1} \cdot \begin{bmatrix}1\\0\end{bmatrix} &= \begin{bmatrix} -r\sin\theta & \cos\theta \\ r\cos\theta & \sin\theta \end{bmatrix}\begin{bmatrix}1\\0\end{bmatrix} = \begin{bmatrix} -r\sin\theta \\ r\cos\theta \end{bmatrix}, \\ T\phi^{-1} \cdot \begin{bmatrix}0\\1\end{bmatrix} &= \begin{bmatrix} -r\sin\theta & \cos\theta \\ r\cos\theta & \sin\theta \end{bmatrix}\begin{bmatrix}0\\1\end{bmatrix} = \begin{bmatrix} \cos\theta \\ \sin\theta \end{bmatrix} \end{aligned}$$

which in turn gives us

$$\begin{aligned} & dx \wedge dy\left(T\phi^{-1} \cdot \begin{bmatrix}1\\0\end{bmatrix}, T\phi^{-1} \cdot \begin{bmatrix}0\\1\end{bmatrix}\right) \\ &= dx \wedge dy\left(\begin{bmatrix} -r\sin\theta \\ r\cos\theta \end{bmatrix}, \begin{bmatrix} \cos\theta \\ \sin\theta \end{bmatrix}\right) \\ &= \begin{vmatrix} -r\sin\theta & \cos\theta \\ r\cos\theta & \sin\theta \end{vmatrix} \\ &= -r\sin^2\theta - r\cos^2\theta \\ &= -r \end{aligned}$$

which means that $g(r,\theta) = -r$. Hence

$$T^*\phi^{-1} \cdot \big(dx \wedge dy\big) = -rd\theta \wedge dr.$$

We had of course already found in the previous chapter.

Now we have all the ingredients we need to rewrite the integral in polar coordinates. Of course, we have gone through a lot of trouble to show all the steps in excruciating detail. Once you have done this a few times the end results are available and you can simply write down the integral in polar coordinates,

$$\begin{aligned}\int_{-\sqrt{3}}^{\sqrt{3}} \int_{-\sqrt{3-y^2}}^{\sqrt{3-y^2}} 3 - x^2 - y^2 \, dxdy &= \int_R f \, dx \wedge dy \\ &= \int_{\phi(R)} f \circ \phi^{-1} \, T^* \phi^{-1} \cdot (dx \wedge dy) \\ &= \int_0^{\sqrt{3}} \int_0^{2\pi} (3 - r^2) \, (-r) d\theta \wedge dr \\ &= \int_0^{\sqrt{3}} \int_0^{2\pi} -3r + r^3 \, d\theta dr.\end{aligned}$$

This is an altogether easier integral to compute than the one we started with. The only caveat is the negative sign in front of the r in the pull-back of the area form $dx \wedge dy$, which reflects the change in orientation.

Question 7.4 Do this integration. Is the answer the same as in Question 7.2, if you managed to complete that integration?

7.3.2 Cylindrical Coordinates Example

Now we are going to do an integration problem where a change of variables to cylindrical coordinates makes the integral easier. Suppose we want to integrate the function $f(x, y) = x^2 + y^2$ in the region bounded by the cone $z = \sqrt{x^2 + y^2}$ and the plane $z = 2$. In other words, we want to find the integral

$$\int_R (x^2 + y^2) \;\; dx \, dy \, dz.$$

In order to draw a picture representing this integral accurately we would need four dimensions, so we will not try to do that. Simply imagine a cartoon along the lines of Fig. 7.1. One way of writing the region R in Cartesian coordinates would be

$$R = \Big\{ -2 \leq x \leq 2, \;\; -\sqrt{4 - x^2} \leq y \leq \sqrt{4 - x^2}, \;\; \sqrt{x^2 + y^2} \leq z \leq 2 \Big\},$$

which results in the integral

$$\int_{-2}^{2} \int_{-\sqrt{4-x^2}}^{\sqrt{4-x^2}} \int_{\sqrt{x^2+y^2}}^{2} (x^2 + y^2) \;\; dz \, dy \, dx.$$

However, like before, we can tell that doing this integral would be quite complicated.

Question 7.5 Show that the region bounded by the cone $z = \sqrt{x^2 + y^2}$ and the plane $z = 2$ can indeed be written as $\{-2 \leq x \leq 2, \;\; -\sqrt{4 - x^2} \leq y \leq \sqrt{4 - x^2}, \;\; \sqrt{x^2 + y^2} \leq z \leq 2\}$.

Question 7.6 Attempt this calculation with the techniques you remember from calculus class without performing a change of variable and see how far you get.

But when we consider the cylindrical coordinate change

$$\begin{aligned} \mathbb{R}^3_{r\theta z} &\xrightarrow{\phi^{-1}} \mathbb{R}^3_{xyz} \\ (r, \theta, z) &\longmapsto (r \cos\theta, \;\; r \sin\theta, \;\; z) \end{aligned}$$

it is fairly obvious that the region has a much simpler description in cylindrical coordinates,

$$\phi(R) = \Big\{ 0 \leq \theta \leq 2\pi, \;\; 0 \leq r \leq 2, \;\; r \leq z \leq 2 \Big\}.$$

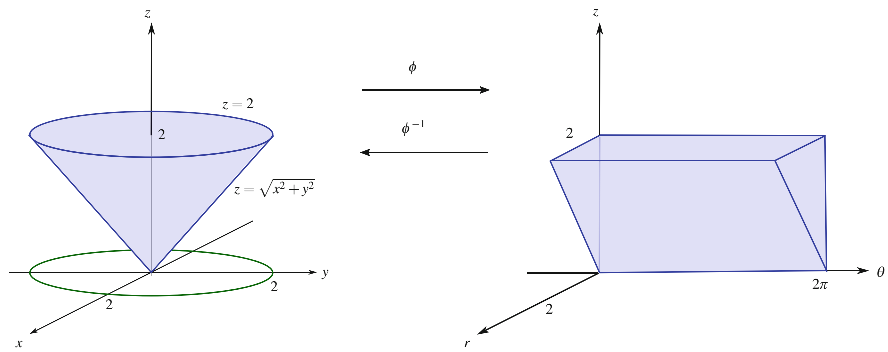


Fig. 7.13 The affect of the mapping $\phi : \mathbb{R}^3_{xyz} \longrightarrow \mathbb{R}^2_{r\theta z}$ on region we are integrating over

Question 7.7 Show that the region $\phi(R)$ is indeed $\{0 \leq \theta \leq 2\pi, \quad 0 \leq r \leq 2, \quad r \leq z \leq 2\}$.

We point out that to keep our notation in this section consistent with the last section, we are denoting the mapping from $\mathbb{R}^3_{r\theta z}$ to $\mathbb{R}^3_{xyz}$ as ϕ^{-1} instead of ϕ as we would have in the last chapter. As long as we restrict our mappings to the appropriate domain of $\mathbb{R}^3_{r\theta z}$ so our inverse is well-defined then this should hopefully not cause any confusion. See Fig. 7.13. We also have that $f \circ \phi^{-1} = r^2$ and $T^*\phi^{-1} \cdot (dx \wedge dy \wedge dz) = -r d\theta \wedge dr \wedge dz$.

Question 7.8 Find both $f \circ \phi^{-1}$ and $T^*\phi^{-1} \cdot (dx \wedge dy \wedge dz)$.

Finally, using the identity developed in the last section we get

$$\begin{aligned}
\int_R f \, dx \wedge dy \wedge dz &= \int_{\phi(R)} f \circ \phi^{-1} \;\; T^*\phi^{-1} \cdot \big(dx \wedge dy \wedge dz\big) \\
&= \int_{\phi(R)} (r^2) \;\; (-r) d\theta \wedge dr \wedge dz \\
&= \int_{\phi(R)} (r^2) \;\; r d\theta \wedge dz \wedge dr \\
&= \int_0^{2\pi} \int_0^2 \int_r^2 r^3 \;\; d\theta \, dz \, dr.
\end{aligned}$$

Question 7.9 Complete this integration.

7.3.3 *Spherical Coordinates Example*

We finish off this section with an example of a coordinate change using spherical coordinates. Suppose we wanted to integrate the function $f(x, y, z) = e^{(x^2+y^2+z^2)^{3/2}}$ over the unit ball

$$R = \Big\{x^2 + y^2 + z^2 \leq 1\Big\}.$$

One integral that would do this is

$$\int_{-1}^{1} \int_{-\sqrt{1-x^2}}^{\sqrt{1-x^2}} \int_{-\sqrt{1-x^2-y^2}}^{\sqrt{1-x^2-y^2}} e^{(x^2+y^2+z^2)^{3/2}} \;\; dz \, dy \, dx,$$

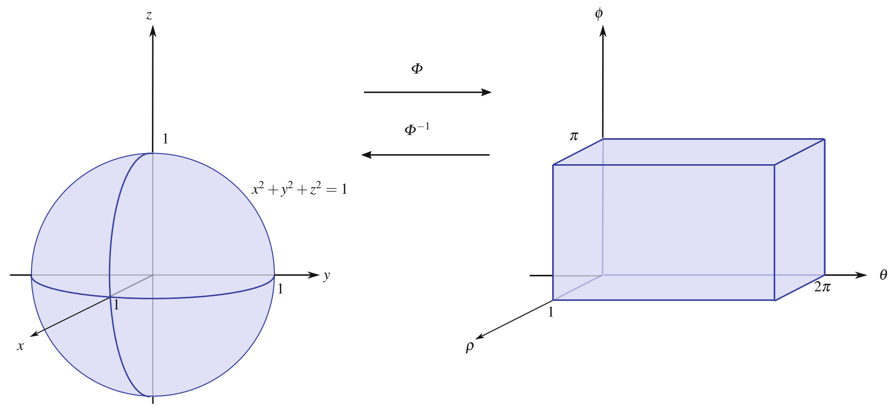


Fig. 7.14 The effect of the mapping $\Phi : \mathbb{R}^3_{xyz} \longrightarrow \mathbb{R}^2_{\rho\theta\phi}$ on region we are integrating over

which would clearly be difficult to do at best. Under the spherical coordinate transformation

$$\mathbb{R}^3_{\rho\theta\phi} \xrightarrow{\Phi^{-1}} \mathbb{R}^3_{xyz}$$
$$(\rho, \theta, \phi) \longmapsto (\rho \sin\phi \cos\theta, \quad \rho \sin\phi \sin\theta, \quad \rho\cos\phi)$$

our region becomes

$$\Phi(R) = \Big\{ 0 \le \rho \le 1, \quad 0 \le \theta \le 2\pi, \quad 0 \le \phi \le \pi \Big\}.$$

Since ϕ is one of the standard variables used in spherical coordinates, we have denoted our mapping by the upper-case Φ. See Fig. 7.14.

Question 7.10 Show that the region $\Phi(R)$ is indeed $\{0 \le \rho \le 1, \quad 0 \le \theta \le 2\pi, \quad 0 \le \phi \le \pi\}$.

We also get $f \circ \Phi^{-1} = e^{(\rho^2)^{3/2}} = e^{\rho^3}$ and $T^*\Phi^{-1} \cdot (dx \wedge dy \wedge dz) = -\rho^2 \sin\phi \; d\rho \wedge d\theta \wedge d\phi$. Again, using the identity developed in the last section we have

$$\begin{aligned}
\int_R f \; dx \wedge dy \wedge dz &= \int_{\Phi(R)} f \circ \Phi^{-1} \;\; T^*\Phi^{-1} \cdot \big(dx \wedge dy \wedge dz\big) \\
&= \int_{\phi(R)} (e^{\rho^3}) \;\; (-\rho^2 \sin\phi) \; d\rho \wedge d\theta \wedge d\phi \\
&= \int_0^{\pi} \int_0^{2\pi} \int_0^1 \; -\rho^2 e^{\rho^3} \sin\phi \;\; d\rho \; d\theta \; d\phi.
\end{aligned}$$

Question 7.11 Complete this integration.

7.4 Integration of Differential Forms on Parameterized Surfaces

So far we have looked at a specific kind of integration problem, problems where the region we want to integrate over has a much simpler representation in some other coordinate system. There are, however, other kinds of integration problems we can do using differential forms.

As long as we have an n-form on an n-dimensional manifold we can integrate it. For example, we can integrate one-forms on curves, two-forms on surfaces, three-forms on three-dimensional spaces, et cetera. As long as we have a parametrization of the manifold $\Sigma_k \subset \mathbb{R}^n$, $k < n$, that is, an invertible mapping $\phi^{-1} : U \subset \mathbb{R}^k \longrightarrow \Sigma_k \subset \mathbb{R}^n$, we can integrate an k-form on Σ_k by pulling the k-form back to $U \subset \mathbb{R}^k$ using $T^*\phi^{-1}$ and integrating there. It is a little unusual to name a parametrisation with ϕ^{-1} instead of ϕ but we want to keep notation consistent with the last chapter. The rest of this section will be about integrating forms over submanifolds of $\mathbb{R}^n$ that are given as parameterized surfaces Σ_k. The equality we had developed earlier was

$$\int_R f\, dx_1 \wedge \cdots \wedge dx_n = \int_{\phi(R)} f \circ \phi^{-1}\; T^*\phi^{-1} \cdot \big(dx_1 \wedge \cdots \wedge dx_n\big).$$

The version of this equality we generally want to use is

$$\boxed{\begin{array}{c}\text{Integral of}\\ k-\text{Form on}\\ k-\text{Dimensional}\\ \text{Parameterized}\\ \text{Surface } \Sigma_k\end{array} \qquad \int_{\Sigma_k} \alpha = \int_{\phi(\Sigma_k)} T^*\phi^{-1} \cdot \alpha}$$

When $k = 1$ then we have a curve and we we often write $\Sigma_1 = C$.

Question 7.12 Define this formula using an argument involving Riemann sums similar to the argument in Sect. 7.1.

But before doing that we also want to reiterate, you can, and often do, integrate all sorts of things that are not differential forms. For example, to find the surface area of a 2-dimensional surface in $\mathbb{R}^3$ you would need to integrate

$$\sqrt{(dx \wedge dy)^2 + (dy \wedge dz)^2 + (dz \wedge dx)^2}$$

which is not a differential form since it is not multi-linear, even though it is composed of the differential forms $dx \wedge dy$, $dy \wedge dz$, and $dz \wedge dx$. However, for simple surfaces sometimes integrals like this can still be done by hand. Otherwise a numerical answer can generally be found using numerical methods. One of the things about forms that makes them so useful is the generalized Stokes' theorem. However, the generalized Stokes' theorem does not apply to integrands that are not forms.

7.4.1 Line Integrals

Here we will look at some examples that involve integrating one-forms on a one-dimensional manifold. We will make use of the identity

$$\int_C \alpha = \int_{\phi(C)} T^*\phi^{-1} \cdot \alpha.$$

Example One

We begin with an example in $\mathbb{R}^2$. We will integrate the one-form $y^2 dx + x dy$ along two different curves. The first curve C_1 will be the line segment from $(-5, -3)$ to $(0, 2)$. The second curve C_2 will be the arc of the parabola $x = 4 - y^2$ from $(-5, -3)$ to $(0, 2)$. We begin with the integral along C_1. In order to do this we first of all parameterize the path C_1 by $\phi^{-1}(t) = \big(x(t), y(t)\big) = (5t - 5, 5t - 3)$. Letting $R = [0, 1] \subset \mathbb{R}$ we have $\phi^{-1}(R) = \big\{(x, y) \mid x = 5t - 5, y = 5t - 3, \quad 0 \le t \le 1\big\} = C_1$. See Fig. 7.15. Using the mapping ϕ^{-1} we find $T\phi^{-1}$,

$$T\phi^{-1} = \begin{bmatrix} \frac{\partial x}{\partial t} \\ \frac{\partial y}{\partial t} \end{bmatrix} = \begin{bmatrix} 5 \\ 5 \end{bmatrix}$$

so we have the induced tangent and cotangent mappings,

$$\begin{array}{ccc} T_p(\mathbb{R}) & \xrightarrow{T\phi^{-1}} & T_{\phi^{-1}(p)}\big(\mathbb{R}^2\big) \\ [1] \in T_p(\mathbb{R}) & \longmapsto & T\phi^{-1} \cdot [1] = \begin{bmatrix} 5 \\ 5 \end{bmatrix} [1] = \begin{bmatrix} 5 \\ 5 \end{bmatrix} \in T_{\phi^{-1}(p)}\big(\mathbb{R}^2\big) \end{array}$$

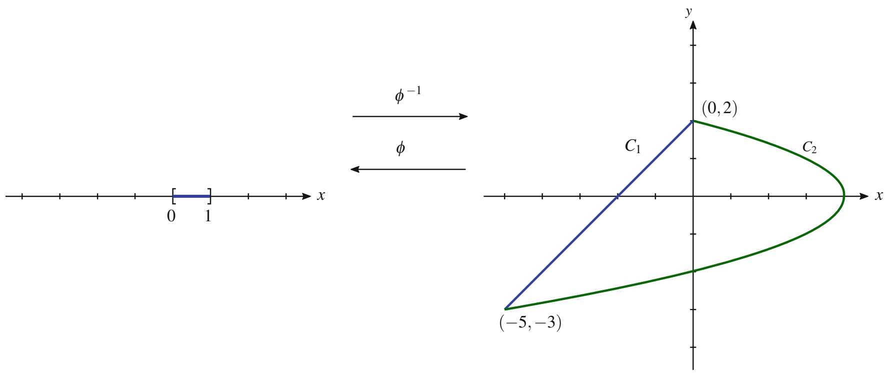


Fig. 7.15 The mapping ϕ^{-1} sends [0, 1] to curve C_1

$$\begin{array}{ccc} T_p^*(\mathbb{R}) & \xleftarrow{T^*\phi^{-1}} & T_{\phi^{-1}(p)}\left(\mathbb{R}^2\right) \\ f(t)dt = T^*\phi^{-1}\cdot(y^2dx + xdy) & \longleftarrow & y^2dx + xdy \end{array}$$

for some $f(t)$. This function of t is what we want to find. Notice that $f(t)dt([1]) = f(t)$ so we have

$$\begin{aligned} f(t) &= f(t)dt([1]) \\ &= \left(T^*\phi^{-1}\cdot\left(y^2dx + xdy\right)\right)([1]) \\ &= \left(y^2dx + xdy\right)\left(T\phi^{-1}\cdot[1]\right) \\ &= \left(y^2dx + xdy\right)\left(\begin{bmatrix}5\\5\end{bmatrix}\right) \\ &= y^2dx\left(\begin{bmatrix}5\\5\end{bmatrix}\right) + xdy\left(\begin{bmatrix}5\\5\end{bmatrix}\right) \\ &= y^2(5) + x(5) \\ &= (5t-3)^2(5) + (5t-5)(5) \\ &= 125t^2 - 125t + 20 \end{aligned}$$

so we have

$$T^*\phi^{-1}\cdot\left(y^2dx + xdy\right) = (125t^2 - 125t + 20)dt.$$

Using the following equality

$$\int_C \alpha = \int_{\phi(C)} T^*\phi^{-1}\cdot\alpha$$

our integral becomes

$$\begin{aligned} \int_{C_1}\left(y^2dx + xdy\right) &= \int_{[0,1]}\left(125t^2 - 125t + 20\right)\,dt \\ &= \int_0^1\left(125t^2 - 125t + 20\right)\,dt \\ &= \frac{-5}{6}. \end{aligned}$$

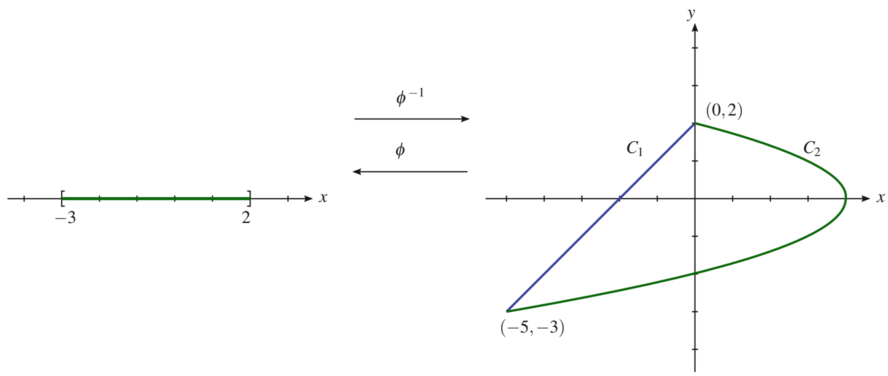


Fig. 7.16 The mapping ϕ^{-1} sends $[-3, 2]$ to curve C_2

Now we will find the integral of the one-form $y^2dx + xdy$ along the second path C_2.In order to do this we first of all parameterize the path C_2 by $\phi^{-1}(t) = \big(x(t), y(t)\big) = (4-t^2, t)$. Letting $R = [-3, 2] \subset \mathbb{R}$ we have $\phi^{-1}(R) = \{\ (x, y) \mid x = 4 - t^2,\ \ y = t,\ \ -3 \le t \le 2\} = C_2$. See Fig. 7.16. Using the mapping ϕ^{-1} we find $T\phi^{-1}$,

$$T_p\phi^{-1} = \begin{bmatrix} \frac{\partial x}{\partial t} \\ \frac{\partial y}{\partial t} \end{bmatrix}_p = \begin{bmatrix} -2t \\ 1 \end{bmatrix}_p$$

which is actually the mapping

$$\begin{aligned} T_p(\mathbb{R}) &\xrightarrow{T\phi^{-1}} T_{\phi^{-1}(p)}\left(\mathbb{R}^2\right) \\ [1] \in T_p(\mathbb{R}) &\longmapsto T_p\phi^{-1} \cdot [1] = \begin{bmatrix} -2t \\ 1 \end{bmatrix}[1] = \begin{bmatrix} -2t \\ 1 \end{bmatrix} \in T_{\phi^{-1}(p)}\left(\mathbb{R}^2\right). \end{aligned}$$

From this we can clearly see the mapping $T_p\phi^{-1}$ depends on the point our vector is at. To make this more concrete we do a few examples showing what the vector [1] based at differing points gets pushed to.

$$\begin{aligned} [1] \in T_{-3}(\mathbb{R}) &\implies T_{-3}\phi^{-1} \cdot [1] = \begin{bmatrix} -2(-3) \\ 1 \end{bmatrix}_{-3} \cdot [1] = \begin{bmatrix} 6 \\ 1 \end{bmatrix} \in T_{(-5,-3)}\left(\mathbb{R}^2\right), \\ [1] \in T_{-2}(\mathbb{R}) &\implies T_{-2}\phi^{-1} \cdot [1] = \begin{bmatrix} -2(-2) \\ 1 \end{bmatrix}_{-2} \cdot [1] = \begin{bmatrix} 4 \\ 1 \end{bmatrix} \in T_{(0,-2)}\left(\mathbb{R}^2\right), \\ [1] \in T_{-1}(\mathbb{R}) &\implies T_{-1}\phi^{-1} \cdot [1] = \begin{bmatrix} -2(-1) \\ 1 \end{bmatrix}_{-1} \cdot [1] = \begin{bmatrix} 2 \\ 1 \end{bmatrix} \in T_{(3,-1)}\left(\mathbb{R}^2\right), \end{aligned}$$

et cetera. We also have the cotangent mapping

$$\begin{aligned} T_p^*(\mathbb{R}) &\xleftarrow{T^*\phi^{-1}} T_{\phi^{-1}(p)}\left(\mathbb{R}^2\right) \\ f(t)dt = T^*\phi^{-1} \cdot (y^2dx + xdy) &\longleftarrow y^2dx + xdy \end{aligned}$$

for some $f(t)$. This $f(t)$ is what we want to find. Similar to before we have

$$\begin{aligned} f(t) &= f(t)dt([1]) \\ &= \left(T^*\phi^{-1} \cdot \left(y^2dx + xdy\right)\right)([1]) \end{aligned}$$

$$
\begin{aligned}
&= \left(y^2dx + xdy\right)\left(T\phi^{-1} \cdot [1]\right) \\
&= \left(y^2dx + xdy\right)\left(\begin{bmatrix} -2t \\ 1 \end{bmatrix}\right) \\
&= y^2dx\left(\begin{bmatrix} -2t \\ 1 \end{bmatrix}\right) + xdy\left(\begin{bmatrix} -2t \\ 1 \end{bmatrix}\right) \\
&= y^2(-2t) + x(1) \\
&= (t)^2(-2t) + (4 - t^2)(1) \\
&= -2t^3 - t^2 + 4
\end{aligned}
$$

which means that

$$T^*\phi^{-1} \cdot \left(y^2dx + xdy\right) = \left(-2t^3 - t^2 + 4\right) dt.$$

Again, using the following equality

$$\int_C \alpha = \int_{\phi(C)} T^*\phi^{-1} \cdot \alpha$$

our integral becomes

$$
\begin{aligned}
\int_{C_2} \left(y^2dx + xdy\right) &= \int_{[-3,2]} \left(-2t^3 - t^2 + 4\right)\, dt \\
&= \int_{-3}^{2} \left(-2t^3 - t^2 + 4\right)\, dt \\
&= 40\frac{5}{6}.
\end{aligned}
$$

Notice that our answers are not the same. Even though the start point and the end point for paths C_1 and C_2 are the same the integral of the one-form $y^2dx + xdy$ depends on the path taken.

Example Two

In this example we will do something similar to what we did in example one. Recall that in example one we integrated a one-form along two different curves C_1 and C_2 that had the same starting point and ending point and discovered that the two integrals were different, that is, the integral depended on the path taken. We will repeat this using a different one-form. The one-form we integrate this time will be a special kind of one-form, it will be the exterior derivative of a zero-form.

We will choose the zero-form (function) on $\mathbb{R}^2$, $f(x, y) = x^2y$. Taking the exterior derivative of this zero-form gives us the one-form

$$df = \frac{\partial f}{\partial x}\, dx + \frac{\partial f}{\partial y}\, dy = 2xy\, dx + x^2\, dy.$$

This is the one-form that we will integrate. We will use the same paths and parameterizations as in example one. The procedure is essentially the same as the previous examples. We know that $T^*\phi^{-1} \cdot \left(2xydx + x^2dy\right)$ is a one-form on $\mathbb{R}$ and so it must take the form $f(t)dt$ for some function $f(t)$. For curve C_1 we find

$$
\begin{aligned}
f(t) &= f(t)dt([1]) \\
&= \left(T^*\phi^{-1} \cdot \left(2xydx + x^2dy\right)\right)([1]) \\
&= \left(2xydx + x^2dy\right)\left(\begin{bmatrix} 5 \\ 5 \end{bmatrix}\right)
\end{aligned}
$$

$$
\begin{aligned}
&= 10xy + 5x^2 \\
&= 10(5t-5)(5t-3) + 5(5t-5)^2 \\
&= 5(75t^2 - 130t + 55)
\end{aligned}
$$

giving us $T^*\phi^{-1} \cdot \big(2xydx + x^2dy\big) = 5(75t^2 - 130t + 55)\ dt$. We use the identity $\int_{C_1} \alpha = \int_{\phi(C_1)} T^*\phi^{-1} \cdot \alpha$ to get

$$
\begin{aligned}
\int_{C_1} \Big(2xydx + x^2dy\Big) &= \int_{\phi(C_1)} T^*\phi^{-1} \cdot \Big(2xydx + x^2dy\Big) \\
&= \int_{[0,1]} 5(75t^2 - 130t + 55)\ dt \\
&= 75.
\end{aligned}
$$

Now we take the integral along path C_2. Similarly we find

$$
\begin{aligned}
f(t) &= f(t)dt([1]) \\
&= \Big(T^*\phi^{-1} \cdot \Big(2xydx + x^2dy\Big)\Big)([1]) \\
&= \Big(2xydx + x^2dy\Big)\left(\begin{bmatrix} -2t \\ 1 \end{bmatrix}\right) \\
&= 2xy(-2t) + x^2(1) \\
&= 2(4-t^2)(t)(-2t) + (4-t^2)^2 \\
&= 5t^4 - 24t^2 + 16
\end{aligned}
$$

giving us $T^*\phi^{-1} \cdot \big(2xydx + x^2dy\big) = (5t^4 - 24t^2 + 16)\ dt$. The integral is

$$
\begin{aligned}
\int_{C_2} \Big(2xydx + x^2dy\Big) &= \int_{\phi(C_2)} T^*\phi^{-1} \cdot \Big(2xydx + x^2dy\Big) \\
&= \int_{[-3,2]} (5t^4 - 24t^2 + 16)\ dt \\
&= 75.
\end{aligned}
$$

Something interesting has just happened. When our one-form was the exterior derivative of a zero-form then the result of our integration does not seem to depend on the path taken. In fact this is true in general, not just in this example. We will look at this in more detail later on.

Example Three

Our next example involves integrating the one-form $ydx + zdy + xdz$ along the straight path C in $\mathbb{R}^3$ from the point $(2, 0, 0)$ to the point $(3, 4, 5)$. We find a parameterizations $\phi^{-1} : \mathbb{R} \to \mathbb{R}^3$ of this path given by $\phi^{-1}(t) = \big(x(t), y(t), z(t)\big) = (2+t, 4t, 5t)$. Letting $R = [0, 1] \subset \mathbb{R}$ we have $\phi^{-1}(R) = \Big\{ (x, y, z) \Bigm| x = 2+t, y = 4t, z = 5t, \ \ 0 \le t \le 1 \Big\} = C$. This gives

$$
T_p\phi^{-1} = \begin{bmatrix} \frac{\partial x}{\partial t} \\ \frac{\partial y}{\partial t} \\ \frac{\partial z}{\partial t} \end{bmatrix}_p = \begin{bmatrix} 1 \\ 4 \\ 5 \end{bmatrix}.
$$

Clearly, $T_p\phi^{-1}$ does not depend on the base point $p \in R$. As before we know $T^*\phi^{-1} \cdot (ydx + zdy + xdz) = f(t)dt$ for some $f(t)$, which we wish to find. The computation proceeds just like before

$$\begin{aligned}
f(t) &= f(t)dt([1]) \\
&= \Big(T^*\phi^{-1} \cdot (ydx + zdy + xdz)\Big)([1]) \\
&= (ydx + zdy + xdz)\Big(T\phi^{-1} \cdot [1]\Big) \\
&= (ydx + zdy + xdz)\left(\begin{bmatrix} 1 \\ 4 \\ 5 \end{bmatrix} \cdot [1]\right) \\
&= ydx\left(\begin{bmatrix} 1 \\ 4 \\ 5 \end{bmatrix}\right) + zdy\left(\begin{bmatrix} 1 \\ 4 \\ 5 \end{bmatrix}\right) + xdz\left(\begin{bmatrix} 1 \\ 4 \\ 5 \end{bmatrix}\right) \\
&= y(1) + z(4) + x(5) \\
&= (4t) + 4(5t) + 5(2 + t) \\
&= 29t + 10
\end{aligned}$$

giving us $T^*\phi^{-1} \cdot (ydx + zdy + xdz) = (29t + 10)dt$. Using $\int_C \alpha = \int_{\phi(C)} T^*\phi^{-1} \cdot \alpha$ we have the integral

$$\begin{aligned}
\int_C (ydx + zdy + xdz) &= \int_{\phi(C)} T^*\phi^{-1} \cdot (ydx + zdy + xdz) \\
&= \int_{[0,1]} (29t + 10)\, dt \\
&= 24\frac{1}{2}.
\end{aligned}$$

7.4.2 Surface Integrals

Now we will do some examples where we integrate a two-form over a parameterized two-dimensional surface.

Example One
We will integrate the two-form $z^2 dx \wedge dy$ over the top half of the unit sphere. We will use the parametrization given by

$$\begin{aligned}
\phi^{-1}(r, \theta) &= \big(x(r, \theta), y(r, \theta), z(r, \theta)\big) \\
&= \big(r\cos\theta, r\sin\theta, \sqrt{1 - r^2}\big).
\end{aligned}$$

Letting $R = \left\{ (r, \theta) \;\middle|\; 0 \le r \le 1, 0 \le \theta \le 2\pi \right\}$ we have that $\phi^{-1}(R)$ is the top half of the unit sphere. See Fig. 7.17. First we find $T\phi^{-1}$,

$$T\phi^{-1} = \begin{bmatrix} \frac{\partial x}{\partial r} & \frac{\partial x}{\partial \theta} \\ \frac{\partial y}{\partial r} & \frac{\partial y}{\partial \theta} \\ \frac{\partial z}{\partial r} & \frac{\partial z}{\partial \theta} \end{bmatrix}$$

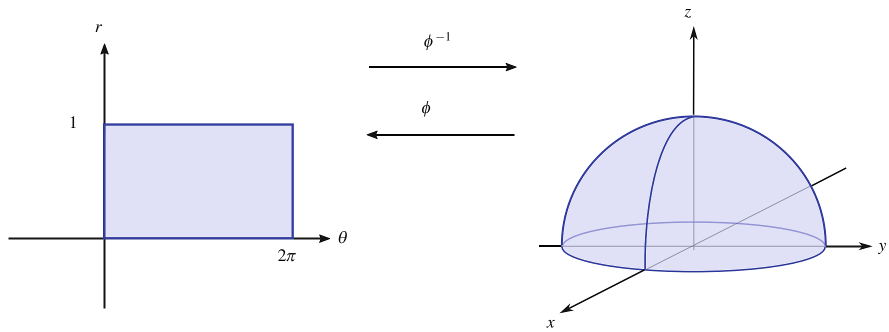


Fig. 7.17 The mapping ϕ^{-1} sends $R = \{ (r,\theta) \mid 0 \le r \le 1, 0 \le \theta \le 2\pi \}$ to the top of the unit sphere in $\mathbb{R}^3$

$$= \begin{bmatrix} \frac{\partial(r\cos\theta)}{\partial r} & \frac{\partial(r\cos\theta)}{\partial\theta} \\ \frac{\partial(r\sin\theta)}{\partial r} & \frac{\partial(r\sin\theta)}{\partial\theta} \\ \frac{\partial(\sqrt{1-r^2})}{\partial r} & \frac{\partial(\sqrt{1-r^2})}{\partial\theta} \end{bmatrix}$$

$$= \begin{bmatrix} \cos\theta & -r\sin\theta \\ \sin\theta & r\cos\theta \\ \frac{-r}{\sqrt{1-r^2}} & 0 \end{bmatrix}.$$

Now we want to find $T^*\phi^{-1} \cdot \left(z^2 dx \wedge dy\right)$. We know this is a two-form on $\mathbb{R}^2$ so it must have the form $g(r,\theta)dr \wedge d\theta$ for some function $g(r,\theta)$. Our goal is to find that function. Notice that

$$g(r,\theta)dr \wedge d\theta \left(\begin{bmatrix} 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \end{bmatrix} \right) = g(r,\theta) \begin{vmatrix} 1 & 0 \\ 0 & 1 \end{vmatrix} = g(r,\theta).$$

Therefore,

$$\begin{aligned} g(r,\theta) &= g(r,\theta)dr \wedge d\theta \left(\begin{bmatrix} 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \end{bmatrix} \right) \\ &= T^*\phi^{-1} \cdot \left(z^2 dx \wedge dy\right) \left(\begin{bmatrix} 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \end{bmatrix} \right) \\ &= z^2 dx \wedge dy \left(T\phi^{-1} \cdot \begin{bmatrix} 1 \\ 0 \end{bmatrix}, T\phi^{-1} \cdot \begin{bmatrix} 0 \\ 1 \end{bmatrix} \right) \\ &= z^2 dx \wedge dy \left(\begin{bmatrix} \cos\theta \\ \sin\theta \\ \frac{-r}{\sqrt{1-r^2}} \end{bmatrix}, \begin{bmatrix} -r\sin\theta \\ r\cos\theta \\ 0 \end{bmatrix} \right) \\ &= z^2 \begin{vmatrix} \cos\theta & -r\sin\theta \\ \sin\theta & r\cos\theta \end{vmatrix} \\ &= z^2\, r \\ &= \left(\sqrt{1-r^2}\right)^2 r \\ &= r - r^3 \end{aligned}$$

$$\implies \quad T^*\phi^{-1} \cdot \left(z^2 dx \wedge dy\right) = (r - r^3)\, dr \wedge d\theta.$$

Using the identity $\int_R \alpha = \int_{\phi(R)} T^*\phi^{-1} \cdot \alpha$ we have

$$\begin{aligned}
\int_R z^2\, dx \wedge dy &= \int_{\phi(R)} (r - r^3)\, dr \wedge d\theta \\
&= \int_0^{2\pi} \int_0^1 (r - r^3)\, dr d\theta \\
&= \int_0^{2\pi} \left[\frac{r^2}{2} - \frac{r^4}{4}\right]_0^1 d\theta \\
&= \int_0^{2\pi} \frac{1}{4}\, d\theta \\
&= \frac{\pi}{2}.
\end{aligned}$$

Example Two

Now we will integrate the two-form $\alpha = \frac{1}{x} dy \wedge dz - \frac{1}{y} dx \wedge dz$ on the top half of the unit sphere using the following three parameterizations. We will do the first parametrization and leave the other two as an exercise.

(a) $(r, \theta) \mapsto \left(r\cos\theta, r\sin\theta, \sqrt{1-r^2}\right), \quad 0 \le \theta \le 2\pi,\ 0 \le r \le 1$
(b) $(\theta, \phi) \mapsto (\sin\phi\cos\theta, \sin\phi\sin\theta, \cos\theta), \quad 0 \le \theta \le 2\pi,\ 0 \le \phi \le \frac{\pi}{2}$
(c) $(x, y) \mapsto \left(x, y, \sqrt{1-x^2-y^2}\right), \quad \sqrt{x^2+y^2} \le 1$

We should expect that we will get the same answer regardless of the parameterizations. We now proceed with the first parametrization which follows example one, except that we have

$$g(r,\theta) dr \wedge d\theta = T^*\phi^{-1} \cdot \left(\frac{1}{x} dy \wedge dz - \frac{1}{y} dx \wedge dz\right) \xleftarrow{T^*\phi^{-1}} \frac{1}{x} dy \wedge dz - \frac{1}{y} dx \wedge dz.$$

Wanting to find $g(r,\theta)$ we let e_1 and e_2 be the unit vectors in the r and θ directions respectively, so we have

$$\begin{aligned}
g(r,\theta) &= g(r,\theta)\, dr \wedge d\theta\, (e_1, e_2) \\
&= T^*\phi^{-1} \cdot \left(\frac{1}{x} dy \wedge dz - \frac{1}{y} dx \wedge dz\right)(e_1, e_2) \\
&= \left(\frac{1}{x} dy \wedge dz - \frac{1}{y} dx \wedge dz\right)(T\phi \cdot e_1, T\phi \cdot e_2) \\
&= \left(\frac{1}{x} dy \wedge dz - \frac{1}{y} dx \wedge dz\right)\left(\begin{bmatrix} \cos\theta \\ \sin\theta \\ \frac{-r}{\sqrt{1-r^2}} \end{bmatrix}, \begin{bmatrix} -r\sin\theta \\ r\cos\theta \\ 0 \end{bmatrix}\right) \\
&= \frac{1}{x} dy \wedge dz \left(\begin{bmatrix} \cos\theta \\ \sin\theta \\ \frac{-r}{\sqrt{1-r^2}} \end{bmatrix}, \begin{bmatrix} -r\sin\theta \\ r\cos\theta \\ 0 \end{bmatrix}\right) - \frac{1}{y} dx \wedge dz \left(\begin{bmatrix} \cos\theta \\ \sin\theta \\ \frac{-r}{\sqrt{1-r^2}} \end{bmatrix}, \begin{bmatrix} -r\sin\theta \\ r\cos\theta \\ 0 \end{bmatrix}\right) \\
&= \frac{1}{x} \begin{vmatrix} \sin\theta & r\cos\theta \\ \frac{-r}{\sqrt{1-r^2}} & 0 \end{vmatrix} - \frac{1}{y} \begin{vmatrix} \cos\theta & -r\sin\theta \\ \frac{-r}{\sqrt{1-r^2}} & 0 \end{vmatrix} \\
&= \frac{1}{x}\left(\frac{r^2\cos\theta}{\sqrt{1-r^2}}\right) - \frac{1}{y}\left(\frac{-r^2\sin\theta}{\sqrt{1-r^2}}\right) \\
&= \frac{1}{r\cos\theta}\left(\frac{r^2\cos\theta}{\sqrt{1-r^2}}\right) - \frac{1}{r\sin\theta}\left(\frac{-r^2\sin\theta}{\sqrt{1-r^2}}\right)
\end{aligned}$$

$$= \frac{2r}{\sqrt{1-r^2}}$$

$$\Longrightarrow \quad T^*\phi^{-1} \cdot \left(\frac{1}{x} dy \wedge dz - \frac{1}{y} dx \wedge dz \right) = \frac{2r}{\sqrt{1-r^2}} dr \wedge d\theta.$$

Now we use the identity $\int_R \alpha = \int_{\phi(R)} T^*\phi^{-1} \cdot \alpha$ to take the integral,

$$\begin{aligned}
&\int_R \left(\frac{1}{x} dy \wedge dz - \frac{1}{y} dx \wedge dz \right) \\
&= \int_{\phi(R)} \frac{2r}{\sqrt{1-r^2}} \; dr \wedge d\theta \\
&= \int_0^{2\pi} \int_0^1 \frac{2r}{\sqrt{1-r^2}} \; dr d\theta \\
&= 2 \left(\int_0^{2\pi} d\theta \right) \cdot \left(\int_0^1 \frac{2r}{\sqrt{1-r^2}} \; dr \right) \\
&\qquad \text{substitution } u = 1 - r^2, \; dr = \frac{du}{-2r} \\
&= 2 \cdot 2\pi \cdot \left(\frac{-1}{2} \int \frac{1}{\sqrt{u}} \; du \right) \\
&= 4\pi \cdot \left(-\sqrt{u} \right) \\
&= 4\pi \cdot \left[-\sqrt{1-r^2} \right]_0^1 \\
&= 4\pi.
\end{aligned}$$

Question 7.13 Complete this example by finding the integral using the second and third parameterizations.

7.5 Summary, References, and Problems

7.5.1 Summary

If finding the integral $\int_R f(x_1, x_2 \ldots, x_n) dx_1 dx_2 \cdots dx_n$ is difficult but it would become easier by doing the change of coordinates

$$\begin{aligned}
\mathbb{R}^n_{x_1 x_2 \cdots x_n} &\xrightarrow{\phi} \mathbb{R}^n_{\phi_1 \phi_2 \cdots \phi_n} \\
(x_1, x_2, \ldots, x_n) &\longmapsto (\phi_1, \phi_2, \ldots, \phi_n)
\end{aligned}$$

then the change of coordinates formula is used

Change of coordinates formula	$\int_R f(x_1, \ldots, x_n) \; dx_1 \wedge \ldots \wedge dx_n = \int_{\phi(R)} f \circ \phi^{-1}(\phi_1, \ldots, \phi_n) \; T^*\phi^{-1} \cdot (dx_1 \wedge \ldots \wedge dx_n)$

The left hand side takes place in $x_1 \cdots x_n$-coordinates and we are integrating the function $f(x_1, \ldots, x_n)$ over the region R using the volume form $dx_1 \wedge \cdots \wedge dx_n$ associated with the $x_1 \cdots x_n$-coordinates. On the right hand side the function $f \circ \phi^{-1} : \mathbb{R}^n \to \mathbb{R}$ is a function of the variables $\phi_1, \phi_2, \cdots \phi_n$. The region we are integrating over in $\mathbb{R}^n_{\phi_1 \phi_2 \cdots \phi_n}$ is the image

$\phi(R)$. Finally, the form on the right hand side is $T^*\phi^{-1} \cdot \big(dx_1 \wedge \cdots \wedge dx_n\big)$, the pull-back of the area form $dx_1 \wedge \cdots \wedge dx_n$ by ϕ^{-1} and NOT the area form $d\phi_1 \wedge \cdots \wedge d\phi_n$. This gives us exactly what you used in multivariable calculus when you used polar, cylindrical, and spherical changes of coordinates.

We can integrate one-forms on curves, two-forms on surfaces, three-forms on three-dimensional spaces, et cetera. As long as we have an n-form on an n-dimensional manifold we can integrate it. Given a parameterizations of the manifold $\Sigma_k \subset \mathbb{R}^n$, $k < n$, that is, a mapping $\phi^{-1} : U \subset \mathbb{R}^k \longrightarrow \Sigma_k \subset \mathbb{R}^n$, where ϕ^{-1} is invertible we can integrate an k-form on Σ_k by pulling the k-form back to $U \subset \mathbb{R}^k$ and integrating there. In this case the version of the change of coordinate formula we want to use is

Integral of k−Form on k−Dimensional Parameterized Surface Σ_k	$\displaystyle\int_{\Sigma_k} \alpha = \int_{\phi(\Sigma_k)} T^*\phi^{-1} \cdot \alpha$

When $k = 1$ then we have a curve and often write $\Sigma_1 = C$, when $k = 2$ then we have a surface and often write $\Sigma_2 = S$, and when $k = 3$ then we have a volume and often write $\Sigma_3 = V$.

7.5.2 *References and Further Reading*

Both Riemann sums and changes of variables are introduced and discussed, often at length, in virtually every book on calculus or introductory analysis, see for example Steward [43], Hubbard and Hubbard [27], or Marsden and Hoffman [31]. However, in this chapter we have generally followed the exposition of Bachman [4] at it relates directly to the differential forms case. Also see Watschap [47] or Renteln [37] for somewhat more theoretical introductions to the same material.

7.5.3 *Problems*

Question 7.14 Let C be an oriented curve in $\mathbb{R}^3$ parameterized by $\gamma(t) = \big(2t + 1, t^2, t^3\big)$ for $1 \le t \le 3$ and let $\alpha = (3x - 1)^2\, dx + 5\, dy + 2\, dz$ be a one-form on $\mathbb{R}^3$. Find $\int_C \alpha$.

Question 7.15 Let C be an oriented curve in $\mathbb{R}^3$ parameterized by $\gamma(t) = \big(1 - t^2t^3 - t, 0\big)$ for $0 \le t \le 2$. Notice that the curve C lies in $\mathbb{R}^2 \subset \mathbb{R}^3$. Let $\alpha = -5xy\, dx + dz$ be a one-form on $\mathbb{R}^3$. Find $\int_C \alpha$.

Question 7.16 Let C be a curve in $\mathbb{R}^2$ parameterized by $\gamma(t) = \big(t^2, 2t + 1\big)$ for $1 \le t \le 2$ and let $\alpha = 4x\, dx + ydy$ be a one-form on $\mathbb{R}^2$. Find $\int_C \alpha$.

Question 7.17 Let C be the curve in the plane $\mathbb{R}^2$ which is the graph of the function $y = 1 + x^2$ for $-1 \le x \le 1$, oriented in the direction of increasing x, and let $\alpha = y\, dx + x\, dy$ be a one-form on the plane. Find $\int_C \alpha$. (While not necessary, one could parameterize the curve C as $\gamma(t) = (t, 1 + t^2)$ for $-1 \le t \le 1$.)

Question 7.18 Let C_1 and C_2 be curves in the plane $\mathbb{R}^2$ parameterized by $\gamma_1(t) = \big(t - 1, 2t - 1\big)$ for $1 \le t \le 2$ and $\gamma_2(u) = \big(2 - u, 5 - 2u\big)$ for $1 \le u \le 2$, respectively. Let $\alpha = y\, dx$ be a one-form on $\mathbb{R}^2$. Find $\int_{C_1} \alpha$ and $\int_{C_2} \alpha$. Compare the two answers and explain their relationship with each other.

Question 7.19 Let C be the unit circle in $\mathbb{R}^2$ parameterized by $\gamma(\theta) = \big(\cos(\theta), \sin(\theta)\big)$ for $0 \le \theta \le 2\pi$. Let $\alpha_1 = \frac{-y}{x^2+y^2}\, dx + \frac{x}{x^2+y^2}\, dz$ and let $\alpha_2 = \frac{x}{x^2+y^2}\, dx + \frac{y}{x^2+y^2}\, dz$ be one-forms on $\mathbb{R}^2$. Find $\int_C \alpha_1$ and $\int_C \alpha_2$. Notice that one can write $\alpha_1 = -\sin(\theta)\, dx + \cos(\theta)\, dy$ and $\alpha_2 = \cos(\theta)\, dx + \sin(\theta)\, dy$.

Question 7.20 Let C be a curve in $\mathbb{R}^2$. The curve C can be parameterized by either $\gamma_1(t) = \big(t, \sqrt{1 + t^2}\big)$ where $-1 \le t \le 1$ or $\gamma_1(\theta) = \big(\tan(\theta), \sec(\theta)\big)$ where $\frac{-\pi}{4} \le \theta \le \frac{\pi}{4}$. Let $\alpha = y^2\, dx$ be a one-form on $\mathbb{R}^2$. Find $\int_C \alpha$ using each parametrization of C.

Question 7.21 Let C be a curve in $\mathbb{R}^2$. The curve C can be parameterized as $\gamma_1(t) = \big(t, t^2\big),\ -2 \le t \le 2$ or as $\gamma_2(u) = \big(u^3 + u, u^6 + 2u^4 + u^2\big),\ -1 \le u \le 1$ or as $\gamma_3(\theta) = \big(4\sin(\theta), 16\sin^2(\theta)\big),\ \frac{-\pi}{6} \le \theta \le \frac{\pi}{6}$. If $\alpha = 5xy\, dy$ is a one-form on $\mathbb{R}^2$ find $\int_C \alpha$ using all three parameterizations.

Question 7.22 Consider the surface S in $\mathbb{R}^3$ parameterized by $\phi(x, y) = \big(x, y, x^3 + xy^2\big)$ for $1 \le x \le 3$, $1 \le y \le 2$ and the two-form $\beta = 2xyz\, dx \wedge dy + 6(2x^3 - y^3)\, dy \wedge dz + z\, dz \wedge dx$ on $\mathbb{R}^3$. Find $\int_S \beta$.

Question 7.23 Let U be the top half of the unit disk, $U = \big\{(x, y)\big| -1 \le x \le 1,\ 0 \le y \le \sqrt{1 - x^2}\big\}$. Let the surface S in $\mathbb{R}^3$ be parameterized by $\phi(x, y) = \big(x, y, x^2 + y^2\big)$, where $(x, y) \in U$ and let $\beta = 6y\, dx \wedge dy$ be a two-form on $\mathbb{R}^3$. Find $\int_S \beta$.

Question 7.24 Let U be the triangular region $U = \big\{(u, v)\big|0 \le u \le 1,\ 0 \le v \le 1 - u\big\}$ in $\mathbb{R}^2$. Let the surface S in $\mathbb{R}^3$ be parameterized by $\phi(u, v) = \big(u - v, uv, u^2 + (v+1)^2\big)$ for $(u, v) \in U$ and let $\beta = -(1 + x^2)\, dx \wedge dy + dy \wedge dz$. Find $\int_S \beta$.

Question 7.25 Let $U = \big\{(u, v)\big| -1 \le u \le 1,\ -2 \le v \le 2\big\} \subset \mathbb{R}^2$ and let the surface S in $\mathbb{R}^3$ be parameterized by $\phi(u, v) = \big(u, v, u^3 + 3uv^2\big)$ where $(u, v) \in U$. Let $\beta = -dy \wedge dz + dz \wedge dx$. Find $\int_S \beta$.

Question 7.26 Let $U = \big\{(u, v)\big|0 \le u \le 1,\ 0 \le v \le 2\big\} \subset \mathbb{R}^2$ and let the surface S in $\mathbb{R}^3$ be parameterized by $\phi(u, v) = \big(u, v, u^2 + v^3\big)$ where $(u, v) \in U$. Let $\beta = 31x^2y^3\, dx \wedge dy + 5xz\, dy \wedge dz + 7yz\, dz \wedge dx$. Find $\int_S \beta$.

Question 7.27 Let $U = \big\{(u, v)\big|u^2 + v^2 \le 9\big\} \subset \mathbb{R}^2$ and let the surface S in $\mathbb{R}^3$ be parameterized by $\phi(u, v) = \big(u, v, 7 + (u^2 + v^2)\big)$ where $(u, v) \in U$. Let $\beta = z^2\, dx \wedge dy$. Find $\int_S \beta$.

Question 7.28 Let $U = \big\{(u, v)\big|u^2 + v^2 \le 9\big\} \subset \mathbb{R}^2$ and let the surface S in $\mathbb{R}^3$ be parameterized by $\phi(u, v) = \big(u, v, 25 - (u^2 + v^2)\big)$ where $(u, v) \in U$. Let $\beta = z^2\, dx \wedge dy$. Find $\int_S \beta$.

Question 7.29 Let $U = \big\{(u, v)\big|0 \le u \le 1,\ 0 \le v \le 2\big\} \subset \mathbb{R}^2$ and let the surface S in $\mathbb{R}^3$ be parameterized by $\phi(u, v) = \big(u + v, u - v^2, uv\big)$ where $(u, v) \in U$. Let $\beta = 6z\, dz \wedge dx$. Find $\int_S \beta$.

Question 7.30 Let $U = \big\{(u, v)\big|1 \le u \le 2,\ 0 \le v \le 2\pi\big\} \subset \mathbb{R}^2$ and let the surface S in $\mathbb{R}^3$ be parameterized by $\phi(u, v) = \big(u^3\cos(v), u^3\sin(v), u^2\big)$ where $(u, v) \in U$. Let $\beta = 2z\, dx \wedge dy - x\, dy \wedge dz - y\, dz \wedge dx$. Find $\int_S \beta$.

Question 7.31 Let $\alpha = xy\, dx \wedge dy$ be a two-form on $\mathbb{R}^2$ and let D be the disk centered at the origin with radius 3. Find $\int_D \alpha$ using a polar change of coordinates.

Question 7.32 Let $\alpha = (3x + 4y^2)\, dx \wedge dy$ be a two-form on $\mathbb{R}^2$ and let R be the region in the upper half plane bounded by the circles $x^2 + y^2 = 1$ and $x^2 + y^2 = 9$. Find $\int_R \alpha$ using a polar change of coordinates.

Question 7.33 Let $\alpha = (x + y)\, dx \wedge dy$ be a two-form on $\mathbb{R}^2$ and let R be the region to the left of the y-axis and bounded by the circles $x^2 + y^2 = 1$ and $x^2 + y^2 = 9$. Find $\int_R \alpha$ using a polar change of coordinates.

Question 7.34 Let $\alpha = \cos(x^2 + y^2)\, dx \wedge dy$ be a two-form on $\mathbb{R}^2$ and let R be the region above the x-axis and inside the circle $x^2 + y^2 = 4$. Find $\int_R \alpha$ using a polar change of coordinates.

Question 7.35 Let $V \subset \mathbb{R}^3$ be the region that lies within the cylinder $x^2 + y^2 = 1$ and between the planes $z = 0$ and $z = 5$ and let $\beta = x\, dx \wedge dy \wedge dz$ be a three-form on $\mathbb{R}^3$. Find $\int_V \beta$ using a cylindrical change of coordinates.

Question 7.36 Let $V \subset \mathbb{R}^3$ be the region that lies within the cylinder $x^2 + y^2 = 25$ and between the planes $z = -5$ and $z = 5$ and let $\beta = \sqrt{x^2 + y^2}\, dx \wedge dy \wedge dz$ be a three-form on $\mathbb{R}^3$. Find $\int_V \beta$ using a cylindrical change of coordinates.

Question 7.37 Let $V = \big\{(x, y, z)\big| -2 \le x \le 2,\ -\sqrt{4 - x^2} \le y \le \sqrt{4 - x^2},\ \sqrt{x^2 + y^2} \le z \le 2\big\} \subset \mathbb{R}^3$ and let $\beta = (x^2 + y^2)\, dx \wedge dy \wedge dz$ be a three-form on $\mathbb{R}^3$. Find $\int_V \beta$ using a cylindrical change of coordinates.

Question 7.38 Let $V \subset \mathbb{R}^3$ be the region that lies between the cylinders $x^2 + y^2 = 1$ and $x^2 + y + 2 = 4$ and between the planes $z = 0$ and $z = x + 2$ and let $\beta = y\, dx \wedge dy \wedge dz$ be a three-form on $\mathbb{R}^3$. Find $\int_V \beta$ using a cylindrical change of coordinates.

Question 7.39 Let $V = \big\{(x, y, z)\big|x^2 + y^2 + z^2 \le 1\big\}$, the unit ball in $\mathbb{R}^3$, and let $\beta = e^{(x^2+y^2+z^2)^{3/2}}\, dx \wedge dy \wedge dz$ be a three-form on $\mathbb{R}^3$. Find $\int_V \beta$ using a spherical change of coordinates.

Question 7.40 Let $V = \big\{(x, y, z)\big|x^2 + y^2 + z^2 \le 1\big\}$, the unit ball in $\mathbb{R}^3$, and let $\beta = e^{(x^2+y^2+z^2)^{3/2}}\, dx \wedge dy \wedge dz$ be a three-form on $\mathbb{R}^3$. Find $\int_V \beta$ using a spherical change of coordinates.

Question 7.41 Let $V = \left\{(x, y, z) \middle| x^2 + y^2 + z^2 \le 5\right\}$, the ball in $\mathbb{R}^3$ with the origin as the center and radius 5, and let $\beta = (x^2 + y^2 + z^2)^2 \, dx \wedge dy \wedge dz$ be a three-form on $\mathbb{R}^3$. Find $\int_V \beta$ using a spherical change of coordinates.

Question 7.42 Find the volume of the region in $\mathbb{R}^3$ above the cone $z = \sqrt{x^2 + y^2}$ and below the sphere $x^2 + y^2 + z^2 = z$. This means the three-form we want to integrate is the volume form $dx \wedge dy \wedge dz$. Use a spherical change of coordinates.

Chapter 8
Poincaré Lemma

Briefly put, the Poincaré lemma is as follows.

Theorem 8.1 (Poincaré Lemma) *Every closed form on $\mathbb{R}^n$ is exact.*

Section one gives a detailed introduction to the Poincaré lemma and what this means. Strictly speaking, the whole space $\mathbb{R}^n$ is not necessary. The proof works on any star-shaped region that is contractible to a point, but as this plays no essential role in the proof, to keep everything as clear as possible we will simply assume the manifolds are all of $\mathbb{R}^n$, $n \geq 0$.

The proof of the Poncareé lemma is based on induction. In case you have not yet had a class in mathematical proofs we will proceed slowly and carefully, trying to fully explain each step. While this is not a class on proof techniques, hopefully when you are done with this chapter you will have a good idea of what proofs by induction are like. Section two covers what is called the base case of the induction proof while section three covers the general case of the proof.

The Poincaré lemma plays an important role in a branch of mathematics called de Rham cohomology which is briefly introduced in Appendix B.

8.1 Introduction to the Poincaré Lemma

The Poincaré lemma states that *every closed form on $\mathbb{R}^n$ is exact*. A differential form α is called **closed** if $d\alpha = 0$. A differential form α is called **exact** if there is another differential form β such that $\alpha = d\beta$. Obviously, if α is an exact k-form then β must be a $(k-1)$-form. So, another way of phrasing the Poincaré lemma is to say that *if α is a k-form on $\mathbb{R}^n$ such that $d\alpha = 0$, then there exists some $(k-1)$-form β such that $\alpha = d\beta$.*

We begin by considering the exact forms on some manifold M. A k-form is called exact if it is equal to the exterior derivative of some $(k-1)$-form. Thus the set of all exact k-forms is exactly the set of all $d\alpha$ where α is a $(k-1)$-form. Another way of saying this is that the set of exact forms is the image under d of $\bigwedge^{k-1}(M)$. Figure 8.1 gives a Venn-like diagram of the mapping $d : \bigwedge^{k-1}(M) \rightarrow \bigwedge^{k}(M)$ to give a picture of the exact k-forms.

A k-form on M is called closed if the exterior derivative of that k-form is zero; that is, the zero $(k+1)$-form. The zero $(k+1)$-form is the form that sends every set of $k+1$ vectors to zero and should not be confused with the zero-forms, which are functions. We generally denote the zero k-form, at any point p, simply with 0. Thus the set of all closed k-forms are those k-forms α where $d\alpha = 0$. Another way of saying this is that the set of closed k-forms is the pre-image of the zero $(k+1)$-form under the mapping $d : \bigwedge^{k}(M) \rightarrow \bigwedge^{k+1}(M)$. Figure 8.2 uses a Venn-like diagram to give a picture of the closed k-forms.

Now the question becomes, how do closed k-forms and exact k-forms relate to each other. A first guess may be go along the lines of Fig. 8.3 where a number of different possibilities are shown:

- The exact forms are a subset of the closed forms.
- The exact forms are the same as the closed forms.
- The closed forms are a subset of the exact forms.
- The closed and exact forms are mutually exclusive.
- The closed and exact forms are not subsets of each other but are not mutually exclusive.

J. P. Fortney, *A Visual Introduction to Differential Forms and Calculus on Manifolds*,
https://doi.org/10.1007/978-3-319-96992-3_8

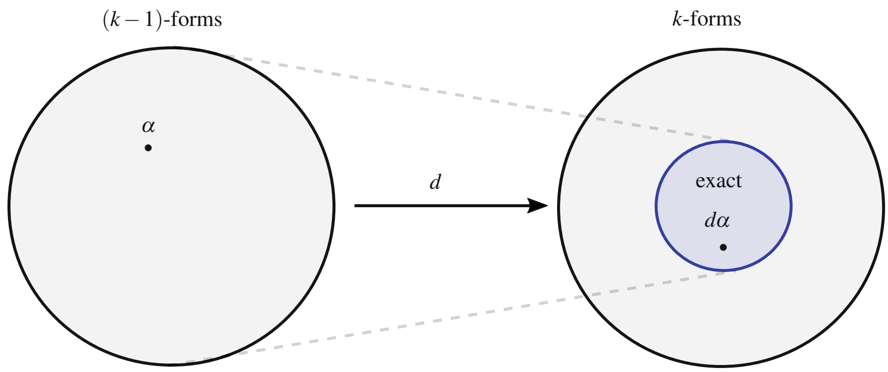


Fig. 8.1 The mapping $d : \bigwedge^{k-1}(M) \rightarrow \bigwedge^{k}(M)$ illustrated using a Venn-like diagram. The exact k-forms are the image, under d, of $\bigwedge^{k-1}(M)$

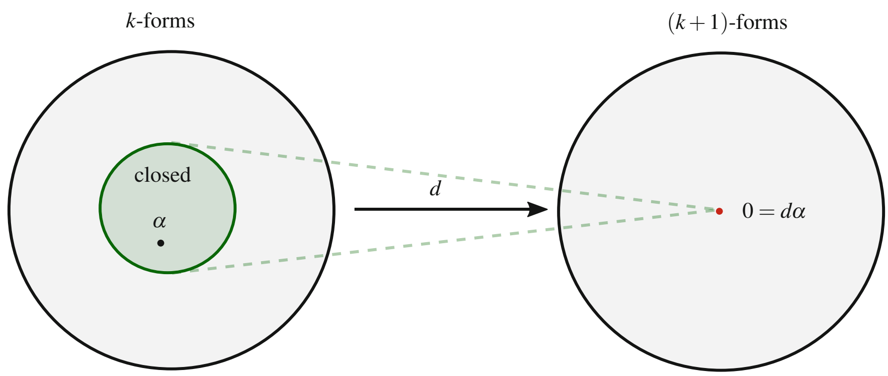


Fig. 8.2 The mapping $d : \bigwedge^{k}(M) \rightarrow \bigwedge^{k+1}(M)$ illustrated using a Venn-like diagram. The closed k-forms are the pre-image, under d, of the zero $(k+1)$-form 0

However, a little reflection will almost immediately eliminate the last three possibilities. All we have to do is recognize two facts. First, every exact k-form is $d\alpha$ for some $(k-1)$-form α, and second that no matter what the form is, we have $dd\alpha = 0$. This means that every exact form is also a closed form, thereby eliminating the last three possibilities of Fig. 8.3. That leaves the first two possibilities; Fig. 8.4.

So, which of the two possibilities in Fig. 8.4 is it? Are exact forms a subset of closed forms or are exact forms the same as closed forms? It turns out that the answer to this question depends on what manifold M the forms are on. The Poincaré lemma tells us that exact forms are the same as closed forms when the manifold is $\mathbb{R}^n$ for any number n. But if the manifold is something different from $\mathbb{R}^n$ then this may not be the case. The branch of mathematics called de Rham cohomology looks at how different the set of closed forms on M is from the set of exact forms on M and uses this information to figure out certain properties of the underlying manifold M.

Figure 8.5 is a commutative diagram showing all the spaces of forms on $\mathbb{R}^n$, which extends to infinity to the bottom and to the right. Clearly, exterior differentiation gives the sequence of maps

$$\bigwedge^0(\mathbb{R}^n) \xrightarrow{d} \bigwedge^1(\mathbb{R}^n) \xrightarrow{d} \bigwedge^2(\mathbb{R}^n) \xrightarrow{d} \bigwedge^3(\mathbb{R}^n) \xrightarrow{d} \cdots$$

for the manifold $\mathbb{R}^n$, for $n = 0, 1, 2, \ldots$. The map $\mathscr{K}$ is needed to prove the Poincaré lemma and will be explained in the next section. This gives the columns in the commutative diagram. The mappings

$$\bigwedge^k(\mathbb{R}^n) \underset{\mathscr{C}}{\overset{\mathscr{L}}{\rightleftarrows}} \bigwedge^k\left(\mathbb{R}^{n+1}\right),$$

where $k = 1, 2, \ldots$ and $n = 0, 1, 2, \ldots$ are mappings that we need to use to prove the Poincaré lemma and which will also be introduced in the next section. As we move through the proof of the Poincaré lemma we will refer back to this commutative diagram. Our goal is to show that every closed k-form α, for $k > 0$, on $\mathbb{R}^n$, for $n > 0$, is exact. In other words, we want to show that for every $\alpha \in \bigwedge^k(\mathbb{R}^n)$, where $d\alpha = 0$, there is a $\beta \in \bigwedge^{k-1}(\mathbb{R}^n)$ such that $\alpha = d\beta$. Actually, we will not be finding this β explicitly, instead we will be using induction on n. That is, if we know it is true for k-forms on $\mathbb{R}^n$ then we

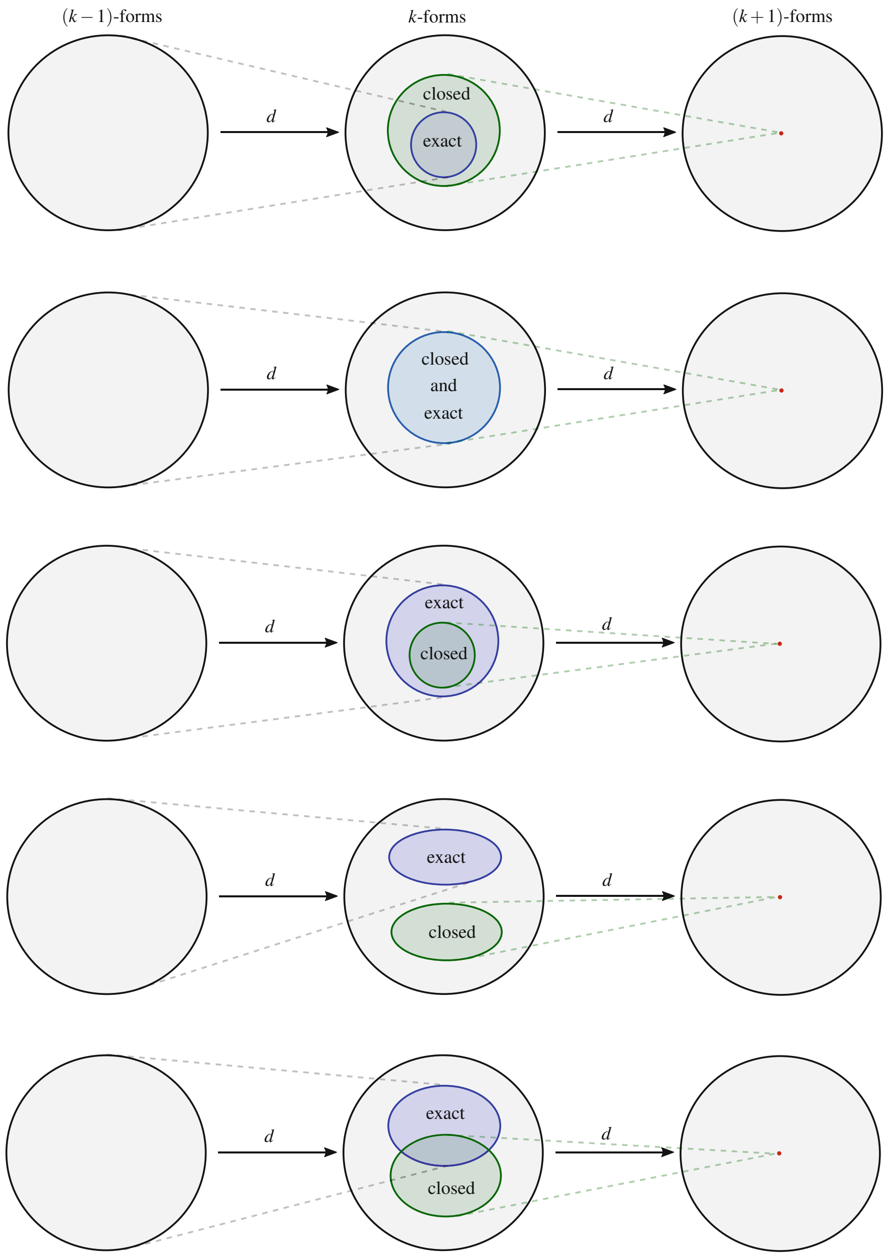


Fig. 8.3 Various possibilities on how closed and exact k-forms may be related to each other

will be able to prove that it is true for $\mathbb{R}^{n+1}$. In Fig. 8.6 we give a Venn-like diagram representation of the columns of the commutative diagram in Fig. 8.5.

8.2 The Base Case and a Simple Example Case

First of all we recall that zero-forms α on $\mathbb{R}^n$ are exactly the real-valued functions on $\mathbb{R}^n$. The zero-forms on $\mathbb{R}^n$ are denoted by $\bigwedge^0(\mathbb{R}^n)$, which is the top row of Fig. 8.5. While it is certainly possible to have closed zero forms on $\mathbb{R}^n$ it is not possible to have an exact zero form on $\mathbb{R}^n$ since there are no such things as (-1)-forms. Hence the Poincaré lemma does not apply to zero forms but only to k-forms where $k > 0$.

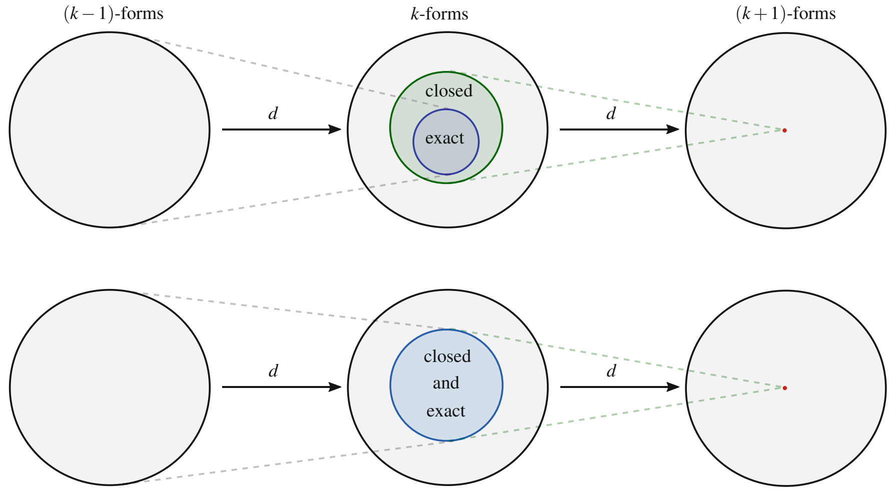


Fig. 8.4 Are exact forms a subset of closed forms (above) or are exact forms the same as closed forms (below)? The answer depends on the manifold M

$$
\begin{tikzcd}
\bigwedge^0(\mathbb{R}^0) \arrow[d, shift left, "d"] & \bigwedge^0(\mathbb{R}^1) \arrow[d, shift left, "d"] & \bigwedge^0(\mathbb{R}^2) \arrow[d, shift left, "d"] & \bigwedge^0(\mathbb{R}^3) \arrow[d, shift left, "d"] & \\
\bigwedge^1(\mathbb{R}^0) \arrow[u, shift left, "\mathscr{K}"] \arrow[d, shift left, "d"] \arrow[r, shift left, "\mathscr{C}"] & \bigwedge^1(\mathbb{R}^1) \arrow[u, shift left, "\mathscr{K}"] \arrow[d, shift left, "d"] \arrow[l, shift left, "\mathscr{Z}"] \arrow[r, shift left, "\mathscr{C}"] & \bigwedge^1(\mathbb{R}^2) \arrow[u, shift left, "\mathscr{K}"] \arrow[d, shift left, "d"] \arrow[l, shift left, "\mathscr{Z}"] \arrow[r, shift left, "\mathscr{C}"] & \bigwedge^1(\mathbb{R}^3) \arrow[u, shift left, "\mathscr{K}"] \arrow[d, shift left, "d"] \arrow[l, shift left, "\mathscr{Z}"] \arrow[r, shift left, "\mathscr{C}"] & \cdots \arrow[l, shift left, "\mathscr{Z}"] \\
\bigwedge^2(\mathbb{R}^0) \arrow[u, shift left, "\mathscr{K}"] \arrow[d, shift left, "d"] \arrow[r, shift left, "\mathscr{C}"] & \bigwedge^2(\mathbb{R}^1) \arrow[u, shift left, "\mathscr{K}"] \arrow[d, shift left, "d"] \arrow[l, shift left, "\mathscr{Z}"] \arrow[r, shift left, "\mathscr{C}"] & \bigwedge^2(\mathbb{R}^2) \arrow[u, shift left, "\mathscr{K}"] \arrow[d, shift left, "d"] \arrow[l, shift left, "\mathscr{Z}"] \arrow[r, shift left, "\mathscr{C}"] & \bigwedge^2(\mathbb{R}^3) \arrow[u, shift left, "\mathscr{K}"] \arrow[d, shift left, "d"] \arrow[l, shift left, "\mathscr{Z}"] \arrow[r, shift left, "\mathscr{C}"] & \cdots \arrow[l, shift left, "\mathscr{Z}"] \\
\bigwedge^3(\mathbb{R}^0) \arrow[u, shift left, "\mathscr{K}"] \arrow[d, shift left, "d"] \arrow[r, shift left, "\mathscr{C}"] & \bigwedge^3(\mathbb{R}^1) \arrow[u, shift left, "\mathscr{K}"] \arrow[d, shift left, "d"] \arrow[l, shift left, "\mathscr{Z}"] \arrow[r, shift left, "\mathscr{C}"] & \bigwedge^3(\mathbb{R}^2) \arrow[u, shift left, "\mathscr{K}"] \arrow[d, shift left, "d"] \arrow[l, shift left, "\mathscr{Z}"] \arrow[r, shift left, "\mathscr{C}"] & \bigwedge^3(\mathbb{R}^3) \arrow[u, shift left, "\mathscr{K}"] \arrow[d, shift left, "d"] \arrow[l, shift left, "\mathscr{Z}"] \arrow[r, shift left, "\mathscr{C}"] & \cdots \arrow[l, shift left, "\mathscr{Z}"] \\
\bigwedge^4(\mathbb{R}^0) \arrow[u, shift left, "\mathscr{K}"] \arrow[d, shift left, "d"] \arrow[r, shift left, "\mathscr{C}"] & \bigwedge^4(\mathbb{R}^1) \arrow[u, shift left, "\mathscr{K}"] \arrow[d, shift left, "d"] \arrow[l, shift left, "\mathscr{Z}"] \arrow[r, shift left, "\mathscr{C}"] & \bigwedge^4(\mathbb{R}^2) \arrow[u, shift left, "\mathscr{K}"] \arrow[d, shift left, "d"] \arrow[l, shift left, "\mathscr{Z}"] \arrow[r, shift left, "\mathscr{C}"] & \bigwedge^4(\mathbb{R}^3) \arrow[u, shift left, "\mathscr{K}"] \arrow[d, shift left, "d"] \arrow[l, shift left, "\mathscr{Z}"] \arrow[r, shift left, "\mathscr{C}"] & \cdots \arrow[l, shift left, "\mathscr{Z}"] \\
\vdots \arrow[u, shift left, "\mathscr{K}"] & \vdots \arrow[u, shift left, "\mathscr{K}"] & \vdots \arrow[u, shift left, "\mathscr{K}"] & \vdots \arrow[u, shift left, "\mathscr{K}"] &
\end{tikzcd}
$$

Fig. 8.5 Commutative diagram that will be helpful in understanding the proof of the Poincaré lemma

Next we will prove our base case. Every proof by induction requires you to first prove a base case. Usually, as will be the case here, the base case is very easy to prove. That is part of the beauty of proofs by induction. Here our base case is the first column of Fig. 8.5, which is also shown as the first row in Fig. 8.6 using Venn-like diagrams. $\mathbb{R}^0$ is a set of exactly one point, $\mathbb{R}^0 = \{0\}$, so the space $\bigwedge^0(\mathbb{R}^0)$ is just the set of real-valued functions on a point, $f : \{0\} \to \mathbb{R}$. This is shown as the first space in the upper left hand corner of Fig. 8.5.

We also have previously seen that $\bigwedge^k(\mathbb{R}^n) = \{0\}$ for $k > n$; that is, all k-forms on $\mathbb{R}^n$, where $k > n$, are the zero k-form. What we mean is that the zero k-form sends every set of k vectors to zero. Thus, for $k > n$ if $\alpha \in \bigwedge^k(\mathbb{R}^n)$ then $\alpha(v_1, \dots v_k) = 0$ for every set of vectors $v_1, \dots, v_k$ and hence we write $\alpha \equiv 0$. Thus $\bigwedge^k(\mathbb{R}^0) = \{0\}$ for $k > 0$. These spaces are all shown as a single point in Fig. 8.6.

Now we proceed to prove our base case. The proof of the base case is essentially trivial but still happens in two simple steps:

1. If $\alpha \in \bigwedge^1(\mathbb{R}^0)$ is closed show that it is exact.
2. If $\alpha \in \bigwedge^k(\mathbb{R}^0)$ for $k > 1$ is closed show that it is exact.

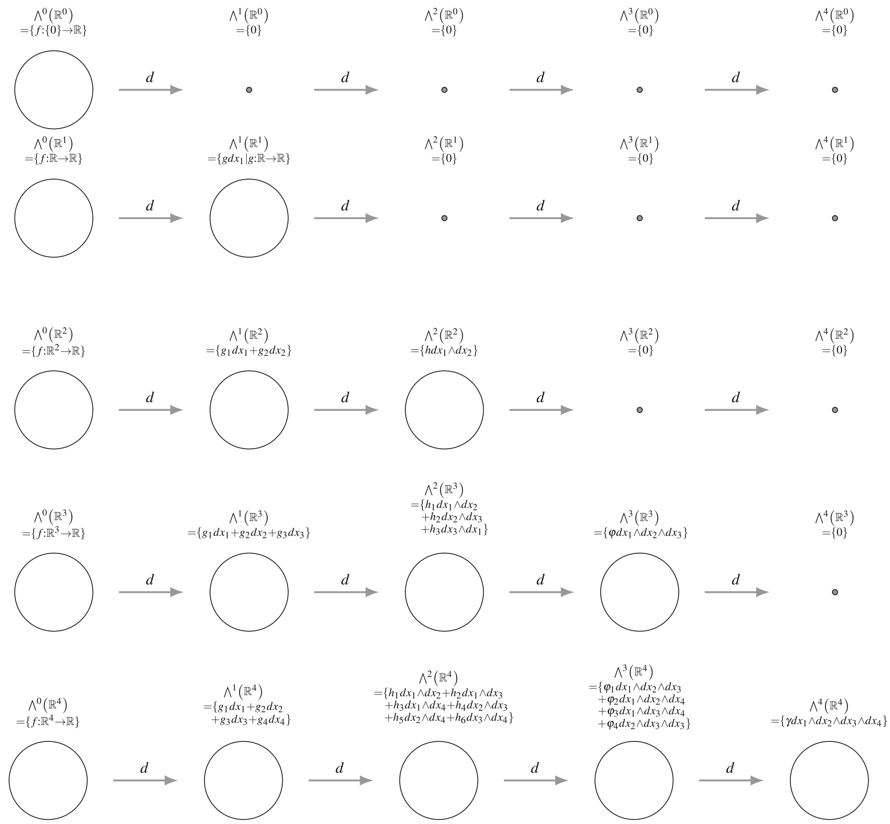


Fig. 8.6 Here the *columns* of 8.5 are shown as *rows* using Venn-like diagrams. The base case for the induction hypothesis is the first column of Fig. 8.5, which is shown as the top row here. The column of the commutative diagram in Fig. 8.5 that is associated with our mid-level sample case is shown as the fourth row here

In the first case, if $\alpha \in \bigwedge^1(\mathbb{R}^0)$ then we must have $\alpha = 0$ since the only form in $\bigwedge^1(\mathbb{R}^0)$ is the zero one-form. And clearly $d\alpha = d0 = 0$ and so it is closed. In order to show it is exact we must show that $\alpha = df$ for some zero-form f. But consider any function $f \in \bigwedge^0(\mathbb{R}^0)$. Since $f : \{0\} \to \mathbb{R}$ then $f(0)$ is simply an unchanging real number. The exterior derivative of f finds how f changes as we move in any direction. But there is no direction to move and no change that can happen so we must have $df = 0 = \alpha$ for every $f \in \bigwedge^0(\mathbb{R}^0)$. Thus the closed α is also exact, thereby proving the first case.

The second case is even easier. It is obvious that $\alpha \in \bigwedge^k(\mathbb{R}^0)$ for $k > 1$ is closed since $d\alpha = d0 = 0$. But is is also obvious that α must be exact since if $\beta \in \bigwedge^{k-1}(\mathbb{R}^0)$ then $d\beta = d0 = 0 = \alpha$. Thus we have also shown that any closed form in $\bigwedge^k(\mathbb{R}^0)$, for any $k > 1$, is also exact. Putting this together we have that any closed form on $\mathbb{R}^0$ is also exact, thereby proving our base case.

Now, instead of proceeding to the general case we will spend the rest of this section doing a somewhat mid-level sample case. We do this in order to help you become familiar with the strategy we will use in the general case. The general case is notationally cumbersome and so it is easier to understand the basic strategy when the notation is not so overwhelming. We shall show that closed one-forms on $\mathbb{R}^3$ are exact. The fourth row of Fig. 8.6 shows the part of the commutative diagram in Fig. 8.5 that is associated with $\mathbb{R}^3$. In order to do our sample case we will need to introduce several mappings. We will have more to say about these mappings after the sample case and before we do the general case. The first mapping we will denote by $\mathscr{Z}$ and the second mapping by $\mathscr{C}$;

$$\mathscr{Z} : \bigwedge^k(\mathbb{R}^n) \longrightarrow \bigwedge^k(\mathbb{R}^{n-1}),$$

$$\mathscr{C} : \bigwedge^k(\mathbb{R}^n) \longrightarrow \bigwedge^k(\mathbb{R}^{n+1}).$$

In essence the mapping $\mathscr{Z}$ "squishes" a k-form on $\mathbb{R}^n$ to a k-form on $\mathbb{R}^{n-1}$ while the mapping $\mathscr{C}$ "expands" a k-form on $\mathbb{R}^n$ to a k-form on $\mathbb{R}^{n+1}$.

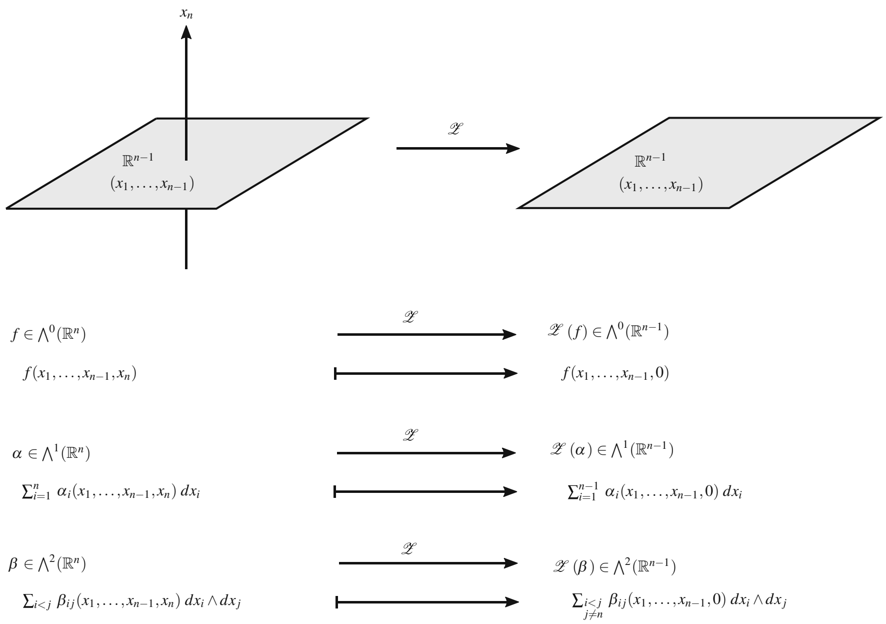


Fig. 8.7 The mapping $\mathscr{Z} : \bigwedge^k(\mathbb{R}^n) \to \bigwedge^k(\mathbb{R}^{n-1})$ is shown. The mapping "squishes" k-forms on $\mathbb{R}^n$ to k-forms on $\mathbb{R}^{n-1}$. This is shown concretely for zero-forms, one-forms, and two-forms. $\mathscr{Z}$ sets $x_n = 0$ and kills all terms that contain dx_n

Suppose that the coordinate functions of $\mathbb{R}^n$ are $x_1, \ldots, x_n$. As usual, we are of course being a bit imprecise and will use the same notation, $x_1, \ldots, x_n$, to both denote real-valued coordinate functions on $\mathbb{R}^n$ as well as the numerical values those functions give, which denotes the point. That is, if $p \in \mathbb{R}^n$ and x_i is a coordinate function $x_i : \mathbb{R}^n \to \mathbb{R}$, then we have $x_i(p) = x_i$, where the x_i on the left hand side is the coordinate function and the x_i on the right hand side is a numerical value. This allows us to write the point $p = (x_1, \ldots, x_n)$. We shall continue with this ambiguity throughout and hope that it will not cause you too much confusion.

A zero-form on $\mathbb{R}^n$ at the point $p = (x_1, \ldots, x_n)$ is simply a function $f(p) = f(x_1, \ldots, x_n)$, one-form α on $\mathbb{R}^n$ at a point p is written as $\alpha = \sum_{i=1}^n \alpha_i(p)\ dx_i = \sum_{i=1}^n \alpha_i(x_1, \ldots, x_n)\ dx_i$, a two-form at the point p can be written is $\beta = \sum_{i<j} \beta_{ij}(p)\ dx_i \wedge dx_j = \sum_{i<j} \beta_{ij}(x_1, \ldots, x_n)\ dx_i \wedge dx_j$, and so on. The mapping $\mathscr{Z} : \bigwedge^k(\mathbb{R}^n) \to \bigwedge^k(\mathbb{R}^{n-1})$ essentially restricts k-forms to the subspace $\mathbb{R}^{n-1}$ of $\mathbb{R}^n$. See Fig. 8.7. It does this by simply replacing the x_n value with 0 and killing every term that has dx_n in it. Thus we have

$$
\begin{aligned}
f(x_1, \ldots, x_{n-1}, x_n) &\overset{\mathscr{Z}}{\longmapsto} f(x_1, \ldots, x_{n-1}, 0), \\
\sum_{i=1}^{n} \alpha_i(x_1, \ldots, x_{n-1}, x_n)\ dx_i &\overset{\mathscr{Z}}{\longmapsto} \sum_{i=1}^{n-1} \alpha_i(x_1, \ldots, x_{n-1}, 0)\ dx_i, \\
\sum_{i<j} \beta_{ij}(x_1, \ldots, x_{n-1}, x_n)\ dx_i \wedge dx_j &\overset{\mathscr{Z}}{\longmapsto} \sum_{\substack{i<j \\ j\neq n}} \beta_{ij}(x_1, \ldots, x_{n-1}, 0)\ dx_i \wedge dx_j,
\end{aligned}
$$

and so on for k-forms where $k > 2$.

The mapping $\mathscr{C}$ "expands" a k-form on $\mathbb{R}^n$ to a k-form on $\mathbb{R}^{n+1}$,

$$
\begin{aligned}
f(x_1, \ldots, x_{n-1}, x_n) &\overset{\mathscr{C}}{\longmapsto} f(x_1, \ldots, x_{n-1}, x_n), \\
\sum_{i=1}^{n} \alpha_i(x_1, \ldots, x_{n-1}, x_n)\ dx_i &\overset{\mathscr{C}}{\longmapsto} \sum_{i=1}^{n} \alpha_i(x_1, \ldots, x_{n-1}, x_n)\ dx_i, \\
\sum_{i<j} \beta_{ij}(x_1, \ldots, x_{n-1}, x_n)\ dx_i \wedge dx_j &\overset{\mathscr{C}}{\longmapsto} \sum_{i<j} \beta_{ij}(x_1, \ldots, x_{n-1}, x_n)\ dx_i \wedge dx_j.
\end{aligned}
$$

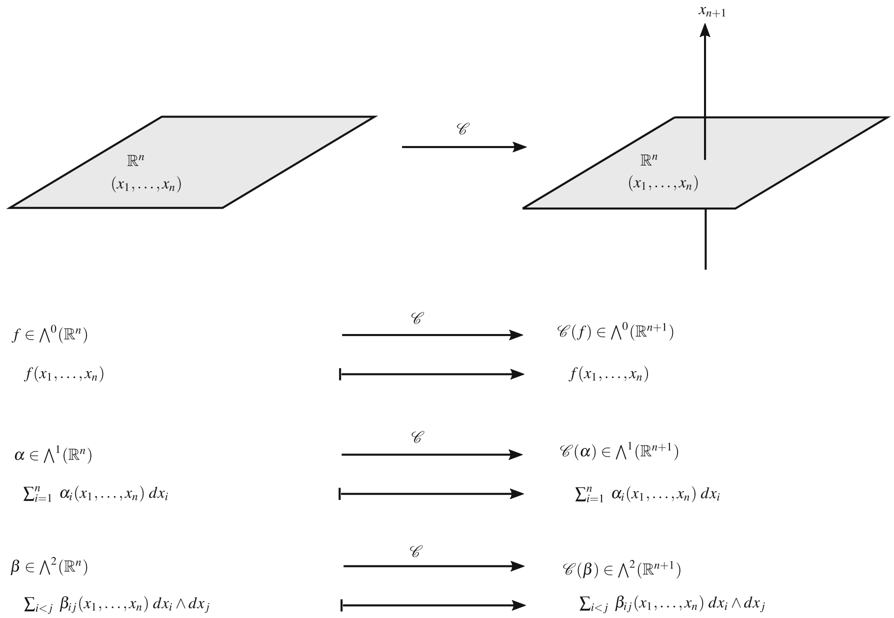


Fig. 8.8 The mapping $\mathscr{C} : \bigwedge^k(\mathbb{R}^n) \to \bigwedge^k(\mathbb{R}^{n+1})$ is shown. The mapping "expands" k-forms on $\mathbb{R}^n$ to k-forms on $\mathbb{R}^{n+1}$. The expanded k-form on $\mathbb{R}^{n+1}$ looks exactly like the k-form on $\mathbb{R}^n$. This is shown concretely for zero-forms, one-forms, and two-forms

See Fig. 8.8. Notice, the way that $f \in \bigwedge^0(\mathbb{R}^n)$ looks exactly the same as $\mathscr{C}(f) \in \bigwedge^1(\mathbb{R}^{n+1})$, which is a function that simply does not depend on the x_{n+1} variable. Similarly, $\alpha \in \bigwedge^1(\mathbb{R}^n)$ looks exactly like $C(\alpha) \in \bigwedge^1(\mathbb{R}^{n+1})$, which is a one-form on $\mathbb{R}^{n+1}$ that does not have a dx_{n+1} term and whose component functions α_i do not depend at all on the x_{n+1} variable. The same is true of the two-form $\beta \in \bigwedge^2(\mathbb{R}^n)$. $\mathscr{C}(\beta)$ looks exactly the same as β; it is two-form on $\mathbb{R}^{n+1}$ which does not have any terms involving dx_{n+1} and whose component functions β_{ij} do not depend on the x_{n+1} variable.

To see this better consider the one-form $\alpha = P(x, y, z)dx + Q(x, y, z)dy + R(x, y, z)dz \in \bigwedge^1(\mathbb{R}^3)$. Then

$$\mathscr{Z}(\alpha) = P(x, y, 0)dx + Q(x, y, 0)dy \in \bigwedge^1(\mathbb{R}^2).$$

Notice that we killed the dz term, which was $R(x, y, z)dz$. Now consider the one-form $\alpha = P(x, y)dx + Q(x, y)dy \in \bigwedge^1(\mathbb{R}^2)$ then

$$\mathscr{C}(\alpha) = P(x, y)dx + Q(x, y)dy \in \bigwedge^1(\mathbb{R}^3).$$

This one-form on $\mathbb{R}^3$ looks exactly like the one-form from $\mathbb{R}^2$, even though it is actually on $\mathbb{R}^3$. It is a one-form on $\mathbb{R}^3$ without a dz term and whose component functions do not depend on the variable z.

Question 8.1 Using $\alpha = P(x, y, z)dx + Q(x, y, z)dy + R(x, y, z)dz \in \bigwedge^1(\mathbb{R}^3)$, show that $d(\mathscr{Z}(\alpha)) = \mathscr{Z}(d(\alpha))$. We often write this as $d\mathscr{Z} = \mathscr{Z}d$.

Question 8.2 Using $\alpha = P(x, y)dx + Q(x, y)dy \in \bigwedge^1(\mathbb{R}^2)$, show that $d(\mathscr{C}(\alpha)) = \mathscr{C}(d(\alpha))$. We often write this as $d\mathscr{C} = \mathscr{C}d$.

Question 8.3 Using $\alpha = P(x, y, z)dx + Q(x, y, z)dy + R(x, y, z)dz \in \bigwedge^1(\mathbb{R}^3)$, find $\mathscr{Z}(\alpha)$ and then find $\mathscr{C}\big(\mathscr{Z}(\alpha)\big)$.

Besides $\mathscr{Z}$ and $\mathscr{C}$ we need to define one more mapping $\mathscr{K} : \bigwedge^1(\mathbb{R}^3) \to \bigwedge^0(\mathbb{R}^3)$. If $\alpha = P(x, y, z)dx + Q(x, y, z)dy + R(x, y, z)dz$ is a one-form on $\mathbb{R}^3$ then $\mathscr{K}(\alpha)$ is a zero-form, that is, a function on $\mathbb{R}^3$, defined by

$$\mathscr{K}(\alpha) = \mathscr{K}\Big(P(x, y, z)dx + Q(x, y, z)dy + R(x, y, z)dz\Big) = \int_0^z R(x, y, t)\, dt.$$

In other words, $\mathscr{K}(\alpha) : \mathbb{R}^3 \to \mathbb{R}$. How does this actually work? If we are given a point $p = (x_0, y_0, z_0)$ on $\mathbb{R}^3$ then

$$\mathscr{K}(\alpha)(p) = \mathscr{K}(\alpha)(x_0, y_0, z_0) = \int_0^{z_0} R(x_0, y_0, t)\, dt \in \mathbb{R}.$$

Since the Poincaré lemma is about showing closed forms are exact, in our proof we will need to make use of some facts that are related to closed forms. In this mid-level sample case we are dealing with a closed one-form α on $\mathbb{R}^3$. We will use the fact that $d\alpha = 0$ to find some identities that will be needed during the course of the proof.

$$\begin{aligned}
d\alpha &= d\big(P(x, y, z)dx + Q(x, y, z)dy + R(x, y, z)dz\big) \\
&= dP \wedge dx + dQ \wedge dy + dR \wedge dz \\
&= \frac{\partial P}{\partial y} dy \wedge dx + \frac{\partial P}{\partial z} dz \wedge dx + \frac{\partial Q}{\partial x} dx \wedge dy + \frac{\partial Q}{\partial z} dz \wedge dy + \frac{\partial R}{\partial x} dx \wedge dz + \frac{\partial R}{\partial y} dy \wedge dz \\
&= \left(\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y}\right) dx \wedge dy + \left(\frac{\partial R}{\partial x} - \frac{\partial P}{\partial z}\right) dx \wedge dz + \left(\frac{\partial R}{\partial y} - \frac{\partial Q}{\partial z}\right) dy \wedge dz \\
&= 0
\end{aligned}$$

thus the coefficients of each term of the two-form are equal to zero, thereby giving us the following three identities

$$\frac{\partial Q}{\partial x} = \frac{\partial P}{\partial y}, \quad \frac{\partial R}{\partial x} = \frac{\partial P}{\partial z}, \quad \frac{\partial R}{\partial y} = \frac{\partial Q}{\partial z}.$$

Now that we have all the necessary preliminaries finally pinned down we are ready to proceed with the proof of the Poincaré lemma in our sample case. We will begin by giving the overall strategy. Once we have done that we will proceed by giving the details of each step. Suppose that we know $\alpha \in \bigwedge^1(\mathbb{R}^3)$ is closed. Our goal is to show that α is also exact. That is, we want to show there is some $\beta \in \bigwedge^0(\mathbb{R}^3)$ such that $d\beta = \alpha$.

First we give the general idea of what we will do. A schematic of what we want to do is presented in Fig. 8.9. We will "squish" α down with $\mathscr{Z}$ to get $\mathscr{Z}(\alpha) \in \bigwedge^1(\mathbb{R}^2)$ which we can then show is exact by the induction hypothesis. We will and then "expand" the exact form $\mathscr{Z}(\alpha)$ with $\mathscr{C}$ to get $\mathscr{C}\big(\mathscr{Z}(\alpha)\big) \in \bigwedge^1(\mathbb{R}^3)$, which we can then also show is exact. Since $\mathscr{C}\big(\mathscr{Z}(\alpha)\big)$ is in some sense part of α we want to show that the rest of α, namely $\alpha - \mathscr{C}\big(\mathscr{Z}(\alpha)\big)$, is also exact. We can do this by using $\mathscr{K}$ on α to get $\mathscr{K}(\alpha) \in \bigwedge^0(\mathbb{R}^3)$ and then taking the exterior derivative of that to get $d\big(\mathscr{K}(\alpha)\big) \in \bigwedge^1(\mathbb{R}^3)$, which

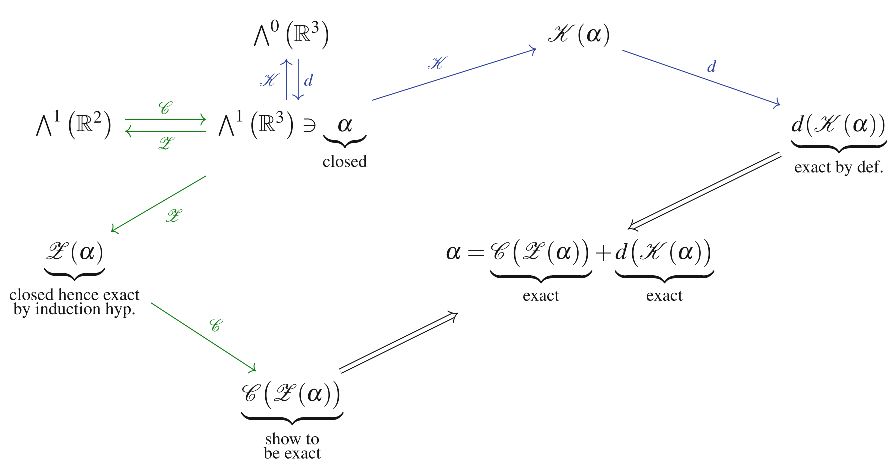


Fig. 8.9 The general idea of our proof for the sample case. For $\alpha \in \bigwedge^1(\mathbb{R}^3)$ we compute $\mathscr{K}(\alpha) \in \bigwedge^0(\mathbb{R}^3)$ and then $d\big(\mathscr{K}(\alpha)\big) \in \bigwedge^1(\mathbb{R}^3)$, which is clearly exact. We also compute $\mathscr{Z}(\alpha) \in \bigwedge^1(\mathbb{R}^2)$, which is closed and hence exact by our induction hypothesis. We then compute $\mathscr{C}\big(\mathscr{Z}(\alpha)\big) \in \bigwedge^1(\mathbb{R}^3)$, which can be shown to be exact. It is then possible to show that $\alpha = \mathscr{C}\big(\mathscr{Z}(\alpha)\big) + d\big(\mathscr{K}(\alpha)\big)$, the sum of two exact terms, which means that α itself is exact

is clearly exact. It turns out that $d\big(\mathscr{K}(\alpha)\big) = \alpha - \mathscr{C}\big(\mathscr{Z}(\alpha)\big)$. In other words we have that

$$\underbrace{d\big(\mathscr{K}(\alpha)\big)}_{\substack{\text{clearly exact} \\ \text{by definition}}} = \alpha - \underbrace{\mathscr{C}\big(\mathscr{Z}(\alpha)\big)}_{\substack{\text{exacty by} \\ \text{induction hyp.}}}.$$

But this means that

$$\begin{aligned}
\alpha &= \underbrace{\mathscr{C}\big(\mathscr{Z}(\alpha)\big)}_{\substack{\text{exact so there} \\ \text{exists some } \omega \\ \text{s.t. } \mathscr{C}(\mathscr{Z}(\alpha))=d\omega}} + \underbrace{d\big(\mathscr{K}(\alpha)\big)}_{\substack{\text{clearly exact} \\ \text{by definition}}} \\
&= d\big(\omega\big) + d\big(\mathscr{K}(\alpha)\big) \\
&= d\big(\omega + \mathscr{K}(\alpha)\big) \qquad \text{by linearity of } d
\end{aligned}$$

and hence α is exact itself, which is what we wanted to show. Notice that we never explicitly showed what ω was. In other words, while we logically reasoned that ω must exist we never actually found an explicit formula for ω.

Now we have given you the overall strategy that we will employ we are ready for the details. We now have to show two things, first that $\mathscr{C}\big(\mathscr{Z}(\alpha)\big)$ is exact by the induction hypotheses and second that $d\big(\mathscr{K}(\alpha)\big) = \alpha - \mathscr{C}\big(\mathscr{Z}(\alpha)\big)$. Showing these two things is actually the guts of the proof. Since α is closed by definition of closed we have that $d\alpha = 0$. Using the fact that $d\mathscr{Z} = \mathscr{Z}d$ we have

$$d\big(\mathscr{Z}(\alpha)\big) = \mathscr{Z}(d\alpha) = \mathscr{Z}(0) = 0 \quad \Rightarrow \quad \mathscr{Z}(\alpha) \in \bigwedge^1(\mathbb{R}^2) \text{ is closed.}$$

The ***induction hypothesis*** *is that all closed one-forms on* $\mathbb{R}^2$ *are already known to be exact*. Since $\mathscr{Z}(\alpha)$ is a one-form on $\mathbb{R}^2$ which is now known to be closed, we know, by the inductive hypothesis, that $\mathscr{Z}(\alpha)$ is exact. that means there is some zero-form ω on $\mathbb{R}^2$ such that $\mathscr{Z}(\alpha) = d\omega$. Now to show that $\mathscr{C}(\mathscr{Z}(\alpha)) \in \bigwedge^1(\mathbb{R}^3)$ is exact we use the fact that $d\mathscr{C} = \mathscr{C}d$,

$$\mathscr{C}\big(\mathscr{Z}(\alpha)\big) = \mathscr{C}(d\omega) = d(\mathscr{C}\omega) \quad \Rightarrow \quad \mathscr{C}\big(\mathscr{Z}(\alpha)\big) \in \bigwedge^1(\mathbb{R}^3) \text{ is exact.}$$

Now we will show that $d\big(\mathscr{K}(\alpha)\big)$ is the part of α that is left after subtracting the exact part $\mathscr{C}\big(\mathscr{Z}(\alpha)\big)$; that is, that $d\big(\mathscr{K}(\alpha)\big) = \alpha - \mathscr{C}\big(\mathscr{Z}(\alpha)\big)$. This is done using an explicit calculation

$$\begin{aligned}
d\big(\mathscr{K}(\alpha)\big) &= d\left(\mathscr{K}\Big(P(x,y,z)dx + Q(x,y,z)dy + R(x,y,z)dz\Big)\right) \\
&= d\left(\int_0^z R(x,y,t)\,dt\right) \\
&= \frac{\partial}{\partial x}\left(\int_0^z R(x,y,t)\,dt\right)dx + \frac{\partial}{\partial y}\left(\int_0^z R(x,y,t)\,dt\right)dy + \frac{\partial}{\partial z}\left(\int_0^z R(x,y,t)\,dt\right)dz \\
&= \underbrace{\left(\int_0^z \frac{\partial}{\partial x}R(x,y,t)\,dt\right)}_{\int \text{ has no depenance on } x}dx + \underbrace{\left(\int_0^z \frac{\partial}{\partial y}R(x,y,t)\,dt\right)}_{\int \text{ has no depenance on } y}dy + \underbrace{R(x,y,z)}_{\substack{\text{fundamental} \\ \text{theorem} \\ \text{of calculus}}}dz \\
&= \underbrace{\left(\int_0^z \frac{\partial P(x,y,t)}{\partial z}\,dt\right)}_{\alpha \text{ closed} \Rightarrow \frac{\partial R}{\partial x} = \frac{\partial P}{\partial z}}dx + \underbrace{\left(\int_0^z \frac{\partial Q(x,y,t)}{\partial z}\,dt\right)}_{\alpha \text{ closed} \Rightarrow \frac{\partial R}{\partial y} = \frac{\partial Q}{\partial z}}dy + R(x,y,z)dz \\
&= \Big[P(x,y,z) - P(x,y,0)\Big]dx + \Big[Q(x,y,z) - Q(x,y,0)\Big]dy + R(x,y,z)dz \\
&= \Big(P(x,y,z)dx + Q(x,y,z)dy + R(x,y,z)dz\Big) - \Big(P(x,y,0)dx + Q(x,y,0)dy\Big)
\end{aligned}$$

$$
\begin{aligned}
&= \alpha - \mathscr{C}\big(\mathscr{Z}(\alpha)\big) \\
\Rightarrow \quad & d(\mathscr{K}(\alpha)) = \alpha - \mathscr{C}(\mathscr{Z}(\alpha)) \\
\Rightarrow \quad & \alpha = \underbrace{d(\mathscr{K}(\alpha))}_{\text{exact by def.}} + \underbrace{\mathscr{C}(\mathscr{Z}(\alpha))}_{\text{exact from above}} \\
\Rightarrow \quad & \alpha \text{ is exact.}
\end{aligned}
$$

Thus we have shown the Poincaré lemma in our sample case. We showed that if $\alpha \in \bigwedge^1(\mathbb{R}^3)$ is closed then α is also exact.

A few comments are now in order. In the general proof of Poincaré's Lemma we will show that if α is a closed k-form on $\mathbb{R}^n$ then it is exact. Our proof requires us to assume what is called the **induction hypothesis**. In this case to show a closed k-form on $\mathbb{R}^n$ is exact we use the fact that a closed k-form on $\mathbb{R}^{n-1}$ is exact. By our base case we have already shown that any closed k-form on $\mathbb{R}^0$ is exact.

The general proof is for any n, so if we let $n = 1$ we can then show that any closed k-form on $\mathbb{R}^1$ is exact. We show this in the general proof of the Poincaré lemma, which requires us to use the induction hypothesis. In the case of $n = 1$ our induction hypothesis is that any closed k-form on $\mathbb{R}^0$ is exact, which we already know. Then, once we know this we can use it to prove that closed k-forms on $\mathbb{R}^1$ are exact. Once this is known the same argument can be used to show closed k-forms on $\mathbb{R}^2$ are exact, which can then be used to show that closed k-forms on $\mathbb{R}^3$ are exact, which can then be used to…and so on. In this way we can bootstrap ourselves up to show that closed k-forms on any $\mathbb{R}^n$ we want are exact.

Of course, we don't want to make the same argument over and over again an infinite number of times, so we make it only once, but do it in a way that is general enough that it holds for any value of k and n we want. This means that we have to be very general with our notation, which can make our argument a little notationally overwhelming. This is one of the reasons we went through a simple case first so you could get a feel for how the argument works without being too distracted by trying to follow all the notation. We will now also say that the general proof will be easier if we make a minor twist to what we did in the sample case. We will explain that when the time comes.

8.3 The General Case

In the general case in the proof of the Poincaré lemma we will show that closed k-forms on $\mathbb{R}^{n+1}$ are exact, for arbitrary k and n. But before launching into the proof of the general case let us revisit the $\mathscr{Z}$ and $\mathscr{C}$ mappings and make sure we understand what they are and how they work for any k and n. It turns out that $\mathscr{Z}$ and $\mathscr{C}$ are the pull-backs of two simple maps. We begin by examining $\mathscr{Z}$.

Consider the stretch mapping denoted by S shown in Fig. 8.10,

$$
\begin{aligned}
\mathbb{R}^n &\xrightarrow{S} \mathbb{R}^{n+1} \\
(x_1, \ldots, x_n) &\longmapsto (x_1, \ldots, x_n, 0)
\end{aligned}
$$

Notice that there is some ambiguity here. Besides having x_i being used as our notation for both a coordinate function and a numerical value, we are also using x_i as the coordinate functions for both the space $\mathbb{R}^n$ and $\mathbb{R}^{n+1}$. We want to be just a little more careful and keep our spaces distinct in our minds, so we will say that $x_1, \ldots, x_n$ are the coordinate functions on $\mathbb{R}^n$

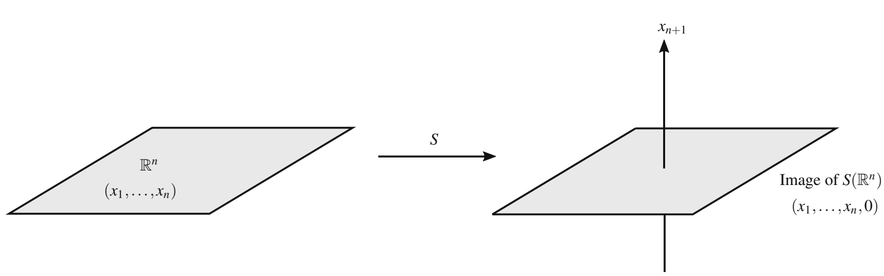


Fig. 8.10 The mapping $S : \mathbb{R}^n \to \mathbb{R}^{n+1}$

and $y_1, \ldots, y_{n+1}$ are the coordinate functions on $\mathbb{R}^{n+1}$. Being precise we would say that

$$\begin{aligned} \mathbb{R}^n &\xrightarrow{S} \mathbb{R}^{n+1} \\ (x_1, \ldots, x_n) &\longmapsto \big(y_1(x_1, \ldots, x_n), \ldots, y_{n+1}(x_1, \ldots, x_n)\big), \end{aligned}$$

where

$$\begin{aligned} y_1(x_1, \ldots, x_n) &= x_1, \\ &\vdots \\ y_n(x_1, \ldots, x_n) &= x_n, \\ y_{n+1}(x_1, \ldots, x_n) &= 0. \end{aligned}$$

Of course the mapping S gives rise to a pull-back mapping $T^*S \equiv S^*$ of k-forms,

$$\begin{aligned} \bigwedge^k(\mathbb{R}^n) &\xleftarrow{T^*S} \bigwedge^k(\mathbb{R}^{n+1}) \\ \mathbb{R}^n &\xrightarrow{S} \mathbb{R}^{n+1}. \end{aligned}$$

This is shown in Fig. 8.11.

The building blocks of k-forms on $\mathbb{R}^{n+1}$ are of the form $f_{i_1\cdots i_k} dy_{i_1} \wedge dy_{i_2} \wedge \cdots \wedge dy_{i_k}$, where by convention we have $i_1 < i_2 < \cdots < i_k \leq n+1$. Relying on the second and third identities in Sect. 6.7 and knowing what the pull-back of a zero-form is, also covered in Sect. 6.7, we compute the pull-back for one of these k-form building blocks,

$$\begin{aligned} &T^*S \cdot \big(f_{i_1\cdots 1_k} dy_{i_1} \wedge dy_{i_2} \wedge \cdots \wedge dy_{i_k}\big) \\ &= T^*S f_{i_1\cdots 1_k} \cdot T^*S dy_{i_1} \wedge T^*S dy_{i_2} \wedge \cdots \wedge T^*S dy_{i_k} \\ &= (f_{i_1\cdots 1_k} \circ S) \cdot d(S^* y_{i_1}) \wedge d(S^* y_{i_2}) \wedge \cdots \wedge d(S^* y_{i_k}) \\ &= (f_{i_1\cdots 1_k} \circ S) \cdot d(y_{i_1} \circ S) \wedge d(y_{i_2} \circ S) \wedge \cdots \wedge d(y_{i_k} \circ S) \end{aligned}$$

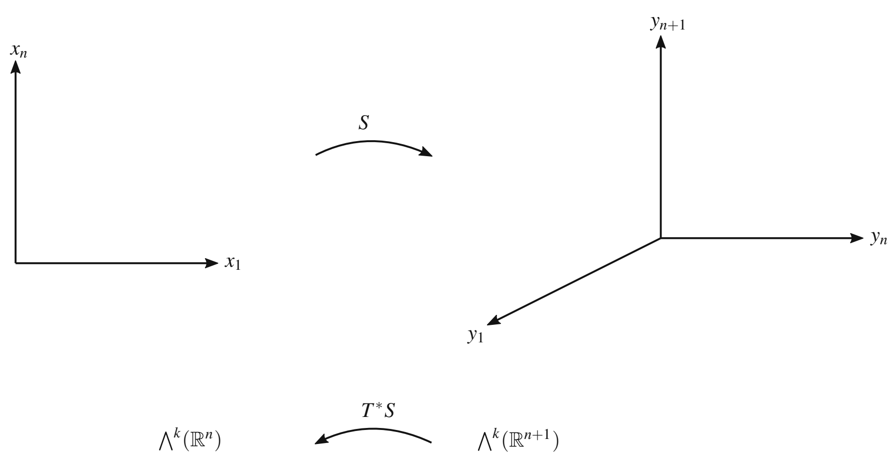


Fig. 8.11 The mapping $S : \mathbb{R}^n \to \mathbb{R}^{n+1}$ shown using the more appropriate coordinate function notation, along with the pullback mapping $T^*S \equiv S^* : \bigwedge^k(\mathbb{R}^{n+1}) \to \bigwedge^k(\mathbb{R}^n)$

$$= (f_{i_1\cdots i_k} \circ S) \cdot \left(\sum_{j=1}^{n} \frac{\partial y_{i_1}(x_1,\ldots,x_n)}{\partial x_j} dx_j\right) \wedge \cdots \wedge \left(\sum_{j=1}^{n} \frac{\partial y_{i_k}(x_1,\ldots,x_n)}{\partial x_j} dx_j\right)$$

$$= \begin{cases} f_{i_1\cdots i_k}(x_1,\ldots,x_n,0)\cdot dx_{i_1}\wedge\cdots\wedge dx_{i_k} \\ \quad 0 \qquad \text{if some } i_\ell = n+1 \end{cases}$$

The last equality follows from the definition of $y_{i_\ell}(x_1,\ldots,x_n)$. For each i_ℓ, $1 \le i_l \le n+1$, we have

$$\sum_{j=1}^{n} \frac{\partial y_{i_\ell}(x_1,\ldots,x_n)}{\partial x_j} dx_j = \begin{cases} \sum_{j=1}^{n} \frac{\partial\, x_{i_\ell}}{\partial x_j} dx_j & \text{if } i_\ell \neq n+1 \\ \sum_{j=1}^{n} \frac{\partial\, 0}{\partial x_j} dx_j & \text{if } i_\ell = n+1 \end{cases} = \begin{cases} dx_{i_\ell} & \text{if } i_\ell \neq n+1 \\ 0 & \text{if } i_\ell = n+1 \end{cases}.$$

This is exactly the map $\mathscr{Z}$ that we had before, hence $\mathscr{Z} = T^*S$. Pulling back a full k-form can be done using this formula and the linearity of the pull-back which is the first identity in Sect. 6.7. In other words, just apply this formula to each term of the k-form.

Question 8.4 For $S : \mathbb{R}^2 \to \mathbb{R}^3$ and $\alpha = P(x,y,z)dx + Q(x,y,z)dy + R(x,y,z)dz \in \bigwedge^1(\mathbb{R}^3)$ show that $T^*S \cdot \alpha$ as defined above is the same as $\mathscr{Z}(\alpha)$.

Now we turn our attention to $\mathscr{C}$. Now we consider a different map, the projection map, denoted by P and shown in Fig. 8.12,

$$\mathbb{R}^{n+1} \xrightarrow{P} \mathbb{R}^n$$
$$(y_1,\ldots,y_{n+1}) \longmapsto \big(x_1(y_1,\ldots,y_{n+1}),\ldots,x_n(y_1,\ldots,y_{n+1})\big)$$

where

$$x_1(y_1,\ldots,y_{n+1}) = y_1,$$
$$\vdots$$
$$x_n(y_1,\ldots,y_n,y_{n+1}) = y_n.$$

This gives us the pullback map $T^*P \equiv P^*$,

$$\bigwedge^k(\mathbb{R}^{n+1}) \xleftarrow{T^*P} \bigwedge^k(\mathbb{R}^n)$$
$$\mathbb{R}^{n+1} \xrightarrow{P} \mathbb{R}^n.$$

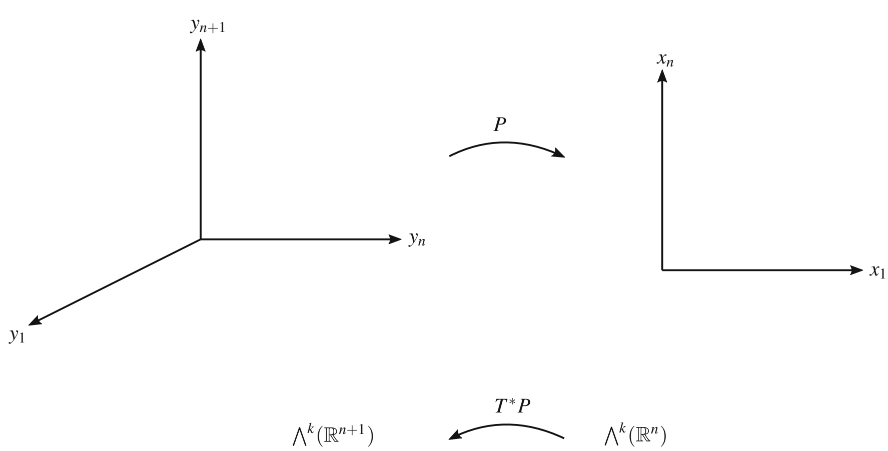


Fig. 8.12 The mapping $P : \mathbb{R}^{n+1} \to \mathbb{R}^n$ along with the pullback mapping $T^*P \equiv P^* : \bigwedge^k(\mathbb{R}^n) \to \bigwedge^k(\mathbb{R}^{n+1})$

The building blocks of k-forms on $\mathbb{R}^n$ are $f_{i_1\cdots i_k} dx_{i_1} \wedge \cdots \wedge dx_{i_k}$ where $i_1 < i_2 < \cdots < i_k$. We compute

$$\begin{aligned}
&T^*P \cdot \left(f_{i_1\cdots i_k} dx_{i_1} \wedge \cdots \wedge dx_{i_k}\right) \\
&= (T^*Pf_{i_1\cdots i_k}) \cdot T^*Pdx_{i_1} \wedge \cdots \wedge T^*Pdx_{i_k} \\
&= (f_{i_1\cdots i_k} \circ P) \cdot d(T^*Px_{i_1}) \wedge \cdots \wedge d(T^*Px_{i_k}) \\
&= (f_{i_1\cdots i_k} \circ P) \cdot d(x_{i_1} \circ P) \wedge \cdots \wedge d(x_{i_k} \circ P) \\
&= f_{i_1\cdots i_k}(y_1, \ldots, y_{n+1})\, dy_{i_1} \wedge \cdots \wedge dy_{i_k}.
\end{aligned}$$

Thus we see that our mapping is exactly $\mathscr{C} = P^*$.

Question 8.5 Show the last equality above, that is, show that $dx_{i_l} = dy_{i_l}$ for $1 \le l \le k \le n$.

Question 8.6 Show that $\mathscr{Z}\big(\mathscr{C}(\beta)\big) = \beta$ for any k-form β.

Since $\mathscr{C} = P^*$ and $\mathscr{Z} = S^*$ and since we already know from the properties of pull-backs and exterior derivatives that $dP^* = P^*d$ and $dS^* = S^*d$ we automatically have $d\mathscr{C} = \mathscr{C}d$ and $d\mathscr{Z} = \mathscr{Z}d$.

In the general case the mapping $\mathscr{K} : \bigwedge^k(\mathbb{R}^{n+1}) \to \bigwedge^{k-1}(\mathbb{R}^{n+1})$ is given by

$$\mathscr{K}\Big(f(x_1, \ldots, x_{n+1})dx_{i_1} \wedge \cdots \wedge dx_{i_k}\Big) = \begin{cases} 0 & \text{if } i_k \neq n+1, \\ \Big(\int_0^{x_{n+1}} f(x_1, \ldots, x_n, t)dt\Big)\, dx_{i_1} \wedge \cdots \wedge dx_{i_{k-1}} & \text{if } i_k = n+1. \end{cases}$$

The mapping $\mathscr{K} : \bigwedge^{k+1}(\mathbb{R}^{n+1}) \to \bigwedge^k(\mathbb{R}^{n+1})$ is similar, but adjusted accordingly.

We are now ready for the proof of the general case of the Poincaré lemma. Suppose we have $\alpha \in \bigwedge^k(\mathbb{R}^{n+1})$, which is closed. That is, $d\alpha = 0$. The mappings we will use are shown at the top of Fig. 8.13 and the general schematic of what we

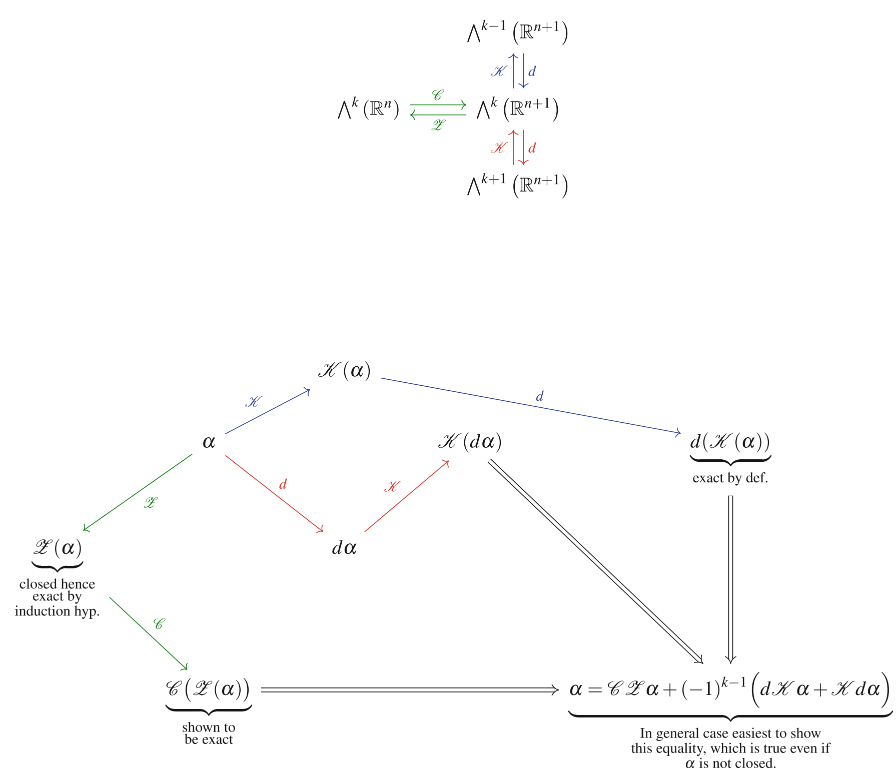


Fig. 8.13 The various mappings which are important in the proof of the general case of the Poincaré lemma are shown above along with a schematic of the general case of the proof of the Poincaré lemma, shown below. Recall that a k-form α has the form $\alpha = \sum \alpha_{i_1\cdots i_k} dx_{i_1} \wedge \cdots \wedge dx_{i_k}$, where each $\alpha_{i_1\cdots i_k}$ is a function on the manifold. While the full k-form α may be closed, each of the terms $\alpha_{i_1\cdots i_k} dx_{i_1} \wedge \cdots \wedge dx_{i_k}$ that makes up α may not be. The derived formula $\alpha = \mathscr{C}\mathscr{Z}\alpha + (-1)^{k-1}\big(d\mathscr{K}\alpha + \mathscr{K}d\alpha\big)$ applies to each of these terms. Using linearity the formula applies to all of α, and if α is closed then the formula becomes $\alpha = \mathscr{C}\mathscr{Z}\alpha + (-1)^{k-1}d\mathscr{K}\alpha$, which is shown to be exact

do is shown on the bottom. In the sample case we showed $\alpha = \mathscr{C}\mathscr{Z}\alpha + d(\mathscr{K}\alpha)$. We also showed that $\mathscr{C}\mathscr{Z}\alpha$ was exact and clearly $d(\mathscr{K}\alpha)$ is exact and so it followed that α was exact. It turns out that in the general case it is actually easier to prove something more general, that for any closed k-form α we have

$$\begin{aligned}\alpha &= \mathscr{C}\mathscr{Z}\alpha + (-1)^{k-1}\Big(d(\mathscr{K}\alpha) + \mathscr{K}d\alpha\Big)\\ &= \mathscr{C}\mathscr{Z}\alpha + (-1)^{k-1}\big(d\mathscr{K} + \mathscr{K}d\big)\alpha.\end{aligned}$$

Why would we do this? The problem is that in general the k-form α on $\mathbb{R}^{n+1}$ has the form

$$\alpha = \sum \alpha_{j_1\cdots j_k}(x_1, x_2, \ldots, x_{n+1})dx_{j_1} \wedge \cdots \wedge dx_{j_k},$$

where each $\alpha_{j_1\cdots j_k}$ is a real-valued function and of course $j_1 < \cdots < j_k$. While α overall may be closed each individual term in the sum may not be closed, and in fact probably is not closed. This is what creates the problem. For simplicity's sake we want to do the computation in our proof on just one term in the sum at a time. That is, we want to do our computation on each term $\alpha_{j_1\cdots j_k}(x_1, x_2, \ldots, x_{n+1})dx_{j_1} \wedge \cdots \wedge dx_{j_k}$ in the sum separately. Being able to do that makes the computations much simpler. We can then just use linearity to apply our result to all of α, which is just a sum of the terms.

The first part of the proof, showing $\mathscr{C}\mathscr{Z}\alpha$ is exact is the same as before. Since α is closed then by definition of closed we have $d\alpha = 0$. which gives us

$$d(\mathscr{C}\alpha) = \mathscr{C}(d\alpha) = \mathscr{C}(0) = 0,$$

so $\mathscr{C}\alpha$ is also closed. But $\mathscr{C}\alpha \in \bigwedge^k(\mathbb{R}^n)$ and our ***induction hypothesis*** *is that all closed forms on* $\mathbb{R}^n$ *are exact*. By the induction hypotheses since $\mathscr{C}\alpha$ is closed then it is also exact so $\mathscr{C}\alpha = d\beta$ for some β. Hence we have

$$\mathscr{Z}\mathscr{C}\alpha = \mathscr{Z}d\beta = d\mathscr{Z}\beta.$$

But this means that $\mathscr{Z}\mathscr{C}\alpha$ is the exterior derivative of the form $\mathscr{Z}\beta$, which by the definition of exactness means the $\mathscr{Z}\mathscr{C}\alpha$ is exact. Thus we have shown that if α is closed then $\mathscr{Z}\mathscr{C}\alpha$ is exact.

Now we turn our attention to the identity we want to show. We will need to break our general proof on a building block element of α into two cases, one where $j_k = n+1$ and one where $j_k \neq n+1$.

Case One: $j_k = n + 1$
We will consider a single term of the form $f(x_1, \ldots, x_{n+1})dx_{j_1} \wedge \cdots \wedge dx_{j_k}$, where $j_1 < j_2 < \cdots < j_k$ and $j_k = n+1$,

$$\begin{aligned}
&d\mathscr{K}\Big(f(x_1, \ldots, x_{n+1})dx_{j_1} \wedge \cdots \wedge dx_{j_k}\Big)\\
&= d\left(\left(\int_0^{x_{n+1}} f(x_1, \ldots, x_n, t)dt\right) dx_{j_1} \wedge \cdots \wedge dx_{j_{k-1}}\right)\\
&= \sum_{i=1}^{n+1} \frac{\partial}{\partial x_i}\left(\int_0^{x_{n+1}} f(x_1, \ldots, x_n, t)dt\right) dx_i \wedge dx_{j_1} \wedge \cdots \wedge dx_{j_{k-1}}\\
&= \sum_{i=1}^{n} \left(\int_0^{x_{n+1}} \frac{\partial}{\partial x_i} f(x_1, \ldots, x_n, t)dt\right) dx_i \wedge dx_{j_1} \wedge \cdots \wedge dx_{j_{k-1}}\\
&\qquad + \frac{\partial}{\partial x_{n+1}}\left(\int_0^{x_{n+1}} f(x_1, \ldots, x_n, t)dt\right) dx_{n+1} \wedge dx_{j_1} \wedge \cdots \wedge dx_{j_{k-1}}\\
&= \sum_{i=1}^{n} \left(\int_0^{x_{n+1}} \frac{\partial}{\partial x_i} f(x_1, \ldots, x_n, t)dt\right) dx_i \wedge dx_{j_1} \wedge \cdots \wedge dx_{j_{k-1}}\\
&\qquad + f(x_1, \ldots, x_{n_1})dx_{n+1} \wedge dx_{j_1} \wedge \cdots \wedge dx_{j_{k-1}}.
\end{aligned}$$

Next we compute

$$\begin{aligned}
&\mathscr{K} d\Big(f(x_1,\dots,x_{n+1})dx_{j_1}\wedge\cdots\wedge dx_{j_k}\Big)\\
&= \mathscr{K}\left(\sum_{i=1}^{n+1}\frac{\partial f(x_1,\dots,x_{n+1})}{\partial x_i}\, dx_i\wedge dx_{j_1}\wedge\cdots\wedge dx_{j_k}\right)\\
&= \mathscr{K}\left(\sum_{i=1}^{n}\frac{\partial f(x_1,\dots,x_{n+1})}{\partial x_i}\, dx_i\wedge dx_{j_1}\wedge\cdots\wedge dx_{j_k}\right)\\
&= \sum_{i=1}^{n}\ \mathscr{K}\left(\frac{\partial f(x_1,\dots,x_{n+1})}{\partial x_i}\, dx_i\wedge dx_{j_1}\wedge\cdots\wedge dx_{j_k}\right)\\
&= \sum_{i=1}^{n}\left(\int_0^{x_{n+1}}\frac{\partial f(x_1,\dots,x_n,t)}{\partial x_i}\,dt\right)\, dx_i\wedge dx_{j_1}\wedge\cdots\wedge dx_{j_k}.
\end{aligned}$$

Subtracting the second from the first, recalling that $j_k = n+1$, and then noting that

$$\mathscr{C}\mathscr{Z}\big(f(x_1,\dots,x_{n_1})dx_{j_1}\wedge\cdots\wedge dx_{j_{k-1}}\wedge dx_{n+1}\big)=0,$$

we get

$$\begin{aligned}
&(d\mathscr{K}-\mathscr{K}d)\big(f(x_1,\dots,x_{n+1})dx_{j_1}\wedge\cdots\wedge dx_{j_k}\big)\\
&= f(x_1,\dots,x_{n_1})dx_{n+1}\wedge dx_{j_1}\wedge\cdots\wedge dx_{j_{k-1}}\\
&= (-1)^{k-1} f(x_1,\dots,x_{n_1})dx_{j_1}\wedge\cdots\wedge dx_{j_{k-1}}\wedge dx_{n+1} - 0\\
&= (-1)^{k-1}\big(1-\mathscr{C}\mathscr{Z}\big)\big(f(x_1,\dots,x_{n_1})dx_{j_1}\wedge\cdots\wedge dx_{j_{k-1}}\wedge dx_{n+1}\big).
\end{aligned}$$

A bit of rearrangement, and relabeling $\omega = f(x_1,\dots,x_{n+1})dx_{j_1}\wedge\cdots\wedge dx_{j_k}$, which may not actually itself be closed, gives us what we wanted, namely

$$\omega = \mathscr{C}\mathscr{Z}\omega + (-1)^{k-1}\big(d\mathscr{K}\omega - \mathscr{K}d\omega\big).$$

Question 8.7 Show that $\mathscr{C}\mathscr{Z}\big(f(x_1,\dots,x_{n_1})dx_{j_1}\wedge\cdots\wedge dx_{j_{k-1}}\wedge dx_{n+1}\big)=0$.

Case Two: $j_k \neq n+1$

We will consider a single term of the form $f(x_1,\dots,x_{n+1})dx_{j_1}\wedge\cdots\wedge dx_{j_k}$, where $j_1 < j_2 < \cdots < j_k$ but where $j_k \neq n+1$. Here, by definition of the mapping $\mathscr{K}$, we have

$$d\mathscr{K}\big(f(x_1,\dots,x_{n+1})dx_{j_1}\wedge\cdots\wedge dx_{j_k}\big)=0.$$

Now we compute

$$\begin{aligned}
&\mathscr{K} d\Big(f(x_1,\dots,x_{n+1})dx_{j_1}\wedge\cdots\wedge dx_{j_k}\Big)\\
&= \mathscr{K}\left(\sum_{i=1}^{n+1}\frac{\partial f(x_1,\dots,x_{n+1})}{\partial x_i}\, dx_i\wedge dx_{j_1}\wedge\cdots\wedge dx_{j_k}\right)\\
&= \mathscr{K}\left(\underbrace{\sum_{i=1}^{n}\frac{\partial f(x_1,\dots,x_{n+1})}{\partial x_i}\, dx_i\wedge dx_{j_1}\wedge\cdots\wedge dx_{j_k}}_{\mathscr{K}\text{ kills all these terms since they not have } dx_{n+1}} + \frac{\partial f(x_1,\dots,x_{n+1})}{\partial x_{n+1}}\, dx_{n+1}\wedge dx_{j_1}\wedge\cdots\wedge dx_{j_k}\right)
\end{aligned}$$

$$
\begin{aligned}
&= \mathscr{K}\left((-1)^k \frac{\partial f(x_1, \dots, x_{n+1})}{\partial x_{n+1}} \, dx_{j_1} \wedge \cdots \wedge dx_{j_k} \wedge dx_{n+1} \right) \\
&= (-1)^k \left(\int_0^{x_{n+1}} \frac{\partial f(x_1, \dots, x_n, t)}{\partial x_{n+1}} \, dt \right) dx_{j_1} \wedge \cdots \wedge dx_{j_k} \\
&= (-1)^k \Big(f(x_1, \dots, x_n, x_{n+1}) - f(x_1, \dots, x_n, 0) \Big) dx_{j_1} \wedge \cdots \wedge dx_{j_k}. \qquad \text{fundamental theorem of calculus.}
\end{aligned}
$$

Subtracting the second equation from the first give us

$$
\begin{aligned}
&(d\mathscr{K} - \mathscr{K}d)\big(f(x_1, \dots, x_{n+1}) dx_{j_1} \wedge \cdots \wedge dx_{j_k}\big) \\
&= 0 - (-1)^k \Big(f(x_1, \dots, x_n, x_{n+1}) - f(x_1, \dots, x_n, 0) \Big) dx_{j_1} \wedge \cdots \wedge dx_{j_k} \\
&= (-1)^{k-1} \Big(f(x_1, \dots, x_n, x_{n+1}) dx_{j_1} \wedge \cdots \wedge dx_{j_k} - f(x_1, \dots, x_n, 0) dx_{j_1} \wedge \cdots \wedge dx_{j_k} \Big).
\end{aligned}
$$

But if $j_k \neq n+1$ then

$$
\mathscr{C}\mathscr{Z}\big(f(x_1, \dots, x_{n+1}) dx_{j_1} \wedge \cdots \wedge dx_{j_k}\big) = f(x_1, \dots, x_n, 0) dx_{j_1} \wedge \cdots \wedge dx_{j_k}
$$

so letting $\omega = f(x_1, \dots, x_{n+1}) dx_{j_1} \wedge \cdots \wedge dx_{j_k}$ we have

$$
\begin{aligned}
(d\mathscr{K} - \mathscr{K}d)\omega &= (-1)^{k-1}\big(\omega - \mathscr{C}\mathscr{Z}\omega\big) \\
\Rightarrow \omega &= \mathscr{C}\mathscr{Z}\omega + (-1)^{k-1}\big(d\mathscr{K}\omega - \mathscr{K}d\omega\big),
\end{aligned}
$$

which is exactly the formula we had in case one.

Thus, regardless of whether $j_k = n+1$ or if $j_k \neq n+1$ we end up with exactly the same identity for each term ω of the k-form α, namely that

$$
\mathscr{C}\mathscr{Z}\omega + (-1)^{k-1}\big(d\mathscr{K}\omega - \mathscr{K}d\omega\big).
$$

Thus this formula holds for each individual term in the k-form α. Now we are ready to complete the proof of the Poincaré lemma. We will leave a couple of the steps to you in the following questions.

Question 8.8 Show that the mappings $\mathscr{Z}$, $\mathscr{C}$, and $\mathscr{K}$ are linear. That is, show that if α and β are k-forms that $\mathscr{Z}(\alpha + \beta) = \mathscr{Z}(\alpha) + \mathscr{Z}(\beta)$, $\mathscr{C}(\alpha + \beta) = \mathscr{C}(\alpha) + \mathscr{C}(\beta)$, and $\mathscr{K}(\alpha + \beta) = \mathscr{K}(\alpha) + \mathscr{K}(\beta)$.

Question 8.9 Using linearity of $\mathscr{C}$, $\mathscr{Z}$, $\mathscr{K}$, and d show that the identity

$$
\omega = \mathscr{C}\mathscr{Z}\omega + (-1)^{k-1}\big(d\mathscr{K}\omega - \mathscr{K}d\omega\big)
$$

applies to a general k-form

$$
\alpha = \sum_i^{n+1} \alpha_{j_1 \cdots j_k} dx_{j_1} \wedge \cdots \wedge dx_{j_k}.
$$

From the above questions, for a general k-form α we have that $\alpha = \mathscr{C}\mathscr{Z}\omega + (-1)^{k-1}\big(d\mathscr{K}\alpha - \mathscr{K}d\alpha\big)$. To show that the closed k-forms on $\mathbb{R}^{n+1}$ are exact we are assuming by the ***induction hypothesis*** *that the closed k-forms on $\mathbb{R}^n$ are exact*. The induction hypothesis was used to show that if α is closed then $\mathscr{C}\mathscr{Z}\alpha$ is exact, so there exists a β such that $\mathscr{C}\mathscr{Z}\alpha = d\beta$. Also, since α is closed we have $\mathscr{K}d\alpha = 0$, so

$$
\begin{aligned}
\alpha &= \mathscr{C}\mathscr{Z}\alpha + (-1)^{k-1}\big(d\mathscr{K}\alpha - \mathscr{K}\underbrace{d\alpha}_{=0}\big) \\
&= d\beta + (-1)^{k-1}\big(d\mathscr{K}\alpha - 0\big)
\end{aligned}
$$

$$= d\beta + d\big((-1)^{k-1}\mathscr{K}\alpha\big)$$
$$= d\Big(\beta + (-1)^{k-1}\mathscr{K}\alpha\Big)$$

so α is exact.

Question 8.10 Explain in more detail, using the diagram in Figs. 8.5 and 8.13, how the induction hypothesis, along with the formula that was derived, allow us to conclude that every closed k-form on any $\mathbb{R}^n$, is exact.

8.4 Summary, References, and Problems

8.4.1 Summary

The Poincaré lemma states that *every closed form on* $\mathbb{R}^n$ *is exact*. A differential form α is called closed if $d\alpha = 0$. A differential form α is called exact if there is another differential form β such that $\alpha = d\beta$. Obviously, if α is an exact k-form then β must be a $(k-1)$-form. So, another way of phrasing the Poincaré lemma is to say that *if α is a k-form on $\mathbb{R}^n$ such that $d\alpha = 0$, then there exists some $(k-1)$-form β such that $\alpha = d\beta$.*

8.4.2 References and Further Reading

The Poincaré lemma is an absolutely essential material for any book that looks at calculus on manifolds. In particular, Munkres [35], Renteln [37], and Abraham, Marsden, and Ratiu [1] are nice presentations. As is often the case in mathematics, with the right mathematical concepts, machinery, and sophistication the proofs of many theorems can be reduced to a mere handful of lines. This is very much the case with the Poincaré lemma. The presentation here, while long, is meant to be readily understood by someone with little more mathematical background than calculus. Generally that is not the case with most other presentations of the Poincaré lemma.

8.4.3 Problems

Question 8.11 Show that each of the below one-forms α_i is closed; that is, show that $d\alpha_i = 0$. By the Poincaré lemma α_i must then be exact. Then find a function f_i such that $\alpha_i = df_i$.

a) $\alpha_1 = y^2\,dx + 2xy\,dy$
b) $\alpha_2 = (3x^2 + y)\,dx + (x + 2y)\,dy$
c) $\alpha_3 = 2x\,dx + 3y^2\,dy + 4z^3\,dz$
d) $\alpha_4 = 2xy\,dx + (x^2 + 2y + z)\,dy + (y - 3z^2)\,dz$
e) $\alpha_5 = (6x^2z - yz)\,dx + (3z^4 - xz)\,dy + (2x^3 + 12yz^3 - xy)\,dz$
f) $\alpha_6 = (3y^2 - 4z^4)\,dx + 6xy\,dy - 16xz^3\,dz$

Question 8.12 Show that each of the below two-forms β_i is closed; that is, show that $d\beta_i = 0$. By the Poincaré lemma β_i must then be exact. Find a one-form α_i such that $\beta_i = d\alpha_i$.

a) $\beta_1 = 24x^3y^2\,dx \wedge dy$
b) $\beta_2 = (6x^2y - 3xy^2)\,dx \wedge dy$
c) $\beta_3 = 8x^3y^3\,dy \wedge dz - 6x^2y^4\,dz \wedge dx$
d) $\beta_4 = (2x - 1)\,dx \wedge dy + (3y^2 - 2z)\,dy \wedge dz + (1 - 3x^2)\,dz \wedge dx$
e) $\beta_5 = -4xy^2\,dx \wedge dy + 2x^2y^2\,dy \wedge dz + 3xz^2\,dz \wedge dx$
f) $\beta_6 = -4xy\,dx \wedge dy + 2yz\,dy \wedge dz - 2xz\,dz \wedge dx$

Question 8.13 Show that each of the below three-forms γ_i is closed; that is, show that $d\gamma_i = 0$. By the Poincaré lemma γ_i must then be exact. Find a two-form β_i such that $\gamma_i = d\beta_i$.

a) $\gamma_1 = 12x^2y^3z^4\,dx \wedge dy \wedge dz$
b) $\gamma_2 = (x^3 - 4y + z^2)\,dx \wedge dy \wedge dz$
c) $\gamma_3 = (2xz + 3yz^2)\,dx \wedge dy \wedge dz$
d) $\gamma_4 = 2(x + y + z)\,dx \wedge dy \wedge dz$

Chapter 9
Vector Calculus and Differential Forms

In sections one through three we take a careful look at divergence, curl, and gradient from vector calculus, introducing and defining all three from a geometrical point of view. In section four we introduce some notation and consider two important operators, the sharp and flat operators. These operators are necessary to understand the relationship between divergence, curl, gradient and differential forms, which is looked at in section five. Also in section five we see how the fundamental theorem of line integrals, the divergence theorem, and the vector calculus version of Stokes' theorem can all be written in the same way using differential forms notation. These three theorems are all special cases of what is called the generalized Stokes' theorem. In more advanced mathematics classes Stokes' theorem always refers to the generalized Stokes' theorem and not to the version of Stokes' theorem you learned in vector calculus, while in physics Stokes' theorem may refer to either version depending on context. But keep in mind, they are actually the same theorem from a more abstract perspective.

9.1 Divergence

In vector calculus a vector field $\mathbf{F}$ on $\mathbb{R}^3$ was generally written as $\mathbf{F} = P(x, y, z)\hat{i} + Q(x, y, z)\hat{j} + R(x, y, z)\hat{k}$, where $P, Q, R : \mathbb{R}^3 \to \mathbb{R}$. Here $\hat{i}$, $\hat{j}$, and $\hat{k}$ are the unit vectors in the x, y, and z directions respectively. If we were to denote our variables as x_1, x_2, x_3 then the unit vectors are generally written as e_1, e_2, e_2. Recall that we always write a vector as a column vector,

$$\mathbf{F} = \begin{bmatrix} P \\ Q \\ R \end{bmatrix}.$$

In vector calculus classes the divergence of a vector field $\mathbf{F}$ is usually defined to be

$$\operatorname{div} \mathbf{F} = \frac{\partial P}{\partial x} + \frac{\partial Q}{\partial y} + \frac{\partial R}{\partial z}.$$

Often an operator ∇ is defined as

$$\nabla = \frac{\partial}{\partial x}e_1 + \frac{\partial}{\partial y}e_2 + \frac{\partial}{\partial z}e_3.$$

If we do this then we can consider the "dot product" of the operator ∇ with a vector $\mathbf{F}$ as

$$\operatorname{div} \mathbf{F} = \nabla \cdot \mathbf{F} = \left(\frac{\partial}{\partial x}e_1 + \frac{\partial}{\partial y}e_2 + \frac{\partial}{\partial z}e_3\right) \cdot (Pe_1 + Qe_2 + Re_3) = \frac{\partial P}{\partial x} + \frac{\partial Q}{\partial y} + \frac{\partial R}{\partial z}.$$

This does seem a little silly since dot products require two vectors and clearly ∇ is not really a vector, it is something else. It is generally called an operator and is said to "operate" on the vector $\mathbf{F}$. Similarly, one can think of the ∇ operator as being a

J. P. Fortney, *A Visual Introduction to Differential Forms and Calculus on Manifolds*,
https://doi.org/10.1007/978-3-319-96992-3_9

row vector which is matrix multiplied by the column vector **F**,

$$\nabla \cdot \mathbf{F} = \left[\frac{\partial}{\partial x}, \frac{\partial}{\partial y}, \frac{\partial}{\partial z}\right]\begin{bmatrix} P \\ Q \\ R \end{bmatrix}.$$

This so-called definition of div **F** as $\nabla \cdot \mathbf{F}$ is really more of a mnemonic device to help us remember the form that div **F** takes when using Cartesian coordinates.

A quick word about terminology. The symbol ∇ is actually called *nabla* after the Hebrew word for harp, a stringed musical instrument that has roughly a ∇ shape. When ∇ is used to represent the divergence operator then it is generally called *del*. One would read $\nabla \cdot \mathbf{F}$ as "del dot F," or as "div F," or as "divergence of F."

We will take a different approach here and define div **F** in a way that will allow the divergence theorem to just fall out of the definition. Using this we will derive the standard formula for div **F** in Cartesian coordinates. We take this approach because we believe it provides a somewhat more geometrical meaning behind the divergence. But before we do that we first need to understand what the surface integral of the normal component of a vector field is. Since we actually assume you have already taken a vector calculus course we will cover this material quickly and without too much detail; we are more interested in conceptual understanding than in mathematical rigor.

Suppose we are given a vector field $\mathbf{F} = P\hat{i} + Q\hat{j} + R\hat{k}$ defined on $\mathbb{R}^3$ along with a two dimensional surface S in $\mathbb{R}^3$ as in Fig. 9.1. We want to define the integral of **F** over the surface S. In a sense we are asking how much of the vector field **F** goes through the surface S. This makes more intuitive sense if **F** represents the velocity of fluid particles. Then we are basically asking the rate of fluid flow through the surface S. See Fig. 9.2.

Suppose for some fluid with density $\rho(x, y, z)$ we have that $\mathbf{v}(x, y, z)$ is the velocity at which the particle of fluid at point (x, y, z) is flowing. We want to find how much fluid flows through some small surface ΔS. If we suppose that $\mathbf{v}$ is perpendicular to ΔS. Then we have that over a small time period Δt the mass of the fluid that flows through ΔS is given by $\rho(x, y, z)\ \mathbf{v}(x, y, z)\ \Delta t\ \Delta S$, as is seen in Fig. 9.2. The rate of flow is given by the mass of the fluid that passes through the surface divided by the time over which this mass passed through the surface,

$$\begin{array}{c}\text{rate of}\\ \text{fluid flow}\\ \text{through}\\ \text{surface } \Delta S\end{array} = \frac{\rho\ \mathbf{v}\ \Delta t\ \Delta S}{\Delta t} = \rho\ \mathbf{v}\ \Delta S.$$

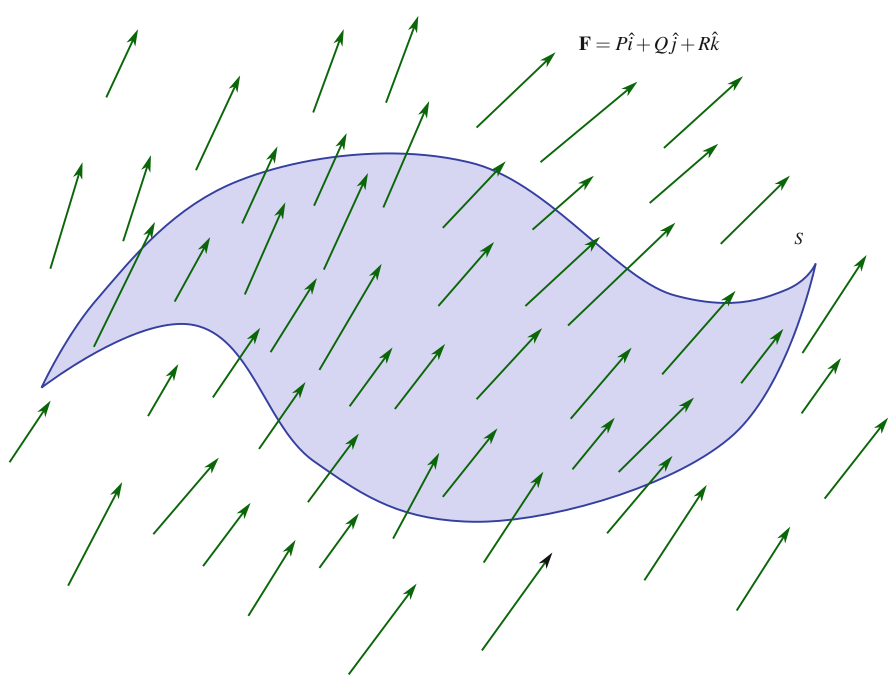
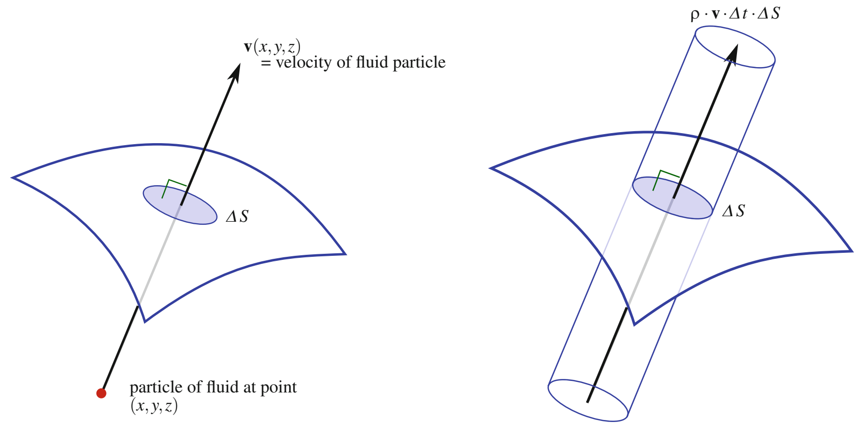


Fig. 9.1 A vector field **F** in $\mathbb{R}^3$ along with a two dimensional surface S



A vector $\mathbf{v}$ with base point (x,y,z) that "penetrates" the surface S. It is helpful to imagine a particle at point (x,y,z) traveling with velocity $\mathbf{v}$.

Estimating the mass of fluid flow through a small bit of surface, ΔS, over a small time period Δt assuming all particles are traveling with velocity $\mathbf{v}$ and the particles have density ρ.

Fig. 9.2 A natural way to think about integrating a vector field $\mathbf{v}$ over a surface S is to think of it as finding the rate of flow of a fluid over that surface traveling with velocity $\mathbf{v}$

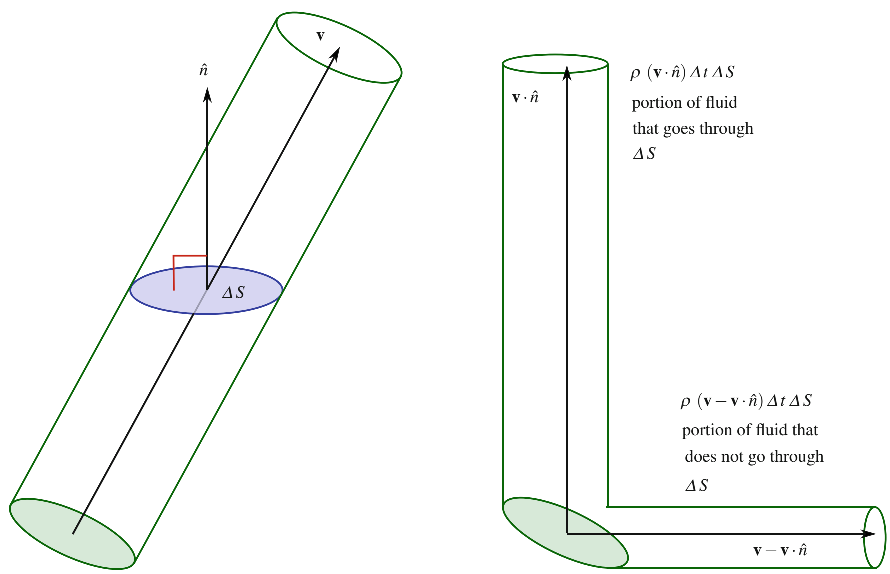


Fig. 9.3 Suppose $\mathbf{v}$ is not perpendicular to the surface S (left). Then we "decompose" the fluid flow into into two parts, a portion that goes through the surface S and a portion that does not go through S (right). This necessitates taking the dot product of $\mathbf{v}$ with the normal unit vector to S, denoted by $\hat{n}$

If $\mathbf{v}$ is not perpendicular to the surface ΔS then the flow through ΔS is given by $\rho(x,y,z)\,\mathbf{v}(x,y,z)\cdot\hat{n}\ \Delta t\ \Delta S$. That is, we replace $\mathbf{v}$ in the product with $\mathbf{v}\cdot\hat{n}$, the dot product with the unit vector normal to the surface element ΔS. This dot product gives us the component of $\mathbf{v}$ in the $\hat{n}$ direction. See Fig. 9.3. In other words, it gives us the velocity of the fluid flow in the $\hat{n}$ direction,

$$\begin{array}{c}\text{rate of}\\ \text{fluid flow}\\ \text{through}\\ \text{surface } \Delta S\end{array} = \frac{\rho\ (\mathbf{v}\cdot\hat{n})\ \Delta t\ \Delta S}{\Delta t} = \rho\ (\mathbf{v}\cdot\hat{n})\ \Delta S.$$

We can approximate the fluid flow through the whole surface S by breaking S into small pieces ΔS, finding the flow through each piece and then summing. Then to find the exact flow we can take the limit as the size of the ΔS shrinks to zero,

$|\Delta S| \to 0$,

$$\begin{array}{c}\text{rate of}\\ \text{fluid flow}\\ \text{through}\\ \text{surface } S\end{array} = \lim_{|\Delta S_i| \to 0} \sum_i \rho\, (\mathbf{v} \cdot \hat{n})\, \Delta S \;=\; \int_S \rho\, (\mathbf{v} \cdot \hat{n})\, dS.$$

We have been dealing with a concrete situation, a fluid flow, to help us understand what is going on better. But the same reasoning applies to any vector field $\mathbf{F}$; we can find the "flow" of the vector field $\mathbf{F}$ through the surface S. But when we take away our mental crutch of the fluid and deal just with a vector field there is actually no fluid density ρ, so we must drop the ρ from the above equations. But in this more pure situation where we just have a vector field and not a fluid then we are not actually finding the rate of a fluid flow through S. This admittedly abstract thing we are finding is called the *flux of the vector field* $\mathbf{F}$ *through the surface* S. Flux is the Latin word for flow. What flux actually represents in a physical situation depends upon the particular problem and on how the vector field is being interpreted. Is the vector a fluid flow? An electric field? A magnetic field? A force? In each case the flux through a surface would be interpreted differently. So the flux is actually an abstract mathematical concept that can be interpreted differently in different physical situations. Later on we will see examples where the field $\mathbf{F}$ represents an electric field or a magnetic field and we will find the electric and the magnetic flux through a surface. However, in general probably the best way to think about what a flux actually is is to imagine a fluid without a density. Thus we would have

$$\text{Flux of } \mathbf{F} = \lim_{|\Delta S_i| \to 0} \sum_i \mathbf{F} \cdot \hat{n}\, \Delta S \;=\; \int_S \mathbf{F} \cdot \hat{n}\, dS.$$

Note that this integral makes sense, at each point p on the surface S we have that $\mathbf{F}_p \cdot \hat{n}_p$ is a real number. That is, we can think of $\mathbf{F} \cdot \hat{n}$ as a real-valued function on S, $\mathbf{F} \cdot \hat{n} : S \to \mathbb{R}$, which means we can easily integrate it over the surface S. In vector calculus you learned various techniques for doing just this. In summary, the flux of the vector field $\mathbf{F}$ through the surface S is given by the integral

$$\text{Flux of } \mathbf{F} \;=\; \int_S \mathbf{F} \cdot \hat{n}\, dS.$$

A couple notes on notation, sometimes this integral is written as

$$\int_S \mathbf{F} \cdot dS \qquad \text{or} \qquad \int_S \mathbf{F} \cdot d\mathbf{S}.$$

So when you see either $\int_S \mathbf{F} \cdot dS$ or $\int_S \mathbf{F} \cdot d\mathbf{S}$ this means exactly $\int_S \mathbf{F} \cdot \hat{n}\, dS$. This notation is explained in more depth in Sect. 9.5.3. Also, since S is a surface and not function then clearly dS does not represent the exterior derivative of a zero-from. But dS comes from the ΔS, which represents the area of a small bit of surface S, so roughly you can think of dS as being an "area form" on the surface S, which allows you to find the surface area of the surface S. We will not delve into the details here.

Now we will define the divergence of $\mathbf{F}$ at a point (x_0, y_0, z_0). Suppose we have a small three dimensional region V about the point (x_0, y_0, z_0). The boundary of this region V is denoted by ∂V. By closed we mean that the surface S is like a sphere with no edges. We will also denote the volume of the region V by ΔV. Then we define the **divergence** of $\mathbf{F}$ to be given by

$$\boxed{\begin{array}{c}\text{Definition}\\ \text{of}\\ \text{divergence}\end{array} \quad \operatorname{div} \mathbf{F} \;=\; \lim_{\Delta V \to 0} \frac{1}{\Delta V} \int_{\partial V} \mathbf{F} \cdot \hat{n}\, dS.}$$

Understanding what this actually represents may be a little easier if we write the right hand side as

$$\lim_{\Delta V \to 0} \frac{\int_{\partial V} \mathbf{F} \cdot \hat{n}\, dS}{\Delta V}$$

and consider what this means. In the numerator is the flux through ∂V, that is, the flow of the vector field through the surface ∂V. Since ∂V is a closed surface, if we choose the normal vector $\hat{n}$ to point outwards, the flux of $\mathbf{F}$ through ∂V is the net flux of $\mathbf{F}$ out of V minus the net flux of $\mathbf{F}$ into V. In other words, the net flux of $\mathbf{F}$ out of V through ∂V, that is, $\int_{\partial V} \mathbf{F} \cdot \hat{n}\, dS$, can either be positive, zero, or negative. By dividing net flux of $\mathbf{F}$ out of V by ΔV we are finding the net flux of $\mathbf{F}$ out of V per unit volume. When we let the limit as the volume ΔV around the point (x_0, y_0, z_0) go to zero then we are finding the net

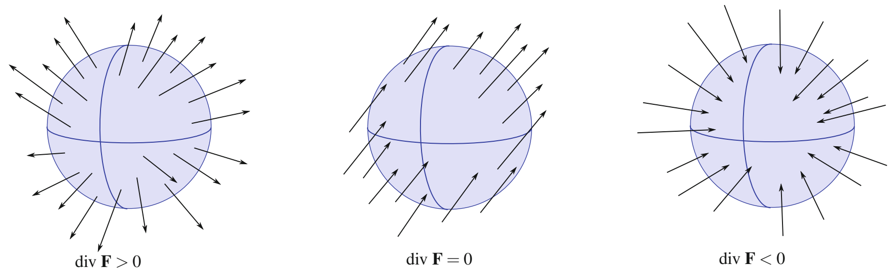


Fig. 9.4 Three examples of the flux of a vector field out of a small volume ΔV about a point (x_0, y_0, z_0). The cases where the flux is positive (left), zero (middle), and negative (right) are shown

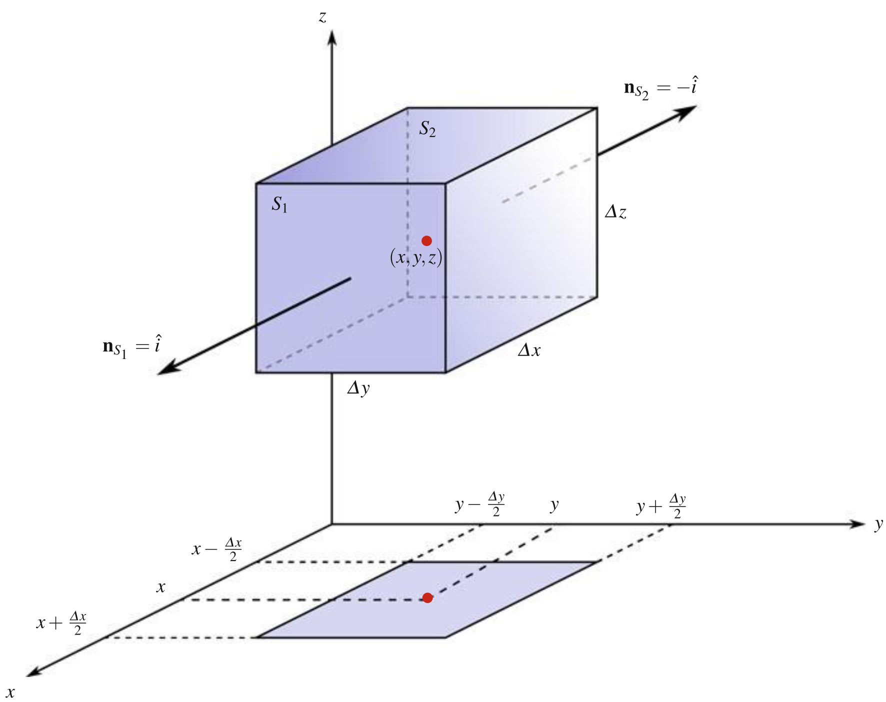


Fig. 9.5 The region V as a cube with sides Δx, Δy, and Δz. Thus we have $\Delta V = \Delta x \; \Delta y \; \Delta z$. The sides S_1 and S_2 (the front and the back of the cube) are perpendicular to the x-axis. The outward pointing unit normal to S_1 is $\hat{i}$ and the outward pointing unit normal to S_2 is $-\hat{i}$

flux out per unit volume at the point (x_0, y_0, z_0). In other words, this is how much the vector field **F** "diverges" at the point (x_0, y_0, z_0).

If div **F** is positive at the point (x_0, y_0, z_0) then that means that the vector field **F** is dissipating from the point (x_0, y_0, z_0). If div **F** is negative at the point (x_0, y_0, z_0) then that means that the vector field **F** is accumulating at the point (x_0, y_0, z_0). If div **F** is zero at the point (x_0, y_0, z_0) then **F** is neither accumulating nor dissipating at the point. See Fig. 9.4.

Now we want to use our definition of the divergence of **F** as $\lim_{\Delta V \to 0} \frac{1}{\Delta V} \int_S \mathbf{F} \cdot \hat{n} \; dS$ to get the standard formula $\frac{\partial P}{\partial x} + \frac{\partial Q}{\partial y} + \frac{\partial R}{\partial z}$ for divergence that you learned in vector calculus. Using the cubical region V that is shown in Fig. 9.5 we have surfaces S_1 and S_2 which are parallel to the y-axis and z-axis and perpendicular to the x-axis. We think of S_1 as the front of the cube and S_2 as the back of the cube. The outward normal for S_1 is $\hat{n}_{S_1} = \hat{i}$ and the outward normal for S_2 is $\hat{n}_{S_2} = -\hat{i}$. We have

$$\begin{aligned} \int_{S_1} \mathbf{F} \cdot \hat{n}_{S_1} \; dS &\approx \mathbf{F}\left(x_0 + \frac{\Delta x}{2}, y_0, z_0\right) \cdot \hat{i} \; \Delta y \; \Delta z \\ &= P\left(x_0 + \frac{\Delta x}{2}, y_0, z_0\right) \; \Delta y \; \Delta z \end{aligned}$$

and

$$
\begin{aligned}
\int_{S_1} \mathbf{F} \cdot \hat{n}_{S_2} \, dS &\approx \mathbf{F}\left(x_0 - \frac{\Delta x}{2}, y_0, z_0\right) \cdot (-\hat{i}) \; \Delta y \; \Delta z \\
&= -P\left(x_0 - \frac{\Delta x}{2}, y_0, z_0\right) \; \Delta y \; \Delta z.
\end{aligned}
$$

Summing we have

$$
\begin{aligned}
\int_{S_1+S_2} \mathbf{F} \cdot \hat{n} \, dS &\approx \left(P\left(x_0 + \frac{\Delta x}{2}, y_0, z_0\right) - P\left(x_0 - \frac{\Delta x}{2}, y_0, z_0\right)\right) \; \Delta y \; \Delta z \\
&= \frac{P\left(x_0 + \frac{\Delta x}{2}, y_0, z_0\right) - P\left(x_0 - \frac{\Delta x}{2}, y_0, z_0\right)}{\Delta x} \; \Delta x \; \Delta y \; \Delta z
\end{aligned}
$$

$$
\Longrightarrow \frac{1}{\Delta x \; \Delta y \; \Delta z} \int_{S_1+S_2} \mathbf{F} \cdot \hat{n} \, dS \approx \frac{P\left(x_0 + \frac{\Delta x}{2}, y_0, z_0\right) - P\left(x_0 - \frac{\Delta x}{2}, y_0, z_0\right)}{\Delta x}.
$$

Recalling that $\Delta V = \Delta x \; \Delta y \; \Delta z$ and then taking the limit as $\Delta V \to 0$ (that is, as $\Delta x \to 0$, $\Delta y \to 0$, and $\Delta z \to 0$) we get

$$
\begin{aligned}
\lim_{\Delta V \to 0} \frac{1}{\Delta V} \int_{S_1+S_2} \mathbf{F} \cdot \hat{n} \, dS &= \lim_{\Delta V \to 0} \frac{P\left(x_0 + \frac{\Delta x}{2}, y_0, z_0\right) - P\left(x_0 - \frac{\Delta x}{2}, y_0, z_0\right)}{\Delta x} \\
&= \lim_{\Delta x \to 0} \frac{P\left(x_0 + \frac{\Delta x}{2}, y_0, z_0\right) - P\left(x_0 - \frac{\Delta x}{2}, y_0, z_0\right)}{\Delta x} \\
&= \left.\frac{\partial P}{\partial x}\right|_{(x_0, y_0, z_0)}.
\end{aligned}
$$

In the limit the approximation becomes an equality. Also, since there is no Δy or Δz on the right hand side then $\Delta V \to 0$ becomes simply $\Delta x \to 0$.

Question 9.1 Let S_3 and S_4 be the sides of the region V that are perpendicular to the y-axis, that is, the right and left sides of the cube. Let S_5 and S_6 be the sides that are perpendicular to the z-axis, that is, the top and bottom of the cube. Repeat these calculations for surfaces $S_3 + S_4$ and $S_5 + S_6$ to get

$$
\begin{aligned}
\lim_{\Delta V \to 0} \frac{1}{\Delta V} \int_{S_3+S_4} \mathbf{F} \cdot \hat{n} \, dS &= \left.\frac{\partial Q}{\partial y}\right|_{(x_0, y_0, z_0)}, \\
\lim_{\Delta V \to 0} \frac{1}{\Delta V} \int_{S_5+S_6} \mathbf{F} \cdot \hat{n} \, dS &= \left.\frac{\partial R}{\partial z}\right|_{(x_0, y_0, z_0)}.
\end{aligned}
$$

Summing all these terms together we have that

$$
\begin{aligned}
\operatorname{div} \mathbf{F} &= \lim_{\Delta V \to 0} \frac{1}{\Delta V} \int_{\partial V} \mathbf{F} \cdot \hat{n} \, dS \\
&= \lim_{\Delta V \to 0} \frac{1}{\Delta V} \int_{S_1+S_2+S_3+S_4+S_5+S_6} \mathbf{F} \cdot \hat{n} \, dS \\
&= \left.\frac{\partial P}{\partial x}\right|_{(x_0, y_0, z_0)} + \left.\frac{\partial Q}{\partial y}\right|_{(x_0, y_0, z_0)} + \left.\frac{\partial R}{\partial z}\right|_{(x_0, y_0, z_0)} \\
&= \frac{\partial P}{\partial x} + \frac{\partial Q}{\partial y} + \frac{\partial R}{\partial z}.
\end{aligned}
$$

In other words, we have just found the formula

$$\text{div}\,\mathbf{F} = \frac{\partial P}{\partial x} + \frac{\partial Q}{\partial y} + \frac{\partial R}{\partial z}.$$

Formula for divergence **F** in Cartesian coordinates

Since we can take the divergence at any point, the point is usually left off of the notation. We use $\nabla \cdot \mathbf{F}$ as a mnemonic device to help us remember this formula.

Now we turn to looking at the divergence theorem (sometimes called Gauss's theorem) from vector calculus. We will "derive" (in a very non-rigorous fashion) the divergence theorem. This is relatively straight-forward given how we defined div **F**. We will make use of the fact that the flux through a surface S is the sum of the fluxes through the subsurfaces S_i of S,

$$\int_S \mathbf{F} \cdot \hat{n}\, dS = \sum_i \int_{S_i} \mathbf{F} \cdot \hat{n}\, dS.$$

In order to understand this consider two adjacent volumes V_1 and V_2 with a common surface S_c as shown in Fig. 9.6. Surface S_c of V_1 has outward pointing normal $\hat{n_1}$ while surface S_c of V_2 has outward pointing normal $\hat{n_2}$. Clearly $\hat{n_2} = -\hat{n_1}$. It is easy to see from Fig. 9.6 that

$$\int_{S_c} \mathbf{F} \cdot \hat{n_1} dS + \int_{S_c} \mathbf{F} \cdot \hat{n_2} dS = \int_{S_c} \mathbf{F} \cdot \hat{n_1} dS - \int_{S_c} \mathbf{F} \cdot \hat{n_1} dS = 0.$$

Given a large volume V with surface ∂V, as shown in Fig. 9.7, it can be subdivided into N smaller volumes V_i. Furthermore, the fluxes out of all the internal surfaces of these smaller volumes cancel with each other,

$$\begin{aligned}
\int_{\partial V} \mathbf{F} \cdot \hat{n}\, dS &\approx \sum_{i=1}^{N} \left(\frac{1}{\Delta V_i} \int_{S_i} \mathbf{F} \cdot \hat{n}\, dS \right) \Delta V_i \\
&= \sum_{i=1}^{N} (\text{div}\,\mathbf{F})_i\ \Delta V_i.
\end{aligned}$$

As we take the limit as $N \to \infty$ and $|\Delta V_i| \to 0$, then $\sum V_i \to V$ and $\sum \partial V_i \to \partial V$ since the inner surfaces all cancel. So when we take the limit we get the following equality,

$$\begin{aligned}
\int_{\partial V} \mathbf{F} \cdot \hat{n}\, dS &= \lim_{\substack{N\to\infty \\ |\Delta V_i|\to 0}} \sum_{i=1}^{N} (\text{div}\,\mathbf{F})_i\ \Delta V_i \\
&= \int_V \text{div}\,\mathbf{F}\, dV,
\end{aligned}$$

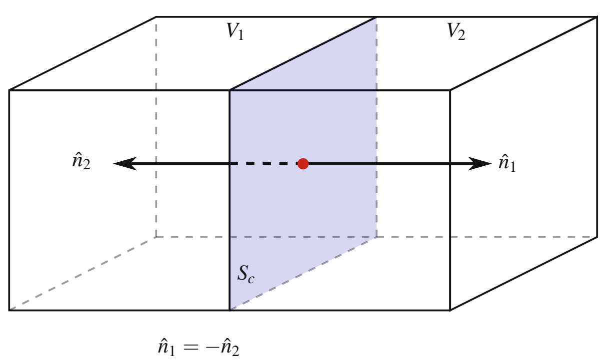


Fig. 9.6 Two adjacent cubical volumes V_1 and V_2 that share a common surface S_c. Surface S_c of V_1 has outward pointing normal $\hat{n_1}$ while surface S_c of V_2 has outward pointing normal $\hat{n_2}$ where $\hat{n_2} = -\hat{n_1}$

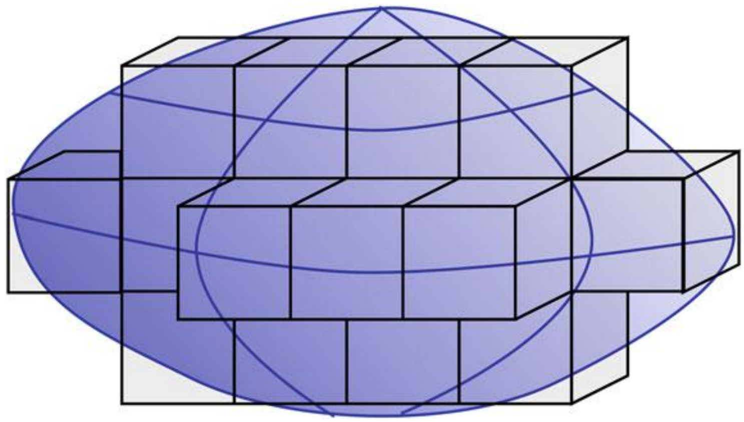


Fig. 9.7 An irregularly shaped volume V covered by smaller cubical volumes V_i. As $\Delta V_i \to 0$ the irregularly shaped volume is approximated better and better

which is exactly the divergence theorem. Recalling that the boundary of V is $\partial V = S$, and writing $\mathbf{F} \cdot d\mathbf{S}$ for $\mathbf{F} \cdot \hat{n}\ dS$, the **divergence theorem** is usually written as

Divergence Theorem
$$\int_{\partial V} \mathbf{F} \cdot d\mathbf{S} = \int_V \operatorname{div} \mathbf{F}\, dV.$$

In this section we have used the definition of the divergence of a vector field $\mathbf{F}$ on $\mathbb{R}^3$ written in Cartesian coordinates to find an expression for div $\mathbf{F}$. But Cartesian coordinates are not the only coordinate system used on $\mathbb{R}^3$. Cylindrical and spherical coordinate systems, introduced in Sect. 6.5, are also very commonly used coordinate systems on $\mathbb{R}^3$.

Question 9.2 Suppose that $\mathbf{F} = F_r\hat{e}_r + F_\theta\hat{e}_\theta + F_z\hat{e}_z$ is a vector field on $\mathbb{R}^3$ written with respect to cylindrical coordinates. The vector $\hat{e}_r$ is the unit vector in the direction of increasing r, $\hat{e}_\theta$ is the unit vector in the direction of increasing θ, and $\hat{e}_z$ is the unit vector in the direction of increasing z. Using the cylindrical volume element shown in Fig. 6.14 and a procedure similar to that of this section, show that

$$\operatorname{div} \mathbf{F} = \frac{1}{r}\frac{\partial(rF_r)}{\partial r} + \frac{1}{r}\frac{\partial F_\theta}{\partial \theta} + \frac{\partial F_z}{\partial z}.$$

Question 9.3 Suppose that $\mathbf{F} = F_r\hat{e}_r + F_\theta\hat{e}_\theta + F_\phi\hat{e}_\phi$ is a vector field on $\mathbb{R}^3$ written with respect to spherical coordinates. The vector $\hat{e}_r$ is the unit vector in the direction of increasing r, $\hat{e}_\theta$ is the unit vector in the direction of increasing θ, and $\hat{e}_\phi$ is the unit vector in the direction of increasing ϕ. Using the spherical volume element shown in Fig. 6.16 and a procedure similar to that of this section, show that

$$\operatorname{div} \mathbf{F} = \frac{1}{r^2}\frac{\partial(r^2F_r)}{\partial r} + \frac{1}{r\sin\theta}\frac{\partial(\sin\theta F_\theta)}{\partial \theta} + \frac{1}{r\sin\theta}\frac{\partial F_\phi}{\partial \phi}.$$

9.2 Curl

Given a vector field $\mathbf{F} = P\hat{i} + Q\hat{j} + R\hat{k}$, in vector calculus classes the curl of the vector field is generally defined as

$$\operatorname{curl} \mathbf{F} = \left(\frac{\partial R}{\partial y} - \frac{\partial Q}{\partial z}\right)\hat{i} + \left(\frac{\partial P}{\partial z} - \frac{\partial R}{\partial x}\right)\hat{j} + \left(\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y}\right)\hat{k}.$$

Often you will also see something like this as well

$$\operatorname{curl} \mathbf{F} = \nabla \times \mathbf{F},$$

where ∇ was defined earlier to be the operator

$$\nabla = \frac{\partial}{\partial x}\hat{i} + \frac{\partial}{\partial y}\hat{j} + \frac{\partial}{\partial z}\hat{k}$$

and the usual mnemonic device for remembering the "cross product" $\times$ of two vectors is employed:

$$\nabla \times \mathbf{F} = \begin{vmatrix} \hat{i} & \hat{j} & \hat{k} \\ \frac{\partial}{\partial x} & \frac{\partial}{\partial y} & \frac{\partial}{\partial z} \\ P & Q & R \end{vmatrix}.$$

It is important to recognize that this formulation of $\nabla \times \mathbf{F}$ is really nothing more than a mnemonic device to help us reconstruct the formula for curl $\mathbf{F}$, assuming we remember how to take the determinant of a 3×3 matrix. Again, as in the case of divergence, we will take a slightly different approach and define curl in a more geometric way that will allow both the formula for curl $\mathbf{F}$ and Stokes' theorem to just fall out. However, before we actually make the definition we need to understand some necessary background material.

We begin by considering the geometrical meaning of the dot product

$$\mathbf{v} \cdot \mathbf{w} = |\mathbf{v}|\ |\mathbf{w}|\ \cos\theta$$

as shown in Fig. 9.8. If $|\mathbf{v}| = 1$, that is, $\mathbf{v}$ is a unit vector, then we have $\mathbf{v} \cdot \mathbf{w} = |\mathbf{w}|\ \cos\theta$, which is the amount of $\mathbf{w}$ that is pointing in the $\mathbf{v}$ direction. Now suppose we want to somehow measure the "amount" of vector field $\mathbf{F}$ that is in the direction of some curve C. We can use the dot product just defined to find this, see Fig. 9.10. After that we will then want to integrate this "amount" we found.

However, we need to first understand line integrals of vector fields. Again, we assume you have had a vector calculus call and so we will cover this material without going into much detail. However, we need to make one notational comment to help you avoid confusion. We will be integrating around closed curves, which we will denote by C. These closed curves will be parameterized by arc length, which we will denote by the lower case s. We will denote by capital S a surface whose boundary is curve C, that is, $C = \partial S$. If the curve C lies in a plane we will assume the surface S also lies in that plane. If this is the case then the unit normal to the surface S will be denoted $\hat{n}$. To make our lives easier we will assume this is the case in the following.

Assume the curve C is parameterized by arc length. A point s on curve C has coordinates $x = x(s)$, $y = y(s)$, and $z = z(s)$ and a point $s + \Delta s$ has coordinates $x + \Delta x = x(s + \Delta s)$, $y + \Delta y = y(s + \Delta s)$, and $z + \Delta z = z(s + \Delta s)$, which gives

$$\begin{aligned} \Delta x &= x(s + \Delta s) - x(s), \\ \Delta y &= y(s + \Delta s) - y(s), \\ \Delta z &= z(s + \Delta s) - z(s). \end{aligned}$$

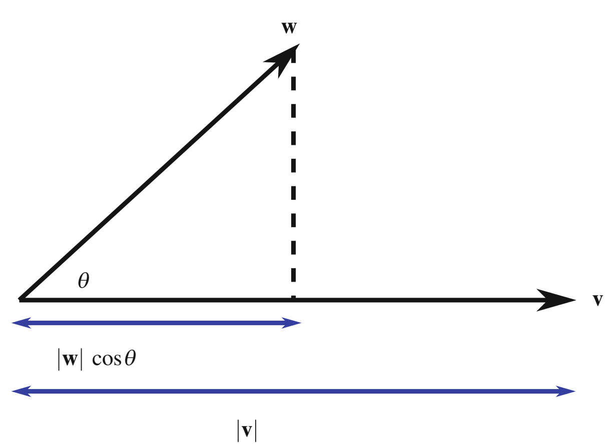


Fig. 9.8 The geometrical meaning of the dot product as $\mathbf{v} \cdot \mathbf{w} = |\mathbf{v}|\ |\mathbf{w}|\ \cos\theta$

Letting Δr be the vector from the point $\big(x(s), y(s), z(s)\big)$ to the point $\big(x(s+\Delta s), y(s+\Delta s), z(s+\Delta s)\big)$ we have

$$\frac{\Delta r}{\Delta s} = \frac{\Delta x}{\Delta s}\hat{i} + \frac{\Delta y}{\Delta s}\hat{j} + \frac{\Delta z}{\Delta s}\hat{k},$$

We define the new vector $\hat{t}$ by taking the limit of the right hand side as $\Delta s \to 0$, which of course means that $\Delta x, \Delta y, \Delta z \to 0$ as well,

$$\begin{aligned}\hat{t} &= \lim_{\Delta s \to 0} \frac{\Delta x}{\Delta s}\hat{i} + \frac{\Delta y}{\Delta s}\hat{j} + \frac{\Delta z}{\Delta s}\hat{k} \\ &= \frac{dx}{ds}\hat{i} + \frac{dy}{ds}\hat{j} + \frac{dz}{ds}\hat{k}.\end{aligned}$$

From Fig. 9.9 we can see that $\hat{t}$ is a tangent vector to the curve C. Furthermore, as $\Delta s \to 0$ it is easy to see that $|\Delta r| \to \Delta s$. Thus $|\hat{t}| = 1$ and so $\hat{t}$ is the unit tangent vector to the curve C. Also, notice that formally we could write $\hat{t}ds = \hat{i}dx + \hat{j}dy + \hat{k}dz$, which we define to be $d\mathbf{s}$.

Now we turn to the line integral of a vector field $\mathbf{F} = P\hat{i} + Q\hat{j} + R\hat{k}$ along a curve C, which is parameterized by arc length s. The amount of $\mathbf{F}$ going in curve C's direction is given by $\mathbf{F}\cdot\hat{t}$, where $\hat{t}$ is the unit tangent vector to C. See Figs. 9.10 and 9.11. Suppose we break curve C into small segments, each of length Δs. Then the amount of $\mathbf{F}$ going in the direction of C over each segment can be approximated by $(\mathbf{F}\cdot\hat{t})\Delta s$,

$$\begin{aligned}(\mathbf{F}\cdot\hat{t})\Delta s &= \left(P\hat{i} + Q\hat{j} + R\hat{k}\right)\cdot\left(\frac{\Delta x}{\Delta s}\hat{i} + \frac{\Delta y}{\Delta s}\hat{j} + \frac{\Delta z}{\Delta s}\hat{k}\right)\Delta s \\ &= \left(P\frac{\Delta x}{\Delta s} + Q\frac{\Delta y}{\Delta s} + R\frac{\Delta z}{\Delta s}\right)\Delta s\end{aligned}$$

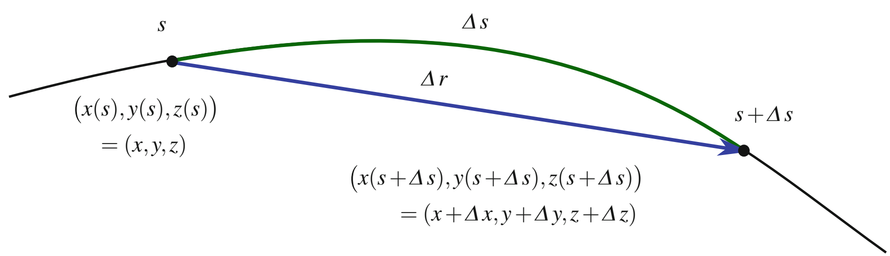


Fig. 9.9 The unit tangent vector $\hat{t}$ to the curve is the limit of $\Delta r/\Delta s = \big(\Delta x/\Delta s\big)\hat{i} + \big(\Delta y/\Delta s\big)\hat{j} + \big(\Delta z/\Delta s\big)\hat{k}$

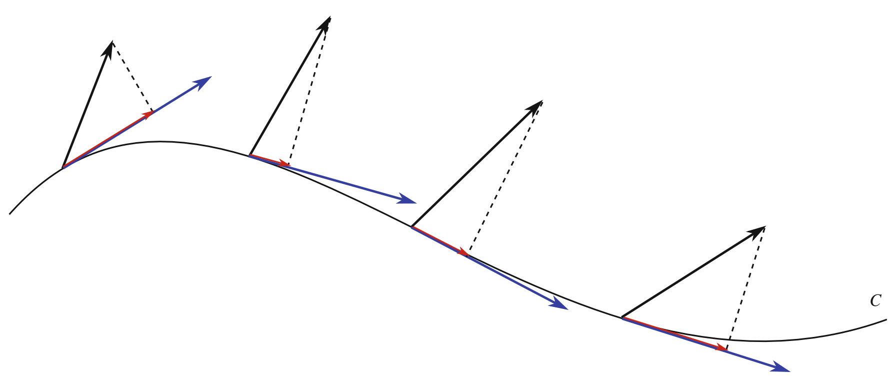


Fig. 9.10 The vector field $\mathbf{F}$ (black) is defined along a curve C. We want to know the "part" of $\mathbf{F}$ that is tangent to C. To find this we consider the unit tangent vectors to $\hat{t}$ to C (blue) and find $\mathbf{F}\cdot\hat{t}$ (red) to find the "amount" of $\mathbf{F}$ that is pointing in the $\hat{t}$ direction and is thus tangent to the curve C. We do this at every point of the curve C. Also see Fig. 9.11

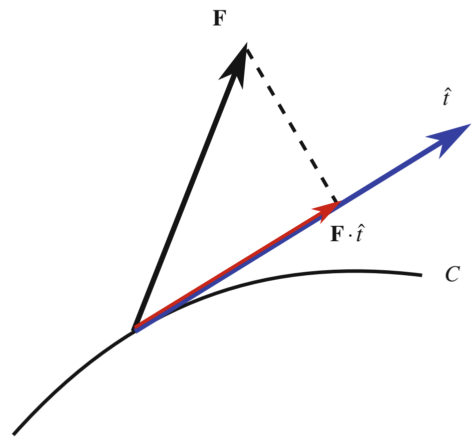


Fig. 9.11 A close-up of a section of Fig. 9.10. We want to know the "part" of **F** that is tangent to C. To find this we consider the unit tangent vectors to $\hat{t}$ to C (blue) and find $\mathbf{F}\cdot\hat{t}$ (red) to find the "amount" of **F** that is pointing in the $\hat{t}$ direction and is thus tangent to the curve C

so we get

$$\int_C \mathbf{F}\cdot\hat{t}\,ds = \lim_{|\Delta s|\to 0}\sum\left(P\frac{\Delta x}{\Delta s} + Q\frac{\Delta y}{\Delta s} + R\frac{\Delta z}{\Delta s}\right)\Delta s.$$

Using the notation from the last paragraph the integral $\int_C \mathbf{F}\cdot\hat{t}\,ds$ can also be written as $\int_C \mathbf{F}\cdot d\mathbf{s}$.

Now that we know what it means to integrate a vector field along a curve we are ready to define **curl F** at a point (x_0, y_0, z_0) as

Definition of curl	$\hat{n}\cdot\operatorname{curl}\mathbf{F} = \lim_{|\Delta S|\to 0}\frac{1}{\Delta S}\int_C \mathbf{F}\cdot\hat{t}\,ds$

where S is the surface bounded by the closed curve $C = \partial S$, ΔS is the area of that surface, $\hat{n}$ is the unit normal vector to that surface, the s in ds is the infinitesimal arc length element, and the surface area ΔS shrinks to zero about the point (x_0, y_0, z_0). Recall, a normal vector is orthogonal to, that is at a right angle to, the plane that surface S is in. The curve C along with **F** and the component of **F** along the curve is shown in Fig. 9.12. Integrating this, dividing by ΔS, and then taking the limit as $\Delta S \to 0$ gives the right hand side of the definition of curl of **F**. In essence, the curl is the "circulation" per unit area of vector field **F** over an infinitesimal path around some point.

Question 9.4 Using Fig. 9.12 and the vector field **F** as shown, estimate $\int_C \mathbf{F}\cdot\hat{t}\,ds$ based on how the integral was defined as the limit of the sum of $(\mathbf{F}\cdot\hat{t})\Delta s$. In other words, is the integral a large or small, positive or negative number, or is it close to zero? Explain your answer in terms of the definition of the integral.

Question 9.5 Draw a picture like that in Fig. 9.12 with a vector field **F** which is circulating around (x_0, y_0, z_0) in a counter-clockwise direction. Assume the tangent vectors to C are also in a counter-clockwise direction. Is the integral $\int_C \mathbf{F}\cdot\hat{t}\,ds$ a large or small, positive or negative number, or is it close to zero? Explain why using the picture and how the integral was defined. Repeat for a vector field **F** which is circulating in the clockwise direction.

We want to try to wrap our heads around this odd definition. As hopefully the last two questions illustrated, we are trying to measure how the vector field **F** is "circulating" around some point. As we take the limit as $|\Delta S| \to 0$ then we are getting the circulation per unit area at the point (x_0, y_0, z_0). But notice the left hand side of the definition, curl **F** is being doted with the unit normal vector $\hat{n}$ to the surface S, which tells us that curl **F** must be a vector itself.

The problem that makes this definition a little tricky is that in $\mathbb{R}^3$, there are an infinite number of planes that pass through the point (x_0, y_0, z_0), and thus an infinite number of possible surfaces S, with corresponding boundaries $\partial S = C$, that can be shrunk to area zero around the point (x_0, y_0, z_0). That is, in essence, the reason for the odd definition of curl that requires the dot product. This definition has to hold true for every possible plane through (x_0, y_0, z_0) and every possible surface S in that plane that shrinks to area zero around (x_0, y_0, z_0). The dot product of $\hat{n}$, the unit normal to surface S, with curl **F** on the left hand side corresponds with, or compensates for, choosing a particular plane in which to place the surface S and curve C on the right hand side.

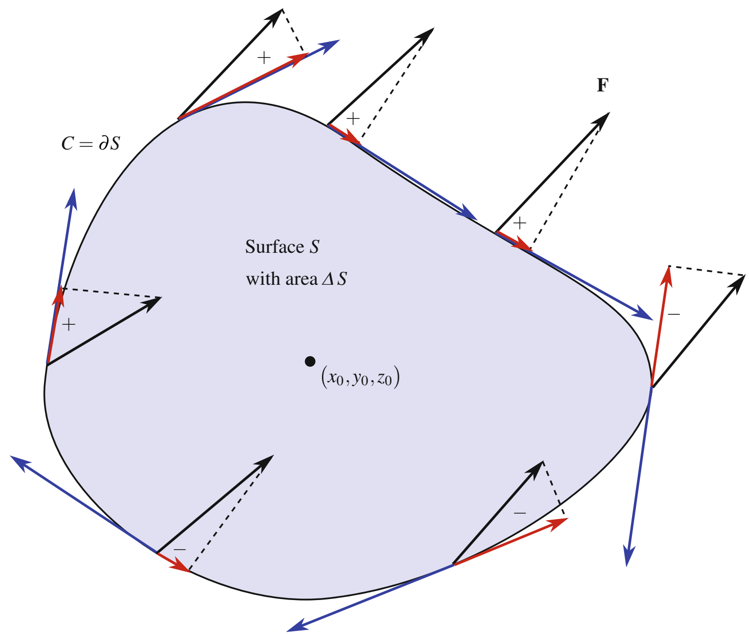


Fig. 9.12 The vector field $\mathbf{F}$ (black) along a closed curve C is shown. Unit tangent vectors to C are shown (blue) along with the "part" of $\mathbf{F}$ along C (red). Notice that sometimes the "part" of $\mathbf{F}$ along C is positive and sometimes it is negative

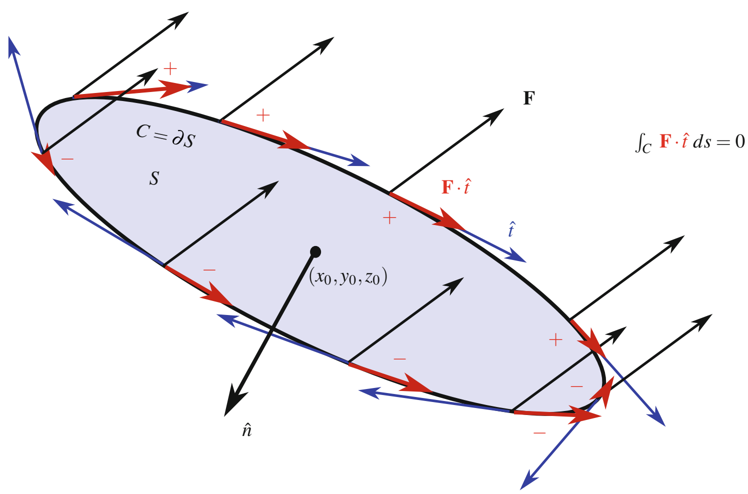


Fig. 9.13 A picture in which $\hat{n} \cdot \text{curl}\, \mathbf{F} = 0$. We have shown all the components that show up in the definition of curl $\mathbf{F}$ except the vector curl $\mathbf{F}$

Let us take a few moments to explore this a little deeper. Consider Fig. 9.13 where a surface S is shown with boundary $\partial S = C$ around the point (x_0, y_0, z_0). The unit tangent vectors $\hat{t}$ to the curve C are shown in blue at several points along C and the vector field $\mathbf{F}$ is shown in black at these same points along C. The portion of each field vector in the direction of $\hat{t}$ is shown as $\mathbf{F} \cdot \hat{t}$ in red, and its sign is also indicated, positive if $\mathbf{F} \cdot \hat{t}$ points in the direction $\hat{t}$ and negative if $\mathbf{F} \cdot \hat{t}$ points in direction $-\hat{t}$. We have drawn all vectors in the vector field $\mathbf{F}$ as parallel to each other in an attempt to show no circulation around the point (x_0, y_0, z_0) is happening. This means the "amount" of positive $\mathbf{F} \cdot \hat{t}$ is exactly the same as the "amount" of negative $\mathbf{F} \cdot \hat{t}$ leading to $\int_C \mathbf{F} \cdot \hat{t}\, ds = 0$. The figure also shows the unit normal $\hat{n}$ to surface S drawn at the point (x_0, y_0, z_0). We draw $\hat{n}$ so it is following the right-hand rule with respect to the unit tangent vectors $\hat{t}$ (blue). We do this just to standardize the positive and negative of $\mathbf{F} \cdot \hat{t}$ between different surfaces. Using our definition of curl we have

$$\hat{n} \cdot \text{curl}\, \mathbf{F} = \lim_{|\Delta S| \to 0} \frac{1}{\Delta S} \int_C \mathbf{F} \cdot \hat{t}\, ds = 0.$$

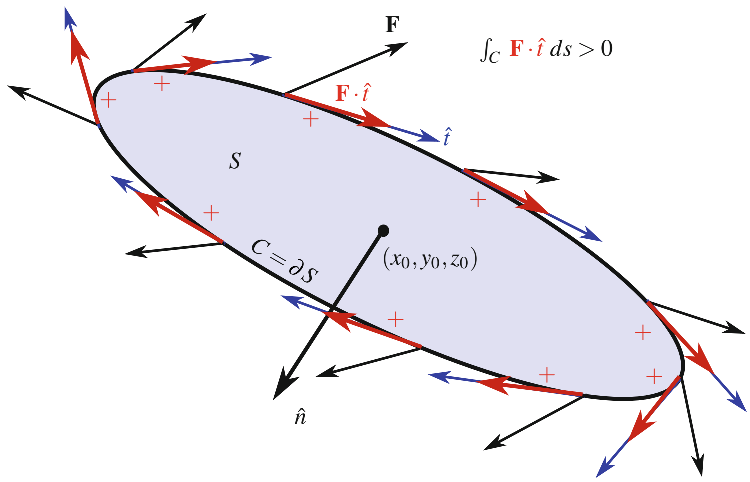


Fig. 9.14 A picture in which $\hat{n} \cdot \text{curl}\,\mathbf{F} > 0$. Again we have shown all the components that show up in the definition of curl **F** except the vector curl **F**

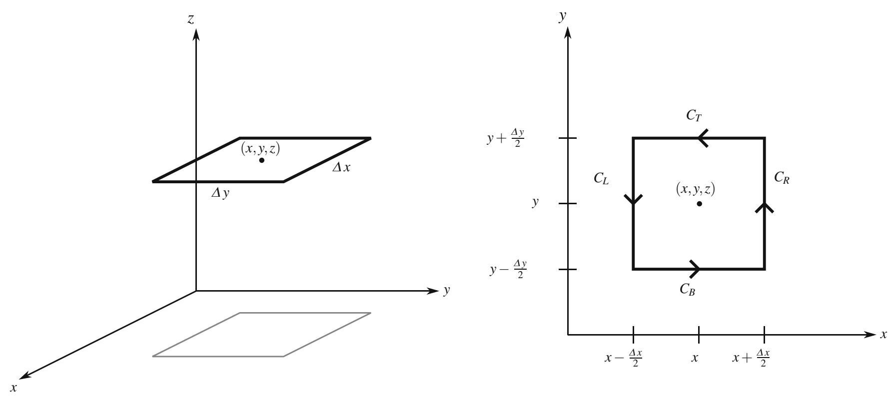


Fig. 9.15 A square curve around the point (x, y, z). This curve is in the plane parallel to the xy-plane (left), which has the unit normal $\hat{k}$. Looking down on this same square curve from the top (right). This square curve is composed of four line segments, C_T, C_B, C_L, and C_R

So, while we may not actually yet know what the vector curl **F** is, we know it must be perpendicular to $\hat{n}$ since $\hat{n} \cdot \text{curl}\,\mathbf{F} = 0$.

Now consider Fig. 9.14 where almost everything is the same as the last figure, except that this time the vectors from the vector field **F** are in some sense circulating around the point (x_0, y_0, z_0). Notice how now all the $\mathbf{F} \cdot \hat{t}$ point in the same direction as $\hat{t}$ and are therefore positive. This means that

$$\hat{n} \cdot \text{curl}\,\mathbf{F} = \lim_{|\Delta S| \to 0} \frac{1}{\Delta S} \int_C \mathbf{F} \cdot \hat{t}\, ds > 0.$$

So, again, while we may not actually yet know what the vector curl **F** is, we know that when dotted with $\hat{n}$ it is positive.

So, actually finding what the vector curl **F** is requires us to choose surfaces, with their corresponding normals, carefully. Clearly, any vector, including the vector curl **F** is a linear combination of the Euclidian unit vectors $\hat{i}$, $\hat{j}$, and $\hat{k}$. In order to find the component of curl **F** that goes in the $\hat{i}$ direction we need to find curl $\mathbf{F} \cdot \hat{i}$. Similarly, to find the component of curl **F** that goes in the $\hat{j}$ direction we need to find curl $\mathbf{F} \cdot \hat{j}$, and to find the component of curl **F** that goes in the $\hat{k}$ direction we need to find curl $\mathbf{F} \cdot \hat{k}$. So by choosing an S in a plane parallel to the yz-plane we get $\hat{n} = \hat{i}$, which allows us to find the $\hat{i}$ term of curl **F**. Similarly, choosing an S in a plane parallel to the xz-plane gives $\hat{n} = \hat{j}$, which allows us to find the $\hat{j}$ term of curl **F** and choosing an S in a plane parallel to the xy-plane gives $\hat{n} = \hat{k}$, which allows us to find the $\hat{k}$ term of curl **F**.

We begin with by choosing a path abound the point (x_0, y_0, z_0) in a plane parallel to the xy-plane, which has unit normal $\hat{k}$. Any path will do, so for simplicity we will choose a square path as shown in Fig. 9.15. Looking at this path from the top allows us to label the top, bottom, left, and right sides of the curve as C_T, C_B, C_L, and C_R as shown. Along C_B we have the unit tangent $\hat{t} = \hat{i}$, which gives $\mathbf{F} \cdot \hat{t} = \mathbf{F} \cdot \hat{i} = P$. So we have

$$\int_{C_B} \mathbf{F} \cdot \hat{t}\, ds \;=\; \int_{C_B} P\, ds \;\approx P\left(x_0, y_0 - \frac{\Delta y}{2}, z_0\right) \Delta x.$$

Along C_T we have the unit the unit tangent $\hat{t} = -\hat{i}$, which gives $\mathbf{F} \cdot \hat{t} = \mathbf{F} \cdot (-\hat{i}) = -P$. So we have

$$\int_{C_T} \mathbf{F} \cdot \hat{t}\, ds \;=\; \int_{C_B} -P\, ds \;\approx -P\left(x_0, y_0 + \frac{\Delta y}{2}, z_0\right) \Delta x.$$

Combining we have

$$\int_{C_T + C_B} \mathbf{F} \cdot \hat{t}\, ds = \frac{-\left(P\left(x_0, y_0 + \frac{\Delta y}{2}, z_0\right) - P\left(z_0, y_0 - \frac{\Delta y}{2}, z_0\right)\right)}{\Delta y} \Delta x \Delta y$$

$$\Longrightarrow \lim_{|\Delta S| \to 0} \frac{1}{\Delta S} \int_{C_T + C_B} \mathbf{F} \cdot \hat{t}\, ds = \lim_{\Delta y \to 0} \frac{-\left(P\left(x_0, y_0 + \frac{\Delta y}{2}, z_0\right) - P\left(z_0, y_0 - \frac{\Delta y}{2}, z_0\right)\right)}{\Delta y} = -\frac{\partial P}{\partial y}.$$

Along C_R we have the unit tangent $\hat{t} = \hat{j}$, which gives $\mathbf{F} \cdot \hat{t} = \mathbf{F} \cdot \hat{j} = Q$. So we have

$$\int_{C_R} \mathbf{F} \cdot \hat{t}\, ds \;=\; \int_{C_R} Q\, ds \;\approx Q\left(x_0 + \frac{\Delta x}{2}, y_0, z_0\right) \Delta y.$$

Along C_L we have the unit tangent $\hat{t} = \hat{-j}$, which gives $\mathbf{F} \cdot \hat{t} = \mathbf{F} \cdot (-\hat{j}) = -Q$. So we have

$$\int_{C_L} \mathbf{F} \cdot \hat{t}\, ds \;=\; \int_{C_L} -Q\, ds \;\approx -Q\left(x_0 - \frac{\Delta x}{2}, y_0, z_0\right) \Delta y.$$

Combining we have

$$\int_{C_R + C_L} \mathbf{F} \cdot \hat{t}\, dx = \frac{Q\left(x_0 + \frac{\Delta x}{2}, y_0, z_0\right) - Q\left(x_0 - \frac{\Delta x}{2}, y_0, z_0\right)}{\Delta x} \Delta x \Delta y$$

$$\Longrightarrow \lim_{|\Delta S| \to 0} \frac{1}{\Delta S} \int_{C_R + C_L} \mathbf{F} \cdot \hat{t}\, ds = \lim_{\Delta x \to 0} \frac{Q\left(x_0, y_0 + \frac{\Delta y}{2}, z_0\right) - Q\left(x_0, y_0 - \frac{\Delta y}{2}, z_0\right)}{\Delta x} = \frac{\partial Q}{\partial x}.$$

Since the normal to the xy-plane is $\hat{k}$ we get

$$\hat{k} \cdot \operatorname{curl} \mathbf{F} = \lim_{|\Delta S| \to 0} \frac{1}{\Delta S} \int_{C_L + C + R + C_B + C_T} \mathbf{F} \cdot \hat{t}\, ds = \frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y}.$$

So this is the component of curl $\mathbf{F}$ in the $\hat{k}$ direction.

Question 9.6 Repeat the above calculations to find that the component of curl $\mathbf{F}$ in the $\hat{i}$ direction is $\frac{\partial R}{\partial y} - \frac{\partial Q}{\partial z}$ and the component of curl $\mathbf{F}$ in the $\hat{j}$ direction is $\frac{\partial P}{\partial z} - \frac{\partial R}{\partial x}$.

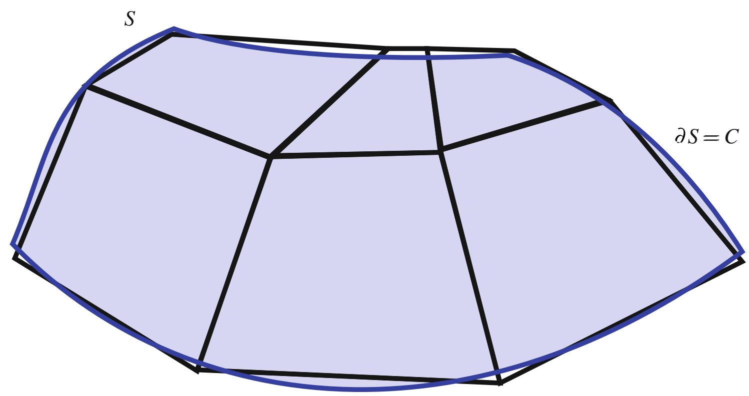


Fig. 9.16 A surface S with boundary the closed curve $\partial S = C$ covered by smaller surfaces S_i. As $\Delta S_i \to 0$ the surface S is approximated better and better

Thus, from our definition of curl we have derived the formula

$$\boxed{\text{Formula for curl } \mathbf{F} \text{ in Cartesian Coordinates} \quad \text{curl } \mathbf{F} = \left(\frac{\partial R}{\partial y} - \frac{\partial Q}{\partial z}\right)\hat{i} + \left(\frac{\partial P}{\partial z} - \frac{\partial R}{\partial x}\right)\hat{j} + \left(\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y}\right)\hat{k}.}$$

Stokes' theorem basically falls out of this definition. We will "derive" the theorem in a very non-rigorous way. Given any surface S, not necessarily in a plane, whose boundary is the closed curve C, we can break up at surface into subsurfaces S_i with boundaries C_i, as in Fig. 9.16. By a line of reasoning similar to that in the last section, if two C_i share and edge we have

$$\int_A^B \mathbf{F} \cdot \hat{t}\, ds = -\int_B^A \mathbf{F} \cdot \hat{t}\, ds$$

so these terms cancel out, see Fig. 9.17. Thus

$$\begin{aligned}
\int_{\partial S = C} \mathbf{F} \cdot \hat{t}\, ds &= \lim_{\substack{N\to\infty \\ |\Delta S_i|\to 0}} \sum_{i=0}^{N} \int_{C_i} \mathbf{F} \cdot \hat{t}\, ds \\
&= \lim_{\substack{N\to\infty \\ |\Delta S_i|\to 0}} \sum_{i=0}^{N} \left(\frac{1}{\Delta S_i} \int_{C_i} \mathbf{F} \cdot \hat{t}\, ds\right) \Delta S_i \\
&= \lim_{\substack{N\to\infty \\ |\Delta S_i|\to 0}} \sum_{i=0}^{N} \underbrace{\lim_{|\Delta S_i|\to 0} \left(\frac{1}{\Delta S_i} \int_{C_i} \mathbf{F} \cdot \hat{t}\, ds\right)}_{(\hat{n}_i \cdot \text{curl } \mathbf{F})_i} \Delta S_i \\
&= \lim_{\substack{N\to\infty \\ |\Delta S_i|\to 0}} \sum_{i=0}^{N} \left(\hat{n}_i \cdot \text{curl } \mathbf{F}\right)_i \Delta S_i \\
&= \int_S (\hat{n} \cdot \text{curl } \mathbf{F})\, dS,
\end{aligned}$$

which is exactly Stokes' theorem,

$$\boxed{\text{Stokes' Theorem} \quad \int_{\partial S} \mathbf{F} \cdot \hat{t}\, ds = \int_S (\hat{n} \cdot \text{curl } \mathbf{F})\, dS.}$$

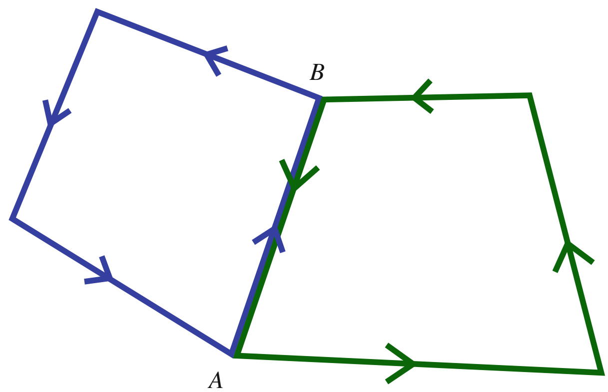


Fig. 9.17 Two adjacent subsurfaces that share a boundary C_i, which is the line segment between points A and B

An alternative way of writing Stokes' theorem is

Stokes' Theorem
$$\int_{\partial S} \mathbf{F} \cdot d\mathbf{s} = \int_S \text{curl}\,\mathbf{F} \cdot d\mathbf{S}.$$

In this section we have used the definition of the curl of a vector field $\mathbf{F}$ on $\mathbb{R}^3$ written in Cartesian coordinates to find an expression for curl $\mathbf{F}$. But of course cylindrical and spherical coordinate systems, introduced in Sect. 6.5, are also very commonly used coordinate systems on $\mathbb{R}^3$.

Question 9.7 Suppose that $\mathbf{F} = F_r\hat{e}_r + F_\theta\hat{e}_\theta + F_z\hat{e}_z$ is a vector field on $\mathbb{R}^3$ written with respect to cylindrical coordinates. The vector $\hat{e}_r$ is the unit vector in the direction of increasing r, $\hat{e}_\theta$ is the unit vector in the direction of increasing θ, and $\hat{e}_z$ is the unit vector in the direction of increasing z. Using curves that correspond to the edges of each face of the cylindrical volume element shown in Fig. 6.14 and a procedure similar to that of this section, show that

$$\text{curl}\,\mathbf{F} = (\text{curl}\,\mathbf{F})_r\hat{e}_r + (\text{curl}\,\mathbf{F})_\theta\hat{e}_\theta + (\text{curl}\,\mathbf{F})_z\hat{e}_z,$$

where

$$\begin{aligned}(\text{curl}\,\mathbf{F})_r &= \frac{1}{r}\frac{\partial F_z}{\partial \theta} - \frac{\partial F_\theta}{\partial z},\\ (\text{curl}\,\mathbf{F})_\theta &= \frac{\partial F_r}{\partial z} - \frac{\partial F_z}{\partial r},\\ (\text{curl}\,\mathbf{F})_z &= \frac{1}{r}\frac{\partial (rF_\theta)}{\partial r} - \frac{1}{r}\frac{\partial F_r}{\partial \theta}.\end{aligned}$$

Question 9.8 Suppose that $\mathbf{F} = F_r\hat{e}_r + F_\theta\hat{e}_\theta + F_\phi\hat{e}_\phi$ is a vector field on $\mathbb{R}^3$ written with respect to spherical coordinates. The vector $\hat{e}_r$ is the unit vector in the direction of increasing r, $\hat{e}_\theta$ is the unit vector in the direction of increasing θ, and $\hat{e}_\phi$ is the unit vector in the direction of increasing ϕ. Using curves that correspond to the edges of each face of the spherical volume element shown in Fig. 6.16 and a procedure similar to that of this section, show that

$$\text{curl}\,\mathbf{F} = (\text{curl}\,\mathbf{F})_r\hat{e}_r + (\text{curl}\,\mathbf{F})_\theta\hat{e}_\theta + (\text{curl}\,\mathbf{F})_\phi\hat{e}_\phi,$$

where

$$\begin{aligned}(\text{curl}\,\mathbf{F})_r &= \frac{1}{r\sin\theta}\frac{\partial(\sin\theta F_\phi)}{\partial \theta} - \frac{1}{r\sin\theta}\frac{\partial F_\theta}{\partial \phi},\\ (\text{curl}\,\mathbf{F})_\theta &= \frac{1}{r\sin\theta}\frac{\partial F_r}{\partial \phi} - \frac{1}{r}\frac{\partial (rF_\phi)}{\partial r},\\ (\text{curl}\,\mathbf{F})_\phi &= \frac{1}{r}\frac{\partial (rF_\theta)}{\partial r} - \frac{1}{r}\frac{\partial F_r}{\partial \theta}.\end{aligned}$$

9.3 Gradient

The gradient is more straight-forward than either divergence or curl. In vector calculus the gradient of the function $f : \mathbb{R}^3 \to \mathbb{R}$ is often simply defined to be the vector field

$$\text{grad } f = \frac{\partial f}{\partial x}\hat{i} + \frac{\partial f}{\partial y}\hat{j} + \frac{\partial f}{\partial z}\hat{k}.$$

Generally grad f is written as ∇f. Additionally, in vector calculus we were told that the directional derivative of a function f in the direction of the unit length vector u can be written as $D_u f = \nabla f \cdot u$ or as $u[f] = \nabla f$.

We essentially go backwards and define the **gradient** of the function f to be the vector field, which when dotted with the unit length vector u, gives the directional derivative of f in the direction u,

Definition of gradient
$$\text{grad} f \cdot u = u[f].$$

In Sect. 2.3 we basically showed that one could write

Formula for grad **F** in Cartesian Coordinates
$$\text{grad } f = \frac{\partial f}{\partial x}\hat{i} + \frac{\partial f}{\partial y}\hat{j} + \frac{\partial f}{\partial z}\hat{k}$$

when using the Cartesian coordinate system. Thus we see that the usual vector calculus definition agrees with our definition. Of course, at this point you should see the similarities between the gradient of f, grad f, and the one-form df.

We already discussed in the last section what the integral of a vector field along a curve was. The integral of the vector field $\mathbf{F}$ along a curve C was given by $\int_C \mathbf{F} \cdot \hat{t} ds$, where $\hat{t}$ was the unit tangent vector along the curve C, which was parameterized by arc length s. Writing the curve $C = c(s) = \big(x(s), y(s), z(s)\big)$ then we have

$$\begin{aligned} c'(s) &= \left(\frac{dx}{ds}, \frac{dy}{ds}, \frac{dz}{ds}\right) \\ &= \frac{dx}{ds}\hat{i} + \frac{dy}{ds}\hat{j} + \frac{dz}{ds}\hat{k} \end{aligned}$$

which is clearly the tangent to the curve $c(s)$. Also, in a manor similar to above we could formally write $c'(s)ds = \hat{i}dx + \hat{j}dy + \hat{k}dz$ and define this as $d\mathbf{s}$. Suppose we let $\mathbf{F} = \text{grad } f$, then the integral of grad f along the curve $c(s)$ from $c(a)$ to $c(b)$, where $a, b \in \mathbb{R}$ and $a \leq b$, would be

$$\begin{aligned} \int_C \text{grad } f \cdot c'(s)\, ds &= \int_a^b \left(\frac{\partial f}{\partial x}\hat{i} + \frac{\partial f}{\partial y}\hat{j} + \frac{\partial f}{\partial z}\hat{k}\right) \cdot \left(\frac{dx}{ds}\hat{i} + \frac{dy}{ds}\hat{j} + \frac{dz}{ds}\hat{k}\right)\, ds \\ &= \int_a^b \frac{\partial f}{\partial x}\frac{dx}{ds} + \frac{\partial f}{\partial y}\frac{dy}{ds} + \frac{\partial f}{\partial z}\frac{dz}{ds}\, ds \\ &= \int_a^b \frac{d}{ds} f(c(s))\, ds \\ &= f(c(b)) - f(c(a)). \end{aligned}$$

The second to last equality comes from the chain rule and the last equality comes form the fundamental theorem of calculus. Writing $c'(s)\, ds$ as $d\mathbf{s}$ we get the fundamental theorem of line integrals,

Fundamental theorem of line integrals
$$\int_C \text{grad } f \cdot d\mathbf{s} = f(c(b)) - f(c(a)).$$

In cylindrical coordinates a similar sort of argument would give

$$\text{grad } f = \frac{\partial f}{\partial r}\hat{e}_r + \frac{1}{r}\frac{\partial f}{\partial \theta}\hat{e}_\theta + \frac{\partial f}{\partial z}\hat{e}_z$$

where $\hat{e}_r$ is the unit vector in the direction of increasing r, $\hat{e}_\theta$ is the unit vector in the direction of increasing θ, and $\hat{e}_z$ is the unit vector in the direction of increasing z. In spherical coordinates we would have

$$\text{grad } f = \frac{\partial f}{\partial r}\hat{e}_r + \frac{1}{r}\frac{\partial f}{\partial \theta}\hat{e}_\theta + \frac{1}{r\sin\theta}\frac{\partial f}{\partial \phi}\hat{e}_\phi$$

where of course $\hat{e}_r$ is the unit vector in the direction of increasing r, $\hat{e}_\theta$ is the unit vector in the direction of increasing θ, and $\hat{e}_\phi$ is the unit vector in the direction of increasing ϕ.

Question 9.9 The Laplacian of a function f is defined to be $\text{div}(\text{grad } f) = \nabla \cdot (\nabla f) = \nabla \cdot \nabla f$. Sometimes this is written as $\nabla^2 f$ or as Δf. Find the Laplacian of f in Cartesian, cylindrical, and spherical coordinates.

9.4 Upper and Lower Indices, Sharps, and Flats

Before going any further now is a good time to introduce a particular notational convention. This notational convention will be useful for what follows. So far in this book we have been trying to explain mathematical concepts and so have primarily used mathematical notations. However, as we move toward a number of other examples and topics the notation that is generally used in physics will be useful.

Different notations have different strengths and weaknesses and differential geometry is one of those areas of mathematics that is very notation heavy. Mathematical notation is good for understanding theory, what mathematical objects actually are, the spaces that mathematical objects live in, and what the underlying mathematical operations and actions are. Physics notation is generally much better suited for performing computations and calculating things, even though the notation may obscure the mathematical reality underneath the computations. To be a good mathematician or physicist you really need to be comfortable with both sets of notations.

First we will introduce Einstein summation notation. Simply put, the essence of Einstein summation notation is that when you see repeated upper and lower indices you sum. That is, if you see something like

$$a^i e_i$$

that really means

$$\sum_i a^i e_i.$$

Similarly,

$$a_i e^i \equiv \sum_i a_i e^i.$$

Einstein was once reported to have quipped that the summation notation was his great contribution to mathematics! Summation notation is used extensively in tensor calculus, which we will introduce in Appendix A. Without going into the meaning of the expression at the moment, consider

$$T^{ijk}_{lm} \Lambda^l_{jk}.$$

With Einstein summation notation this actually means

$$\sum_j \sum_k \sum_l T^{ijk}_{lm} \Lambda^l_{jk}.$$

As you can see, every time an upper index and a lower index is repeated we sum over that index. Since there is both an upper and a lower index j, k, and l then we sum over j, k, and l.

Consider the vector space $V = \text{span}\{e_1, e_2, e_3\}$. The elements of of V can be written in terms of the basis vectors e_1, e_2, e_3 with the use of coefficients. That is, $v \in V$ can be written as

$$v = v^1 e_1 + v^2 e_2 + v^3 e_3 = \begin{bmatrix} v^1 \\ v^2 \\ v^3 \end{bmatrix} = v^i e_i,$$

where the vector coefficients v^1, v^2, v^3 are real numbers. Notice that the vector coefficients are written with upper indices. As usual, we continue to think of a vector as a column matrix.

The dual space of V is $V^* = \text{Span}\{e_1^*, e_2^*, e_3^*\}$, where e_1^*, e_2^*, e_3^* are simply the dual elements of e_1, e_2, e_3. That is, we have

$$\langle e_i^*, e_j \rangle = \delta_{ij} = \begin{cases} 1 & i = j \\ 0 & i \neq j. \end{cases}$$

To keep in line with Einstein summation notation from now on we will write the dual basis elements with upper indices as e^1, e^2, e^3. That is, $e_1^* \equiv e^1$, $e_2^* \equiv e^2$, $e_3^* \equiv e^3$, so we have $\langle e^i, e_j \rangle = e^i(e_j) = \delta_{ij}$. Sometimes you will see the Kronecker delta function written as δ_j^i to keep consistent with Einstein summation notation. The elements of $V^* = \text{Span}\{e^1, e^2, e^3\}$, sometimes called co-vectors, are written as

$$\alpha = \alpha_1 e^1 + \alpha_2 e^2 + \alpha_3 e^3 = [\alpha_1, \alpha_2, \alpha_3] = \alpha_i e^i,$$

where the co-vector coefficients $\alpha_1, \alpha_2, \alpha_3$ are real numbers. Again, notice that the co-vector coefficients are written with lower indices. As usual, we continue to think of a co-vector as a row matrix. So,

$$\begin{aligned}
\langle \alpha, v \rangle &= \alpha(v) \\
&= (\alpha_1 e^1 + \alpha_2 e^2 + \alpha_3 e^3)(v^1 e_1 + v^2 e_2 + v^3 e_3) \\
&= \alpha_1 e^1 (v^1 e_1 + v^2 e_2 + v^3 e_3) \\
&\quad + \alpha_2 e^2 (v^1 e_1 + v^2 e_2 + v^3 e_3) \\
&\quad + \alpha_3 e^3 (v^1 e_1 + v^2 e_2 + v^3 e_3) \\
&= \alpha_1 v^1 e^1(e_1) + \alpha_1 v^2 e^1(e_2) + \alpha_1 v^3 e^1(e_3) \\
&\quad + \alpha_2 v^1 e^2(e_1) + \alpha_2 v^2 e^2(e_2) + \alpha_2 v^3 e^2(e_3) \\
&\quad + \alpha_3 v^1 e^3(e_1) + \alpha_3 v^2 e^3(e_2) + \alpha_3 v^3 e^3(e_3) \\
&= \alpha_1 v^1 + \alpha_2 v^2 + \alpha_3 v^3.
\end{aligned}$$

What a messy calculation, not something you would want to do if you had many more then three basis elements, though of course the pattern is pretty clear. The same calculation using traditional summation notation would be

$$\begin{aligned}
\langle \alpha, v \rangle &= \alpha(v) \\
&= \left(\sum_{i=1}^{3} \alpha_i e^i \right) \left(\sum_{j=1}^{3} v^j e_j \right) \\
&= \sum_{i=1}^{3} \alpha_i e^i \left(\sum_{j=1}^{3} v^j e_j \right)
\end{aligned}$$

$$= \sum_{i=1}^{3} \sum_{j=1}^{3} \alpha_i v^j e^i \left(e_j\right)$$

$$= \sum_{i=1}^{3} \alpha_i v^i. \qquad \text{(Only } j = i \text{ terms survive.)}$$

Finally we do the same calculation with Einstein summation notation

$$\begin{aligned}
\langle \alpha, v \rangle &= \alpha(v) \\
&= (\alpha_i e^i)(v^j e_j) \\
&= \alpha_i e^i (v^j e_j) \\
&= \alpha_i v^j e^i (e_j) \\
&= \alpha_i v^i. \qquad \text{(Only } j = i \text{ terms survive.)}
\end{aligned}$$

Once you get comfortable with Einstein summation notation your hand does quite a bit less moving.

Now, let's take a look at all of this in the context of differential forms. For the moment we will assume that $M = \mathbb{R}^3$. We will write the Cartesian coordinate functions of M as x^1, x^2, x^3, with upper indices instead of lower indices. Mathematicians tend to use lower indices for coordinate functions, which is in fact what we have done up till now, but by writing coordinate functions with upper indices Einstein summation notation works nicely. So, at a point $p \in M$ we have the tangent space at p given by

$$\begin{aligned}
T_p\mathbb{R}^3 &= \text{span} \left\{ \left.\frac{\partial}{\partial x^1}\right|_p, \left.\frac{\partial}{\partial x^2}\right|_p, \left.\frac{\partial}{\partial x^3}\right|_p \right\} \\
&= \text{span} \left\{ \frac{\partial}{\partial x^1}, \frac{\partial}{\partial x^2}, \frac{\partial}{\partial x^3} \right\},
\end{aligned}$$

where the second line has the base point suppressed in the notation. The cotangent space at $p \in M$ is given by

$$\begin{aligned}
T_p^*\mathbb{R}^3 &= \text{Span} \left\{ \left. dx^1 \right|_p, \left. dx^2 \right|_p, \left. dx^3 \right|_p \right\} \\
&= \text{Span} \left\{ dx^1, dx^2, dx^3 \right\}.
\end{aligned}$$

Clearly we can consider the index in dx^i as an upper index. By convention we consider the index in $\frac{\partial}{\partial x^i}$ as a lower index because it is in the "denominator" of the expression. Therefore elements $v \in T_p\mathbb{R}^3$ are written as

$$v = v^1 \frac{\partial}{\partial x^1} + v^2 \frac{\partial}{\partial x^2} + v^3 \frac{\partial}{\partial x^3} = v^i \frac{\partial}{\partial x^i},$$

where v^1, v^2, v^3 are real numbers. Again vectors have coefficients with upper indices. Elements $\alpha \in T_p^*\mathbb{R}^3$ are written as

$$\alpha = \alpha_1 dx^1 + \alpha_2 dx^2 + \alpha_2 dx^3 = \alpha_i dx^i,$$

where $\alpha^1, \alpha^2, \alpha^3$ are real numbers. As we can see, covectors, that is, differential forms, have coefficients with lower indices.

In summary, vector basis elements are written with lower indices and vector components are written with upper indices. Covector (differential form) basis elements are written with upper indices and covector (differential form) components are written with lower indices.

Now we introduce the so called **musical isomorphisms**. In reality, the musical isomorphisms depend on something called a metric. Very roughly you can think of a metric as something that allows you to measure the distance between two points. We will discuss the **Euclidian metric** later on, after we have introduced tensors, but it is, in essence, the usual distance function on $\mathbb{R}^n$ that we are familiar with. However, right now we will not be that mathematically rigorous and give a "definition" of the musical isomorphisms that does not explicitly rely on the metric. You should simply keep in mind that at the moment we are sweeping some of the mathematical theory under the rug, so to speak.

The musical isomorphisms are called musical because of the notation that is used, the **flat** $\flat$ and the **sharp** $\sharp$. If you have ever sung or played an instrument you know that the $\flat$ tells you to lower the pitch of a musical note and the $\sharp$ tells you to raise the pitch of a musical note. Similarly, the $\flat$ isomorphism is said to "lower indices" while the $\sharp$ isomorphism is said to "raise indices."

The $\flat$ isomorphism is given by

$$\begin{aligned}\flat : T_pM &\longrightarrow T_p^*M \\ v^i\frac{\partial}{\partial x^i} &\longmapsto v_i dx^i \qquad \text{where } v_i = v^i.\end{aligned}$$

The way this mapping is written is $v^\flat$, just like the flat is written in music, so we would write

$$v^\flat = \left(v^i\frac{\partial}{\partial x^i}\right)^\flat \equiv v_i dx^i,$$

where $v_i = v^i$. So for example we would have

$$\left(7\frac{\partial}{\partial x^1} + 3\frac{\partial}{\partial x^2} - 6\frac{\partial}{\partial x^3}\right)^\flat = 7dx^1 + 3dx^2 - 6dx^3.$$

Notice what is happening, we are turning a vector, where the vector components are written with upper indices, into a one-form, where the one-form components are written with lower indices, so we are "lowering" the indices, which explains the use of the $\flat$ symbol.

The $\sharp$ isomorphism is given by

$$\begin{aligned}\sharp : T_p^*M &\longrightarrow T_pM \\ \alpha_i dx^i &\longmapsto \alpha^i\frac{\partial}{\partial x^i} \qquad \text{where } \alpha^i = \alpha_i.\end{aligned}$$

Again, the way this mapping is written is $\alpha^\sharp$, just like the sharp is written in music, so we would write

$$\alpha^\sharp = \left(\alpha_i dx^i\right)^\sharp \equiv \alpha^i\frac{\partial}{\partial x^i},$$

where $\alpha^i = \alpha_i$. So for example, we would have

$$\left(3dx^1 - 9dx^2 + 8dx^3\right)^\sharp = 3\frac{\partial}{\partial x^1} - 9\frac{\partial}{\partial x^2} + 8\frac{\partial}{\partial x^3}.$$

We are turning a differential one-form, or co-vector, where components are written with lower indices, into a vector, where components are written with upper indices, so we are "raising" the indices, which explains the use of the $\sharp$ symbol.

You should note that these "definitions" of the flat and sharp operators only work if we are writing our vectors and one-forms in Cartesian coordinates. This is the price we pay for not giving the real definitions of the flat and short operators in terms of the metric on a manifold. So, for the moment we can only take the flats of vectors and the sharps of one-forms if they are written in terms of Cartesian coordinates but not if our vectors or one-forms are written with respect to any other coordinate system.

If $\mathbf{F}$ is a vector field on M, that is a section of the tangent bundle TM, then $\mathbf{F}^\flat$ is a differential one-form on M, that is, a section of the cotangent bundle T^*M. And if α is a differential one-form on M, that is a section of the cotangent bundle T^*M, then $\alpha^\sharp$ is a vector field on M, that is, a section of the tangent bundle TM. Recalling that we always write vectors as column matrices and one-forms as row matrices, we can think of $\flat$ as turning a column matrix into into a row matrix and $\sharp$ as turning a row matrix into a column matrix.

Now, let us remind ourselves very briefly of the Hodge star operator from Sect. 5.6. We had computed the following Hodge star mappings

$$
\begin{aligned}
* : \bigwedge^0\left(\mathbb{R}^3\right) \longrightarrow \bigwedge^3\left(\mathbb{R}^3\right)\\
* : \bigwedge^1\left(\mathbb{R}^3\right) \longrightarrow \bigwedge^2\left(\mathbb{R}^3\right)\\
* : \bigwedge^2\left(\mathbb{R}^3\right) \longrightarrow \bigwedge^1\left(\mathbb{R}^3\right)\\
* : \bigwedge^3\left(\mathbb{R}^3\right) \longrightarrow \bigwedge^0\left(\mathbb{R}^3\right).
\end{aligned}
$$

Which were given by

$$
\begin{aligned}
&* 1 = dx^1 \wedge dx^2 \wedge dx^3 = dx \wedge dy \wedge dz,\\
&* dx = dy \wedge dz, \quad *dy = dz \wedge dx, \quad *dz = dx \wedge dy,\\
&* dy \wedge dz = dx, \quad *dz \wedge dx = dy, \quad *dx \wedge dy = dz,\\
&* dx \wedge dy \wedge dz = 1.
\end{aligned}
$$

Suppose we have the vector field $\mathbf{F} = P\frac{\partial}{\partial x} + Q\frac{\partial}{\partial y} + R\frac{\partial}{\partial z}$. We will define the mapping $(* \circ \flat)$ as first flatting the vector field to get a one-form and then Hodge staring that one-form to get a two form;

$$
\begin{aligned}
(* \circ \flat)\mathbf{F} &\equiv *(\mathbf{F}^{\flat})\\
&= *\left(\left(P\frac{\partial}{\partial x} + Q\frac{\partial}{\partial y} + R\frac{\partial}{\partial z}\right)^{\flat}\right)\\
&= *\left(P dx + Q dy + R dz\right)\\
&= P dy \wedge dz + Q dz \wedge dx + R dx \wedge dy.
\end{aligned}
$$

With this final mapping in hand, we are now ready to look at the relationship between standard vector calculus and differential forms.

9.5 Relationship to Differential Forms

9.5.1 Grad, Curl, Div and Exterior Differentiation

Understanding the relationship between differential forms and vector calculus really amounts to little more then seeing that the appropriate mappings lead to everything working out nicely. We will essentially look at each piece, one at a time. As we go through this process we will discover the generalized Stokes' theorem, usually just called Stokes' theorem. The generalized Stokes' theorem, often simply called Stokes' theorem, is a simultaneous generalization of the three vector calculus theorems you already know,

- fundamental theorem of line integrals,
- (vector calculus) Stokes' theorem,
- (vector calculus) divergence theorem.

The generalized Stokes' theorem will allow us to use differential forms to rewrite all three of these theorems in a very nice and compact way.

Now we will show that the following diagram commutes.

$$\begin{array}{ccc} C\left(\mathbb{R}^3\right) & \xrightarrow{\text{grad}} & T\mathbb{R}^3 \\ \Big\downarrow \text{id} & & \Big\downarrow \flat \\ \bigwedge^0\left(\mathbb{R}^3\right) & \xrightarrow{d} & \bigwedge^1\left(\mathbb{R}^3\right) \end{array}$$

First of all we make sure we recall what the spaces in the four corners are. In the upper left hand corner we have the space $C\left(\mathbb{R}^3\right)$, which is simply the space of continuous functions on the manifold $\mathbb{R}^3$. In the upper right hand corner the space $T\mathbb{R}^3$ is simply the tangent bundle of $\mathbb{R}^3$, which contains all the vector fields on $\mathbb{R}^3$. In the lower left hand corner is $\bigwedge^0\left(\mathbb{R}^3\right)$, the zero-forms on $\mathbb{R}^3$, which are just the continuous functions on $\mathbb{R}^3$. Notice that the space $\bigwedge^0\left(\mathbb{R}^3\right)$ is exactly the same space as $C\left(\mathbb{R}^3\right)$. In the lower left hand corner is the space $\bigwedge^1\left(\mathbb{R}^3\right) = T^*\mathbb{R}^3$, the one-forms on $\mathbb{R}^3$. The mappings between these spaces are grad, $\flat$, d, and id, the identity mapping.

"The diagram commutes" is one of those math phrases that you will hear a lot if you are a math major, and even if you are a physics major you may hear it from time to time. Let us take a moment to explain what it means. If you start with a function $f \in C\left(\mathbb{R}^3\right)$ then regardless of the path you take, either across the top then down the right hand side or down the left hand side and then across the bottom, you will get the same element in $\bigwedge^1\left(\mathbb{R}^3\right)$. That is, we have

$$(\text{grad } f)^\flat = d\,(\text{id}\,(f))\,.$$

So saying the above diagram commutes is exactly the same thing as saying $(\text{grad } f)^\flat = d\,(\text{id}\,(f)) = df$. Now we will show this equality. Given a function on $\mathbb{R}^3$, $f \in C(\mathbb{R}^3)$ we know that grad turns that function into a vector field $\nabla f \in T(\mathbb{R})$ according to

$$\begin{aligned} \nabla f &= \frac{\partial f}{\partial x}\hat{i} + \frac{\partial f}{\partial y}\hat{j} + \frac{\partial f}{\partial z}\hat{k} \\ &= \frac{\partial f}{\partial x}\frac{\partial}{\partial x} + \frac{\partial f}{\partial y}\frac{\partial}{\partial y} + \frac{\partial f}{\partial z}\frac{\partial}{\partial z}. \end{aligned}$$

Flatting this we have

$$(\nabla f)^\flat = \frac{\partial f}{\partial x}dx + \frac{\partial f}{\partial y}dy + \frac{\partial f}{\partial z}dz.$$

Notice, this is exactly equal to the exterior derivative of f,

$$df = \frac{\partial f}{\partial x}dx + \frac{\partial f}{\partial y}dy + \frac{\partial f}{\partial z}dz,$$

hence we have shown that

$$df = (\nabla f)^\flat\,.$$

Similarly, we want to show the below diagram commutes as well.

$$\begin{array}{ccc} T\mathbb{R}^3 & \xrightarrow{\text{curl}} & T\mathbb{R}^3 \\ \Big\downarrow \flat & & \Big\downarrow * \circ \flat \\ \bigwedge^1\left(\mathbb{R}^3\right) & \xrightarrow{d} & \bigwedge^2\left(\mathbb{R}^3\right) \end{array}$$

That is, we want to show that $d(\mathbf{F}^\flat) = *\big((\text{curl}\,\mathbf{F})^\flat\big)$. Given a vector field $\mathbf{F} \in T\mathbb{R}^3$. Recalling the definition of the mapping we defined in the last section, $(* \circ \flat)\mathbf{F} \equiv *(\mathbf{F}^\flat)$, we want to find $(* \circ \flat)(\text{curl}\,\mathbf{F})$. First, we know

$$\text{curl}\,\mathbf{F} = \left(\frac{\partial R}{\partial y} - \frac{\partial Q}{\partial z}\right)\frac{\partial}{\partial x} + \left(\frac{\partial P}{\partial z} - \frac{\partial R}{\partial x}\right)\frac{\partial}{\partial y} + \left(\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y}\right)\frac{\partial}{\partial z}.$$

So we get

$$\begin{aligned}(* \circ \flat)(\text{curl}\,\mathbf{F}) &= *\big((\text{curl}\,\mathbf{F})^\flat\big) \\ &= *\left(\left(\frac{\partial R}{\partial y} - \frac{\partial Q}{\partial z}\right)dx + \left(\frac{\partial P}{\partial z} - \frac{\partial R}{\partial x}\right)dy + \left(\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y}\right)dz\right) \\ &= \left(\frac{\partial R}{\partial y} - \frac{\partial Q}{\partial z}\right)dy \wedge dz + \left(\frac{\partial P}{\partial z} - \frac{\partial R}{\partial x}\right)dz \wedge dx + \left(\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y}\right)dx \wedge dy.\end{aligned}$$

Now we want to compute $d(\mathbf{F}^\flat)$. First we find

$$\begin{aligned}\mathbf{F}^\flat &= \left(P\frac{\partial}{\partial x} + Q\frac{\partial}{\partial y} + R\frac{\partial}{\partial z}\right)^\flat \\ &= P\,dx + Q\,dy + R\,dz.\end{aligned}$$

Then we take the exterior derivative of this

$$\begin{aligned}d(\mathbf{F}^\flat) &= d(P\,dx + Q\,dy + R\,dz) \\ &= dP \wedge dx + dQ \wedge dy + dR \wedge dz \\ &= \left(\frac{\partial P}{\partial x}dx + \frac{\partial P}{\partial y}dy + \frac{\partial P}{\partial z}dz\right) \wedge dx \\ &\quad + \left(\frac{\partial Q}{\partial x}dx + \frac{\partial Q}{\partial y}dy + \frac{\partial Q}{\partial z}dz\right) \wedge dy \\ &\quad + \left(\frac{\partial R}{\partial x}dx + \frac{\partial R}{\partial y}dy + \frac{\partial R}{\partial z}dz\right) \wedge dz \\ &= \frac{\partial P}{\partial y}dy \wedge dx + \frac{\partial P}{\partial z}dz \wedge dx + \frac{\partial Q}{\partial x}dx \wedge dy + \frac{\partial Q}{\partial z}dz \wedge dy \\ &\quad + \frac{\partial R}{\partial x}dx \wedge dz + \frac{\partial R}{\partial y}dy \wedge dz \\ &= \left(\frac{\partial R}{\partial y} - \frac{\partial Q}{\partial z}\right)dy \wedge dz + \left(\frac{\partial P}{\partial z} - \frac{\partial R}{\partial x}\right)dz \wedge dx + \left(\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y}\right)dx \wedge dy.\end{aligned}$$

Thus we have just shown that $d(\mathbf{F}^\flat) = (* \circ \flat)(\text{curl}\,\mathbf{F})$. Finally, we would like to show that this diagram commutes.

$$\begin{array}{ccc} T\mathbb{R}^3 & \xrightarrow{\text{div}} & C\left(\mathbb{R}^3\right) \\ \Big\downarrow{\scriptstyle * \circ \flat} & & \Big\downarrow{\scriptstyle *} \\ \bigwedge^2\left(\mathbb{R}^3\right) & \xrightarrow{d} & \bigwedge^3\left(\mathbb{R}^3\right) \end{array}$$

Given $\mathbf{F}$ we have

$$\operatorname{div}\mathbf{F} = \frac{\partial P}{\partial x} + \frac{\partial Q}{\partial y} + \frac{\partial R}{\partial z}.$$

Taking the Hodge star of this we have

$$\begin{aligned} *(\operatorname{div}\mathbf{F}) &= *\left(\frac{\partial P}{\partial x} + \frac{\partial Q}{\partial y} + \frac{\partial R}{\partial z}\right) \\ &= \left(\frac{\partial P}{\partial x} + \frac{\partial Q}{\partial y} + \frac{\partial R}{\partial z}\right) dx \wedge dy \wedge dz. \end{aligned}$$

Next, we find

$$\begin{aligned} (* \circ \flat)\mathbf{F} &= *(\mathbf{F}^\flat) \\ &= *\left(\left(P\frac{\partial}{\partial x} + Q\frac{\partial}{\partial y} + R\frac{\partial}{\partial z}\right)^\flat\right) \\ &= *\,(Pdx + Qdy + Rdz) \\ &= Pdy \wedge dz + Qdz \wedge dx + Rdx \wedge dy \end{aligned}$$

and then taking the exterior derivative of this,

$$\begin{aligned} d\left((* \circ \flat)\mathbf{F}\right) &= d\left(Pdy \wedge dz + Qdz \wedge dx + Rdx \wedge dy\right) \\ &= dP \wedge dy \wedge dz + dQ \wedge dz \wedge dx + dR \wedge dx \wedge dy \\ &= \frac{\partial P}{\partial x} dx \wedge dy \wedge dz + \frac{\partial Q}{\partial y} dy \wedge dz \wedge dx + \frac{\partial R}{\partial z} dz \wedge dx \wedge dy \\ &= \left(\frac{\partial P}{\partial x} + \frac{\partial Q}{\partial y} + \frac{\partial R}{\partial z}\right) dx \wedge dy \wedge dz. \end{aligned}$$

Hence we have found that

$$d\left((* \circ \flat)\mathbf{F}\right) = *(\operatorname{div}\mathbf{F}),$$

which was what we wanted to find. When we connect these three diagrams together what we get is this:

$$\begin{tikzcd} C\left(\mathbb{R}^3\right) \arrow[r, "\text{grad}"] \arrow[d, "\text{id}"] & T\mathbb{R}^3 \arrow[r, "\text{curl}"] \arrow[d, "\flat"] & T\mathbb{R}^3 \arrow[r, "\text{div}"] \arrow[d, "*\circ\flat"] & C\left(\mathbb{R}^3\right) \arrow[d, "*"] \\ \bigwedge^0\left(\mathbb{R}^3\right) \arrow[r, "d"] & \bigwedge^1\left(\mathbb{R}^3\right) \arrow[r, "d"] & \bigwedge^2\left(\mathbb{R}^3\right) \arrow[r, "d"] & \bigwedge^3\left(\mathbb{R}^3\right) \end{tikzcd}$$

So, we have very clear relationship between the three vector calculus operators and the exterior derivative. The gradient, the curl, and the divergence are all nothing other than exterior derivatives in a different guise. Vector calculus turns out to be another way of formulating and presenting exterior derivatives and differential forms on $\mathbb{R}^3$. In vector calculus everything is kept as a vector, yet utilized in the same way that forms are utilized. The problem with vector calculus is that it can not be generalized to $\mathbb{R}^n$ for $n > 3$ or to general manifolds, while our notions of differential forms and exterior derivatives can be generalized to both $\mathbb{R}^n$ for $n > 3$ and to general manifolds.

Question 9.10 Let $\mathbf{F} = P\hat{i} + Q\hat{j} + R\hat{k}$ be a vector field and f a continuous function. Show that

(a) Show curl (grad $\mathbf{F}$)=0.
(b) Show div (curl $\mathbf{F}$)=0.
(c) Explain why you would expect this given what you know about the exterior derivative d.

Question 9.11 Prove the following identities:

(a) grad $f = (df)^\sharp$
(b) curl $\mathbf{F} = [*(df\flat)]^\sharp$
(c) div $\mathbf{F} = *d(*(\mathbf{F}^\flat))$
(d) $v \times w = [*(v^\flat \wedge w^\flat)]^\sharp$ for vector fields v and w.
(e) $(v \cdot w) dx \wedge dy \wedge dz = v^\flat \wedge *(w^\flat)$

9.5.2 Fundamental Theorem of Line Integrals

Now, let's take a look at the fundamental theorem of line integrals. Suppose we are given a curve C, given by $c(s) = \big(x(s), y(s), z(s)\big)$ with end points $c(a) = \big(x(a), y(a), z(a)\big)$ and $c(b) = \big(x(b), y(b), z(b)\big)$, where $a, b \in \mathbb{R}$ and $a \leq b$, then the fundamental theorem of line integrals is given by

$$\int_C (\text{grad } f) \cdot d\mathbf{s} = f(c(b)) - f(c(a)).$$

We will write the boundary of curve C as $\partial C = \big\{c(b)\big\} - \big\{c(a)\big\}$. Boundaries are explained in depth in Chap. 11. By doing that then we can view, or define, $f(c(b)) - f(c(a))$ as the degenerate form of the integral of f on ∂C, that is, as

$$\int_{\partial C} f \equiv f(c(b)) - f(c(a)).$$

We admit, this notation for the zero dimensional case may seem little contrived, but by making this definition we are able to have one consistent notation for the generalized Stokes' theorem. This allows us to write the right hand side of the fundamental theorem of line integrals as $\int_{\partial C} f$.

Let the curve C be parameterized by s and given by c, so we have $C = c(s) = \big(x(s), y(s), z(s)\big)$ and $c'(s) = \frac{dx}{ds}\hat{i} + \frac{dy}{ds}\hat{j} + \frac{dz}{dt}\hat{s}$. If $\mathbf{F}$ is a vector field we have the integral of $\mathbf{F}$ along C as

$$\begin{aligned}
\int_C \mathbf{F} \cdot d\mathbf{s} &\equiv \int_C \mathbf{F} \cdot c'(s)\, ds \\
&= \int_C \left(P\hat{i} + Q\hat{j} + R\hat{k}\right) \cdot \left(\frac{dx}{ds}\hat{i} + \frac{dy}{ds}\hat{j} + \frac{dz}{ds}\hat{k}\right) ds \\
&= \int_C \left(P\frac{dx}{dt} + Q\frac{dy}{dt} + R\frac{dz}{dt}\right) ds \\
&= \int_C \mathbf{F}^\flat.
\end{aligned}$$

If $\mathbf{F} = \text{grad } f$ then we have $(\text{grad } f)^\flat = df$. In other words, we can write the right hand side of the fundamental theorem of line integrals as $\int_C df$. Combining this with what we had above, we could write the fundamental theorem of line integrals as

$$\int_C df = \int_{\partial C} f.$$

Recognizing that f is in fact a zero-form, which we could denote at α and C is a one dimensional manifold M we could write the fundamental theorem of line integrals as

$$\int_M d\alpha = \int_{\partial M} \alpha.$$

9.5.3 *Vector Calculus Stokes' Theorem*

Now we turn to look at the vector calculus version of Stokes' theorem,

$$\int_S \text{curl}\,\mathbf{F} \cdot \hat{n}\; dS = \int_{\partial S} \mathbf{F} \cdot c'(s)\; ds.$$

From above we know we can rewrite the right hand side of the Stokes' theorem as $\int_{\partial S} \mathbf{F}^\flat$.

For the left hand side of Stokes' theorem we want to do something similar to what we did above, which requires us to show that

$$\int_S \text{curl}\,\mathbf{F} \cdot \hat{n}\; dS = \int_S * \left((\text{curl}\,\mathbf{F})^\flat \right).$$

In order to make the notation simpler we will actually show

$$\int_S \mathbf{F} \cdot \hat{n}\; dS = \int_S * \left(F^\flat \right).$$

We then get the required identity simply by replacing the vector field $\mathbf{F}$ by the vector field curl $\mathbf{F}$.

However, instead of a tidy computation like when we showed that $\int_C \mathbf{F} \cdot dr = \int_C \mathbf{F}^\flat$ in the line integral case, here it is easiest to go back to the basic definition of flux through a surface. This argument will not be rigorous but should suffice to help you understand what is going on. An infinitesimal area dS with normal $\hat{n} = n_1\hat{i} + n_2\hat{j} + n_3\hat{k}$ can be decomposed into infinitesimal pieces in the xy-plane, the yz-plane, and the xz-plane. Here $n_1 dS$ is the piece of dS that is in the yz-plane. This piece of dS has the volume form $dy \wedge dz$. Similarly, $n_2 dS$ is the piece of dS that is in the xz-plane and has the volume form $dz \wedge dx$. Finally $n_3 dS$ is the piece of dS in the xy-plane which has volume form $dx \wedge dy$. See Fig. 9.18. Thus we can write

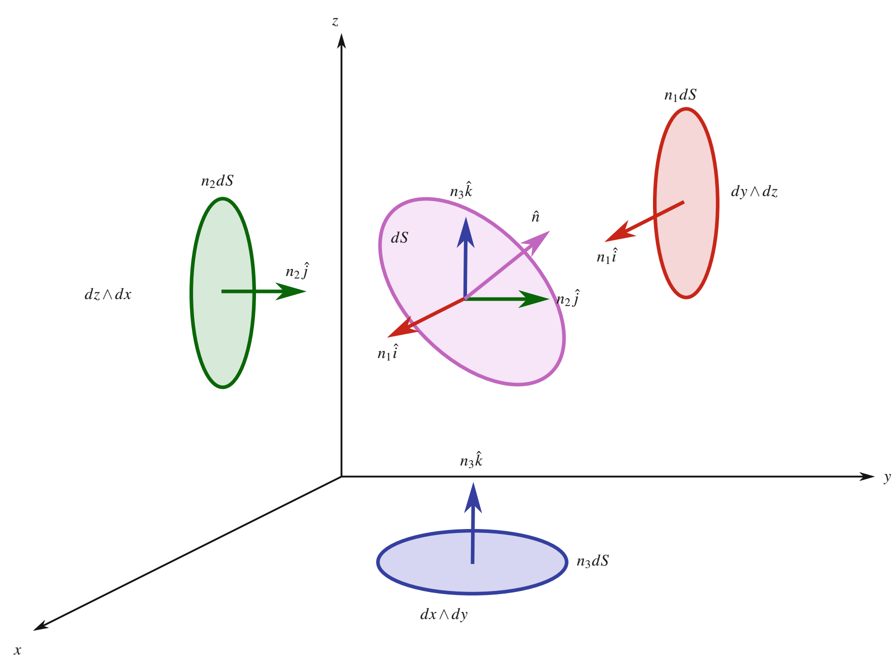


Fig. 9.18 The infinitesimal area dS with the unit normal vector $\hat{n} = n_1\hat{i} + n_2\hat{j} + n_3\hat{k}$ is decomposed into $n_1 dS$, $n_2 dS$, and $n_3 dS$, with unit normals $n_1\hat{i}$, $n_2\hat{j}$, and $n_3\hat{k}$, respectively

$\mathbf{F} \cdot \hat{n}\ dS$ as $\mathbf{F} \cdot d\mathbf{S}$ and so have

$$\begin{aligned}
\int_S \mathbf{F} \cdot d\mathbf{S} &= \int_S \left(P\hat{i} + Q\hat{j} + R\hat{k}\right) \cdot \left(n_1\hat{i} + n_2\hat{j} + n_3\hat{k}\right)\ dS \\
&= \int_S (Pn_1 + Qn_2 + Rn_3)\ dS \\
&= \int_S\ Pn_1\ dS + Qn_2\ dS + Rn_3\ dS \\
&= \int_S\ Pdy \wedge dz + Qdz \wedge dx + Rdx \wedge dy \\
&= \int_S * \left(\mathbf{F}^\flat\right)
\end{aligned}$$

which was what we wanted to show. When we replace $\mathbf{F}$ by curl $\mathbf{F}$ we have

$$\begin{aligned}
\int_S \text{curl}\,\mathbf{F} \cdot d\mathbf{S} &= \int_S\ * \left((\text{curl}\,\mathbf{F})^\flat\right) \\
&= \int_S\ d(\mathbf{F}^\flat).
\end{aligned}$$

Thus we have rewritten the left hand side of Stokes' theorem. Putting the right hand and left hand pieces together together we can rewrite the vector calculus version of Stokes' theorem as

$$\int_S d(\mathbf{F}^\flat) = \int_{\partial S}\ F^\flat.$$

Clearly, $\mathbf{F}^\flat$ is a one-form α and S is a two dimensional manifold M, which means we have written Stokes' theorem as

$$\int_M\ d\alpha = \int_{\partial M}\ \alpha.$$

9.5.4 Divergence Theorem

Finally, we take a look at the divergence theorem from vector calculus,

$$\int_V \text{div}\,\mathbf{F}\,dV = \int_{\partial V}\ \mathbf{F} \cdot d\mathbf{S}.$$

We notice that the integrand on the left hand side is nothing more than $*(\text{div}\,\mathbf{F})$, which can be rewritten as $d\big((* \circ \flat)\mathbf{F}\big)$. We have also seen the right hand side $\int_{\partial V}\ \mathbf{F} \cdot dS$ written as $\int_{\partial V}\ *(\mathbf{F}^\flat)$. This allows us to write the divergence theorem as

$$\int_V\ d\left((* \circ \flat)\mathbf{F}\right) = \int_{\partial V}\ *(\mathbf{F}^\flat).$$

Writing the three dimensional manifold V as M and the three two-form $(* \circ \flat)\mathbf{F}$ as α we have the divergence theorem as

$$\int_M\ d\alpha = \int_{\partial M}\ \alpha.$$

Thus we have written all three of our vector calculus theorems in the same form in differential forms notation. This identity,

$$\boxed{\int_M d\alpha = \int_{\partial M} \alpha}$$

is exactly the generalized Stokes' theorem that we will prove in a couple of chapters.

9.6 Summary, References, and Problems

9.6.1 Summary

Geometrically the divergence measures how much the vector field "diverges", or "spreads out", at the point (x_0, y_0, z_0). Given a small three dimensional region V about the point (x_0, y_0, z_0) with boundary ∂V and volume ΔV the divergence of $\mathbf{F}$ at (x_0, y_0, z_0) is defined by

Definition of divergence	$\text{div}\,\mathbf{F} = \lim_{\Delta V \to 0} \frac{1}{\Delta V} \int_{\partial V} \mathbf{F} \cdot \hat{n}\, dS.$

In essence, the curl measures the "circulation" per unit area of vector field $\mathbf{F}$ over an infinitesimal path around some point. Suppose S is the surface bounded by the closed curve $C = \partial S$, ΔS is the area of that surface, $\hat{n}$ is the unit normal vector to that surface, the s in ds is the infinitesimal arc length element, and the surface area ΔS shrinks to zero about the point (x_0, y_0, z_0). Then the curl $\mathbf{F}$ at a point (x_0, y_0, z_0) is defined as

Definition of curl	$\hat{n} \cdot \text{curl}\,\mathbf{F} = \lim_{\lvert\Delta S\rvert \to 0} \frac{1}{\Delta S} \int_C \mathbf{F} \cdot \hat{t}\, ds.$

We essentially go backwards and define the gradient of the function f to be the vector field, which when dotted with the unit length vector u, gives the directional derivative of f in the direction u,

Definition of gradient	$\text{grad} f \cdot u = u[f].$

Using these geometrical definitions of divergence, curl, and gradient, it is possible to obtain the standard formula definitions from vector calculus. Letting $\mathbf{F} = P\hat{i} + Q\hat{j} + R\hat{k}$ we have

Formula for divergence $\mathbf{F}$ in Cartesian coordinates	$\text{div}\,\mathbf{F} = \frac{\partial P}{\partial x} + \frac{\partial Q}{\partial y} + \frac{\partial R}{\partial z},$

Formula for curl $\mathbf{F}$ in Cartesian Coordinates	$\text{curl}\,\mathbf{F} = \left(\frac{\partial R}{\partial y} - \frac{\partial Q}{\partial z}\right)\hat{i} + \left(\frac{\partial P}{\partial z} - \frac{\partial R}{\partial x}\right)\hat{j} + \left(\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y}\right)\hat{k},$

Formula for grad $\mathbf{F}$ in Cartesian Coordinates	$\text{grad}\, f = \frac{\partial f}{\partial x}\hat{i} + \frac{\partial f}{\partial y}\hat{j} + \frac{\partial f}{\partial z}\hat{k}.$

In keeping with Einstein summation notation vector basis elements are written with lower indices and vector components are written with upper indices. Covector (differential form) basis elements are written with upper indices and covector (differential form) components are written with lower indices. This allows us to give the musical isomorphisms in terms of

Cartesian coordinates. The $\flat$ isomorphism is given by

$$\begin{aligned} \flat : T_pM &\longrightarrow T_p^*M \\ v^i \frac{\partial}{\partial x^i} &\longmapsto v_i dx^i \qquad \text{where } v_i = v^i, \end{aligned}$$

while the $\sharp$ isomorphism is given by

$$\begin{aligned} \sharp : T_p^*M &\longrightarrow T_pM \\ \alpha_i dx^i &\longmapsto \alpha^i \frac{\partial}{\partial x^i} \qquad \text{where } \alpha^i = \alpha_i . \end{aligned}$$

Keep in mind that these are not actually the mathematical definitions of the musical isomorphism but instead the formula for them in Cartesian coordinates. But these isomorphisms, along with the Hodge star operator, provide the link between vector calculus and differential forms. The following diagram commutes.

$$\begin{tikzcd} C\left(\mathbb{R}^3\right) \arrow[r, "\nabla"above, "\text{grad}"] \arrow[d, "\text{id}"] & T\mathbb{R}^3 \arrow[r, "\nabla\times"above, "\text{curl}"] \arrow[d, "\flat"] & T\mathbb{R}^3 \arrow[r, "\nabla\cdot"above, "\text{div}"] \arrow[d, "*\circ\flat"] & C\left(\mathbb{R}^3\right) \arrow[d, "*"] \\ \bigwedge^0\left(\mathbb{R}^3\right) \arrow[r, "d"] & \bigwedge^1\left(\mathbb{R}^3\right) \arrow[r, "d"] & \bigwedge^2\left(\mathbb{R}^3\right) \arrow[r, "d"] & \bigwedge^3\left(\mathbb{R}^3\right) \end{tikzcd}$$

This allows the following identities from vector calculus to be written in terms of the exterior derivative,

$$\begin{array}{ccc} \nabla \times (\nabla f) = 0 & \qquad & \nabla \cdot (\nabla \times \mathbf{F}) = 0 \\ \Updownarrow & & \Updownarrow \\ d(df) = 0 & & d(d\alpha) = 0. \end{array}$$

Notice that $d(df) = 0$ and $d(d\alpha) = 0$ of course just follow from the Poincaré lemma. It also allows the three major theorems from vector calculus to be written in terms of differential forms. All three of these theorems are in fact special cases of the differential forms version of Stokes' theorem,

$$\begin{array}{ccc} \underline{\text{Fund. Thm. Line Integrals}} & \underline{\text{Stokes' Theorem}} & \underline{\text{Divergence Theorem}} \\ f\big(c(b)\big) - f\big(c(a)\big) = \int_C \nabla f \cdot d\mathbf{s} & \int_{\partial S} \mathbf{F} \cdot d\mathbf{s} = \int_S (\nabla \times \mathbf{F}) \cdot d\mathbf{S} & \int_{\partial V} \mathbf{F} \cdot d\mathbf{S} = \int_V \nabla \cdot \mathbf{F}\, dV \\ \Downarrow & \Downarrow & \Downarrow \\ \int_{\partial C} \alpha = \int_C d\alpha & \int_{\partial S} \alpha = \int_S d\alpha & \int_{\partial V} \alpha = \int_V d\alpha. \end{array}$$

9.6.2 *References and Further Reading*

The geometric approach to divergence, curl, and gradient presented in the first half of this chapter is largely inspired by and follows Schey [38] who probably gives the nicest and most intuitively understandable geometrical introductions to these vector calculus operators available. In particular, refer to this book for a more detailed derivation of the divergence, curl, gradient, and Laplacian in both cylindrical and spherical coordinates. The relationship between the divergence, curl, and gradient operators and exterior differentiation is presented almost everywhere, but see in particular Abraham, Marsden, and Ratiu [1], Hubbard and Hubbard [27], or Edwards [18]. The musical isomorphism, sharp and flat, are introduced, for example, in Abraham, Marsden, and Ratiu [1] and in Marsden and Ratiu [32] as well as, very briefly, in Burke [8] and Darling [12].

9.6.3 Problems

$$f_1(x,y,z) = 5xy^2 - 4x^3y \qquad f_5(x,y,z) = x^2y^3z^4 \qquad f_9(x,y,z) = x^2y^3z - 3xyz^2$$
$$f_2(x,y,z) = y\ln(z) \qquad f_6(x,y,z) = \sqrt{x+yz} \qquad f_{10}(x,y,z) = x^2 + y^2 - 5z^2$$
$$f_3(x,y,z) = 1 + 2x\sqrt{y} \qquad f_7(x,y,z) = xe^y + ye^z + ze^x \qquad f_{11}(x,y,z) = -x^2y - y^3z^2 + z^4x^5$$
$$f_4(x,y,z) = x^4 - x^2y^3 \qquad f_8(x,y,z) = xe^{yz} \qquad f_{12}(x,y,z) = 3x - 4y + 5z$$

Question 9.12 For the functions listed above find grad f_i. Then find $(\text{grad} f_i)^\flat$.

Question 9.13 For the functions listed above find df_i. Compare with $(\text{grad } f_i)^\flat$ from Question 9.12.

Question 9.14 For the functions listed above find curl(grad f_i). Then find $d(df)$. Compare.

$$\mathbf{F}_1 = 2x^2\hat{i} + (x+y)\hat{j} \qquad \mathbf{F}_5 = x^2\hat{i} + xy\hat{j} - (x^2+y^2+z^2)\hat{k} \qquad \mathbf{F}_9 = e^{xy}\hat{i} - e^{yz}\hat{j} + e^{xz}\hat{k}$$
$$\mathbf{F}_2 = -x\hat{i} + (x-2y)\hat{j} \qquad \mathbf{F}_6 = yz^2\hat{i} + yz\hat{j} + (2x-2y)\hat{k} \qquad \mathbf{F}_{10} = \sqrt{xyz}\hat{i} + \sqrt{xy}\hat{j} + \sqrt{x}\hat{k}$$
$$\mathbf{F}_3 = 2x^2\hat{i} - 3y^3\hat{j} \qquad \mathbf{F}_7 = y\hat{i} + z\hat{j} + x\hat{k} \qquad \mathbf{F}_{11} = xy^2z\hat{i} - 2x^2y^3z\hat{j} + 3xy^3z^2\hat{k}$$
$$\mathbf{F}_4 = 3\hat{j} + 4\hat{k} \qquad \mathbf{F}_8 = -x^2\hat{i} + y^3\hat{j} - z^4\hat{k} \qquad \mathbf{F}_{12} = y^2\hat{i} + (e^x + e^y)\hat{j} - (x^3 - 2y)\hat{k}$$

Question 9.15 For the vectors listed above, find curl $\mathbf{F}_i$. Then find $(* \circ \flat)(\text{curl } \mathbf{F}_i)$.

Question 9.16 For the vectors listed above find $\mathbf{F}_i^\flat$. Then find $d(\mathbf{F}_i^\flat)$. Compare with $(* \circ \flat)(\text{curl } F_i)$ from Question 9.15.

Question 9.17 For the vectors listed above find div(curl $\mathbf{F}_i$). Then find $d(d\mathbf{F}_i^\flat)$. Compare.

$$\mathbf{G}_1 = -x^2y\hat{i} - xy^2\hat{j} \qquad \mathbf{G}_5 = \sqrt{xy}\hat{i} + xy\hat{j} + (3x+2y+z)\hat{k} \qquad \mathbf{G}_9 = e^{xy}\hat{i} - e^{yz}\hat{j} + e^{xz}\hat{k}$$
$$\mathbf{G}_2 = (x+2y)\hat{i} + 2xy\hat{j} \qquad \mathbf{G}_6 = yz^2\hat{i} + yz\hat{j} - (x^2+y^2+z^2)\hat{k} \qquad \mathbf{G}_{10} = x^3yz\hat{i} + x^2y\hat{j} + x\hat{k}$$
$$\mathbf{G}_3 = x^2y^2\hat{i} - 3x^2y^2\hat{j} \qquad \mathbf{G}_7 = \sqrt{yz}\hat{i} + \sqrt{xz}\hat{j} + \sqrt{xy}\hat{k} \qquad \mathbf{G}_{11} = 3xyz\hat{i} + \sqrt{xyz}\hat{j} + (4y-z)\hat{k}$$
$$\mathbf{G}_4 = 3x\hat{j} + 4y\hat{k} \qquad \mathbf{G}_8 = e^x\hat{i} + e^y\hat{j} - e^z\hat{k} \qquad \mathbf{G}_{12} = (x^2+y^2+z^2)\hat{i} + 3x\hat{j} - e^{xyz}\hat{k}$$

Question 9.18 For the vectors listed above find div$(\mathbf{G}_i)$. Then find $*(\text{div } \mathbf{G}_i)$.

Question 9.19 For the vectors listed above find $(* \circ \flat)\mathbf{G}_i$. Then find $d\big((* \circ \flat)\mathbf{G}_i\big)$. Compare with $*(\text{div } \mathbf{G}_i)$ from Question 9.18.

Question 9.20 Find a function f such that $\nabla f = xy^2\hat{i} + x^2y\hat{j}$ and then use that function, along with the Fundamental Theorem of Line Integrals, to evaluate $\int_C \nabla f \cdot ds$ along the curve C given by $\gamma(t) = \left(t + \sin\left(\frac{1}{2}\pi t\right), t + \cos\left(\frac{1}{2}\pi t\right)\right)$, $0 \le t \le 1$. Write the integral in terms of differential forms.

Question 9.21 Find a function f such that $\nabla f = yz\hat{i} + xz\hat{j} + (xy + 2z)\hat{k}$ and then use that function, along with the Fundamental Theorem of Line Integrals, to evaluate $\int_C \nabla f \cdot ds$ along the line segment from $(1, 0, -2)$ to $(4, 6, 3)$.

Question 9.22 Find a function f such that $\nabla f = (2xz + y^2)\hat{i} + 2xy\hat{j} + (x^2 + 3z^2)\hat{k}$ and then use that function, along with the Fundamental Theorem of Line Integrals, to evaluate $\int_C \nabla f \cdot ds$ along the curve C given by $\gamma(t) = (t^2, t+1, 2t-1)$, $0 \le t \le 1$. Write the integral in terms of differential forms.

Question 9.23 Find a function f such that $\nabla f = y^2\cos(z)\hat{i} + 2xy\cos(z)\hat{j} - xy^2\sin(z)\hat{k}$ and then use that function, along with the Fundamental Theorem of Line Integrals, to evaluate $\int_C \nabla f \cdot ds$ along the curve C given by $\gamma(t) = (t^2, \sin(t), t)$, $0 \le t \le \pi$. Write the integral in terms of differential forms.

Question 9.24 Use Stokes' theorem to evaluate $\int_S (\nabla \times \mathbf{F}) \cdot dS$ where $\mathbf{F} = 2y\cos(z)\hat{\imath} + e^x \sin(z)\hat{\jmath} + xe^y\hat{k}$ and S is the portion of the sphere $x^2 + y^2 + z^2 = 16$, $z \geq 0$, oriented upwards. Then write the integral in terms of differential forms.

Question 9.25 Use Stokes' theorem to evaluate $\int_S (\nabla \times \mathbf{F}) \cdot dS$ where $\mathbf{F} = e^{xy}\cos(z)\hat{\imath} + x^2z\hat{\jmath} + xy\hat{k}$ and S is the hemisphere $x = \sqrt{1 - y^2 - z^2}$, oriented in the direction of the positive x-axis. Then write the integral in terms of differential forms.

Question 9.26 Use Stokes' theorem to evaluate $\int_C \mathbf{F} \cdot ds$ where $\mathbf{F} = yz\hat{\imath} + 2xz\hat{\jmath} + e^{xy}\hat{k}$ and C is the circle $x^2 + y^2 = 9$ in the plane $z = 4$. C is oriented counterclockwise when viewed from above. Then write the integral in terms of differential forms.

Question 9.27 Use the Divergence theorem to calculate the surface integral $\int_S \mathbf{F} \cdot dS$ where $\mathbf{F} = x^2z^3\hat{\imath} + 2xyz^3\hat{\jmath} + xz^4\hat{k}$ and S is the surface of the box with vertices $(\pm 3, \pm 2, \pm 2)$. Then write the integral in terms of differential forms.

Question 9.28 Use the Divergence theorem to calculate the surface integral $\int_S \mathbf{F} \cdot dS$ where $\mathbf{F} = e^x \sin(y)\hat{\imath} + e^x \cos(y)\hat{\jmath} + yz^2\hat{k}$ and S is the surface of the box with bounded by the planes $x = 0$, $x = 1$, $y = 0$, $y = 1$, $z = 0$, $z = 1$. Then write the integral in terms of differential forms.

Question 9.29 Use the Divergence theorem to calculate the surface integral $\int_S \mathbf{F} \cdot dS$ where $\mathbf{F} = 3xy^2\hat{\imath} + xe^z\hat{\jmath} + z^3\hat{k}$ and S is the surface of the solid bounded by the planes $x = -1$, $x = 1$, and the cylinder $y^2 + z^2 = 4$. Then write the integral in terms of differential forms.

Chapter 10
Manifolds and Forms on Manifolds

In this chapter we reintroduce manifolds in a somewhat more mathematically rigorous manor while simultaneously trying not to overwhelm you with details. As always we will still place an emphasis on conceptual understanding and the big picture. Manifold theory is a vast and rich subject and there are numerous books that present manifolds in a completely rigorous manor with all the gory technical details made explicit. When you understand this chapter you will be prepared to tackle these texts.

Section one reintroduces manifolds while section two reintroduces the idea of tangent spaces. Section three gives an abstract presentation of push-forwards of vectors and pull-backs of forms. We will then discuss differential forms and both differentiation and integration on manifolds in section four. Even though general manifolds are a more abstract setting than $\mathbb{R}^n$, you should not be surprised that most of the conceptual ideas you have already become familiar with carry through essentially unchanged. In that sense we are simply repeating what we have already done, only in a more abstract setting.

10.1 Definition of a Manifold

In the beginning of this book we made a big point to make a distinction between the manifold $\mathbb{R}^n$ and the vector space $\mathbb{R}^n$. The tangent space at each point p of the manifold $\mathbb{R}^n$, that is the space $T_p(\mathbb{R}^n)$, was equivalent to the vector space $\mathbb{R}^n$. We told you that we made this distinction because in general most manifolds are not vector spaces at all. Most manifolds lack a lot of the important properties that Euclidian spaces, the manifolds we have dealt with throughout calculus, have. We begin this section by trying to explain this in more detail.

As you already know, manifolds are spaces that you can generally think of as being locally Euclidian; meaning that if you zoom in on a single point of the manifold, a very small neighborhood of that point essentially looks like the manifold $\mathbb{R}^n$ for some n, see Figs. 2.8 through 2.10. Because of this fact one can actually do calculus on manifolds just as one does calculus on the manifold $\mathbb{R}^n$.

However, there are some distinct differences between manifolds and the Euclidian spaces $\mathbb{R}^n$. First, in the Euclidian space $\mathbb{R}^n$ there is a natural way in which tangent vectors can be **parallel transported** to another point. For example when we transport the vector $\begin{bmatrix} 2 \\ 1 \end{bmatrix}_p \in T_p\mathbb{R}^2$ to $\begin{bmatrix} 2 \\ 1 \end{bmatrix}_q \in T_q\mathbb{R}^2$, where $p \neq q$, these two vectors are, in an intuitively obvious way, parallel to each other. See Fig. 10.1. In fact, this is so obvious that we rarely think about what underlying and unstated assumptions we need to make in order to know that two vectors at different points are parallel to each other.

As we have discussed before, S^2 is a manifold since a small neighborhood of every point of S^2 looks like $\mathbb{R}^2$. However, consider S^2 as shown in Fig. 10.2 with the point p being the "north pole" and the point q being on the "equator." Now consider $v_p \in T_pS^2$. What vector in T_qS^2 could we consider as parallel in some sense to v_p? That is, what is the parallel transport of v_p to the point q? It certainly is not obvious from Fig. 10.2. It turns out that some additional structure on the manifold is needed in order to find the parallel transports of vectors on manifolds. This additional structure is called a **connection**. It is essentially a way to "connect" nearby tangent spaces (whatever "nearby" means!) Connections are beyond the scope of this book, but they play an essential role in manifold theory. On a general manifold we have to specify a connection explicitly, while on a Euclidian space manifold a connection is already implicitly built into our intuitive understanding of Euclidian

J. P. Fortney, *A Visual Introduction to Differential Forms and Calculus on Manifolds*,
https://doi.org/10.1007/978-3-319-96992-3_10

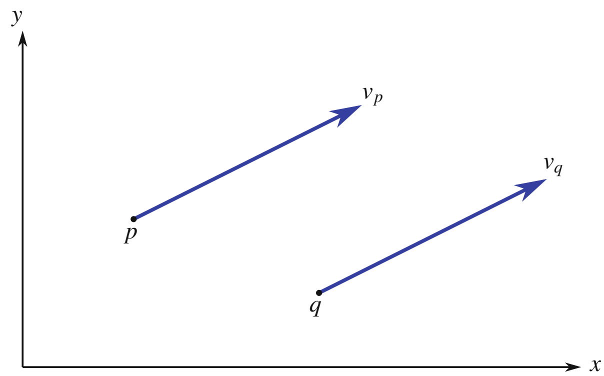


Fig. 10.1 The vector $2e_1 + e_2$ drawn at two different points p and q in the manifold $\mathbb{R}^2$. These two vectors are parallel to each other in an intuitively obvious way

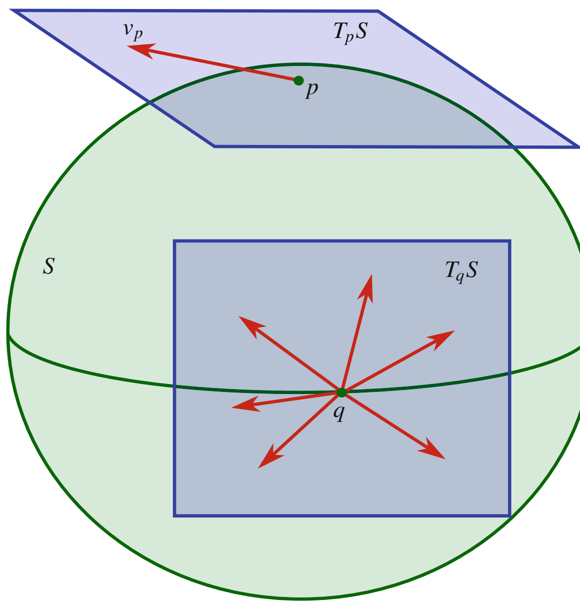


Fig. 10.2 The vector v_p where p is the "north pole." Which of the vectors at q on the "equator" are parallel to v_p?

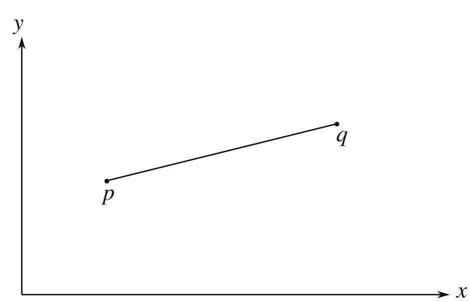


Fig. 10.3 A straight line between two points p and q in the manifold $\mathbb{R}^2$. What straight means in Euclidian space is naturally intuitive

space. This major difference between general manifolds and Euclidian space manifolds $\mathbb{R}^n$ is one of the reasons we have taken such care with the distinction between the two.

Now consider the points p and q in $\mathbb{R}^2$ as shown in Fig. 10.3. There is a very natural and intuitive idea of what a straight line between the two points would look like. And, given the Cartesian coordinate system we know how to measure the distance between the two points. We also know that the torus T^2 is a manifold since a very small neighborhood of every point of T^2 looks like $\mathbb{R}^2$. However, what would a "straight line" between the points p and q on a torus look like? What

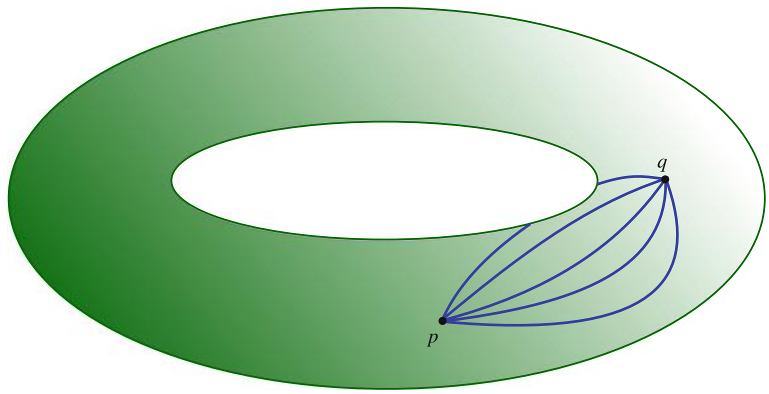


Fig. 10.4 The torus T^2 shown with two points p and q on it. Several lines connecting these points are shown. What would a "straight line" connecting two points on the manifold T^2 look like? What does "straight" even mean in this situation?

does "straight" even mean in this situation? It is not obvious from Fig. 10.4. And for that matter, how would one measure the distance between the two points on a torus? To answer this question is that in addition to a connection you would need a structure on the manifold called a **metric**. We will actually learn a little more about metrics later, though we will not go into much depth. In summary, on a general manifold we have to specify a metric explicitly, while on a Euclidian space manifold a metric, called the Euclidian metric, is already implicitly built into our intuitive understanding of Euclidian space. This is another major difference between general manifolds and our familiar Euclidian space manifolds $\mathbb{R}^n$ and is yet another reason we have to be careful with the distinction between the two.

All of this highlights the real difference between general manifolds and the Euclidian space manifolds $\mathbb{R}^n$ that you are used to. However, in everything we have done we have gone to some effort to make a distinction between the "manifold" $\mathbb{R}^n$ and the tangent spaces of the manifold, which are basically the vector spaces $\mathbb{R}^n$. Since you are already used to this moving into the general case with manifolds should hopefully not be too difficult.

Now we are ready to give a rigorous definition for what a differentiable manifold is. Here we will stick with what are called differentiable manifolds. In reality there are a lot of different types of manifolds that have different properties. Since this is not a class on manifold theory we will stick with the simplest kind. General manifolds retain some of the structure of the Euclidian space manifolds, but will not retain either the connection or the Euclidian metric from the Euclidian space manifolds.

An n-dimensional **manifold** is a space M that can be completely covered by a collection of **local coordinate neighborhoods** U_i with one-to-one mappings $\phi_i : U_i \to \mathbb{R}^n$, which are called a **coordinate maps**. Together U_i and ϕ_i are called **coordinate patch** or a **chart**, which is generally denoted as (U_i, ϕ_i). The set of all the charts together, $\{(U_i, \phi_i)\}$, is called a **coordinate system** or an **atlas** of M. Since the U_i cover all of M we write that $M = \bigcup U_i$. Also, since ϕ_i is one-to-one it is invertible, so ϕ_i^{-1} exists and is well defined. See Fig. 10.5 for a two-dimensional example. The terms atlas and chart are particularly illustrative; what is an atlas of the world but a collection of individual charts? Here the word chart is a nautical word for map. Finally, if two charts have a non-empty intersection, $U_i \cap U_j \neq \varnothing$, then the functions $\phi_j \circ \phi_i^{-1} : \mathbb{R}^n \to \mathbb{R}^n$ are called **transition functions**. There is a lot of important terminology here so we summarize it:

$$
\begin{aligned}
U_i &\quad : \quad \text{coordinate neighborhood,} \\
\phi_i : U_i \to \mathbb{R}^n &\quad : \quad \text{coordinate map,} \\
(U_i, \phi_i) &\quad : \quad \text{coordinate patch/chart,} \\
\big\{(U_i, \phi_i)\big\} &\quad : \quad \text{coordinate system/atlas,} \\
\phi_j \circ \phi_i^{-1} : \mathbb{R}^n \to \mathbb{R}^n &\quad : \quad \text{transition function.}
\end{aligned}
$$

Now we consider what happens to a point $r \in U_i \cap U_i \neq \emptyset$. We have $\phi_i(r) \in \mathbb{R}^n$ and $\phi_j(r) \in \mathbb{R}^n$. Furthermore, we have $\phi_j \circ \phi_i^{-1}$ sending $\phi_i(U_i \cap U_j) \subset \mathbb{R}^n$ to $\phi_j(U_i \cap U_j) \subset \mathbb{R}^n$, that is, $\phi_j \circ \phi_i^{-1}$ is a map of a subset of $\mathbb{R}^n$ to another subset of $\mathbb{R}^n$, and so the mapping $\phi_j \circ \phi_i^{-1} : \mathbb{R}^n \to \mathbb{R}^n$ with domain $\phi_i(U_i \cap U_j)$ and range $\phi_j(U_i \cap U_j)$ is the sort of mapping that we know how to differentiate from multivariable calculus. A **differentiable manifold** is a set M, together with a collection of charts (U_i, ϕ_i), where $M = \bigcup U_i$, such that every mapping $\phi_j \circ \phi_i^{-1}$, where $U_i \cap U_j \neq \emptyset$, is differentiable.

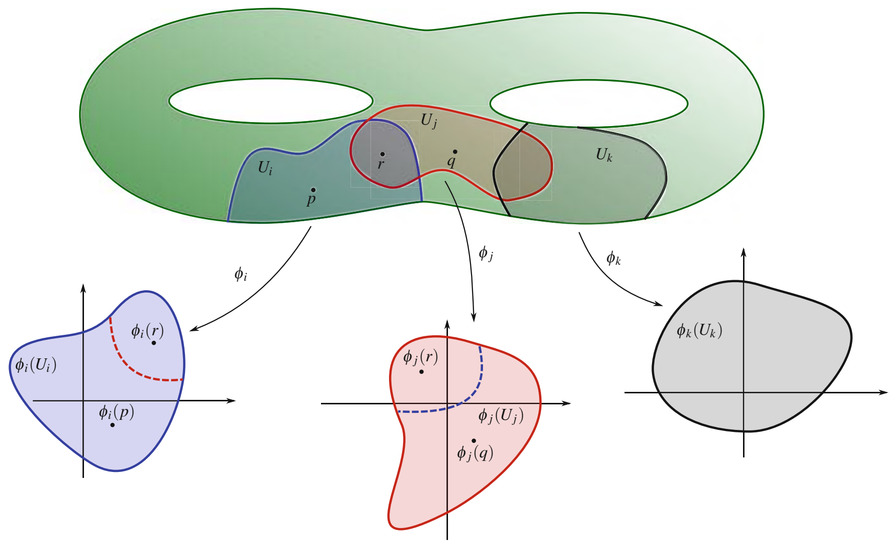


Fig. 10.5 Here a two-dimensional manifold, the double torus, is shown along with three chats (U_i, ϕ_i), (U_j, ϕ_j), and (U_k, ϕ_k). Notice the point $r \in U_i \cap U_i$. Clearly $\phi_j \circ \phi_i^{-1}$ sends $\phi_i(U_i \cap U_j) \subset \mathbb{R}^2$ to $\phi_j(U_i \cap U_j) \subset \mathbb{R}^2$

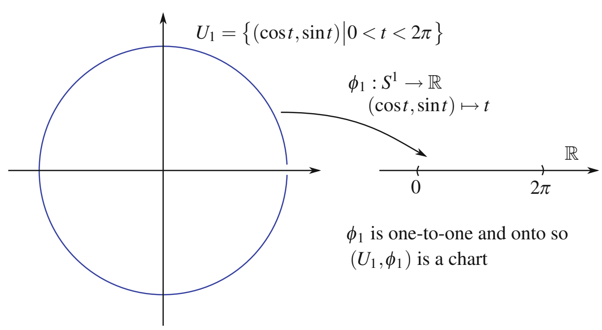


Fig. 10.6 The chart (U_1, ϕ_1) on S^1. U_1 is the subset of S^1 that consists of the points $(\cos t, \sin t)$ for $0 < t < 2\pi$ and $\phi_1 : U_1 \to \mathbb{R}$ is defined by $(\cos t, \sin t) \mapsto t$

Let's see how this works with an simple explicit example. Consider the unit circle S^1 in $\mathbb{R}^2$. We will show that S^1 is indeed a differentiable manifold. Let U_1 be the subset of S^1 that consists of the points $(\cos t, \sin t)$ for $0 < t < 2\pi$. Then $\phi_1 : U_1 \to \mathbb{R}$, defined by $(\cos t, \sin t) \mapsto t$, is a one-to-one mapping onto the open interval $(0, 2\pi) \subset \mathbb{R}$, so (U_1, ϕ_1) is a chart on S^1. See Fig. 10.6.

Similarly, let U_2 be the subset of S^1 that consists of the points $(\cos\tau, \sin\tau)$ for $-\pi < \tau < \pi$. Then $\phi_2 : U_2 \to \mathbb{R}$, defined by $(\cos\tau, \sin\tau) \mapsto \tau$, is a one-to-one mapping onto a subset of $\mathbb{R}$, so (U_2, ϕ_2) is another chart on S^1. See Fig. 10.7. The domains U_1 and U_2 cover all of S^1, that is, $S^1 = U_1 \cup U_2$. We say that (U_1, ϕ_1) and (U_2, ϕ_2) are an atlas for S^1.

Question 10.1 What point of S^1 is not included in U_1? What point of S^1 is not included in U_2? Why were these points not included? Could they have been included? Even if they had been included, what problem still arises?

Notice $U_1 \cap U_2$ has two components, the top open half of S^1 and the bottom open half of S^1. See Fig. 10.8. On the top open half we have $\phi_1^{-1} \circ \phi_2(t) = t$ for $0 < t < \pi$, which is differentiable, and on the bottom half we have $\phi_1^{-1} \circ \phi_2(t) = t - 2\pi$ for $\pi < t < 2\pi$, which is also differentiable. Hence S^1 is a differentiable manifold.

At this point in most books on differential geometry one usually encounters the definitions for both orientated manifolds and manifolds with boundary. Based on Sect. 1.2 you already know that volumes come with a sign. The idea of an oriented manifold is related to this, it is essentially when a manifold comes with a way to determine when a volume is positive and when it is negative. We will discuss oriented manifolds in Sect. 10.4 when we discuss integration on manifolds. Manifolds with a boundary are essentially what you would think them to be, manifolds that have some sort of a boundary or "edge" to them. What a boundary or "edge" actually is can be defined very precisely in terms of coordinate charts. In this book we will

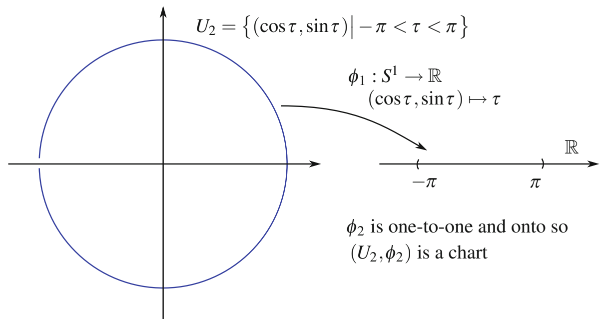


Fig. 10.7 The chart (U_2, ϕ_2) on S^1. U_2 is the subset of S^1 that consists of the points $(\cos\tau, \sin\tau)$ for $-\pi < \tau < \pi$ and $\phi_2 : U_2 \to \mathbb{R}$ is defined by $(\cos\tau, \sin\tau) \mapsto \tau$

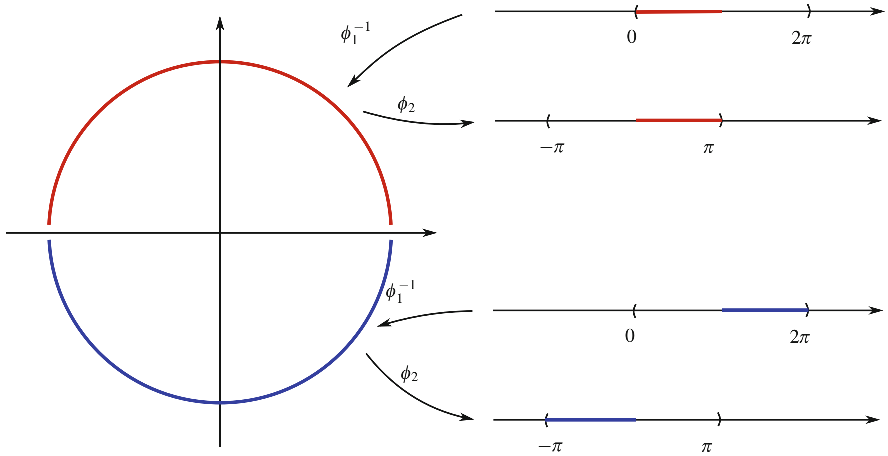


Fig. 10.8 $U_1 \cap U_2$ has two components, the top open half of S^1 (red) and the bottom open half of S^1 (blue). On the top open half we have $\phi_1^{-1} \circ \phi_2(t) = t$ for $0 < t < \pi$ and on the bottom half we have $\phi_1^{-1} \circ \phi_2(t) = t - 2\pi$ for $\pi < t < 2\pi$. Both of these are differentiable making S^1 a differentiable manifold

not discuss manifolds with boundary explicitly, though we will explore the concepts of boundary more closely in the specific context of Stoke's theorem in Chap. 11.

10.2 Tangent Space of a Manifold

Recall that when we introduced the tangent space before we did so by considering the manifold $\mathbb{R}^n$. The examples we considered had $n = 2$ or 3, but the idea generalized to any n. See Figs. 2.14 through 2.17. We chose a point $p \in \mathbb{R}^n$ and then considered all the vectors v_p that originated at that point. We called the set of all vectors v_p originating at the point p the tangent space of $\mathbb{R}^n$ at p and denoted it as $T_p\mathbb{R}^n$. We then saw that $T_p\mathbb{R}^n$ was essentially the same space as the vector space $\mathbb{R}^n$. In this section we want to give a somewhat more abstract presentation of tangent spaces.

For example, given a general manifold M you may wonder what we mean by vectors emanating from a point $p \in M$. After all, $\mathbb{R}^n$ is clearly also a vector space in addition to being a differentiable manifold, which allow us to talk about vectors, whereas a general differential manifold need not be a vector space at all meaning that there are not any vectors in the manifold. Consider the sphere $S^2 \subset \mathbb{R}^3$ as shown in Fig. 10.9. We can draw all sorts of vectors at the point $p \in S^2$ which are tangent to S^2, but these vectors are actually in the space $\mathbb{R}^3$, which is the space that the manifold S^2 is embedded in, and not in the manifold S^2 itself.

When we originally introduced tangent spaces we illustrated the general idea of the tangent space to a manifold by drawing a few manifolds and saying the tangent space at a point p was in essence the tangent plane (or hyperplane) to the manifold at that point. For example, the manifold S^2 along with a few of its tangent planes is shown in Fig. 10.10.

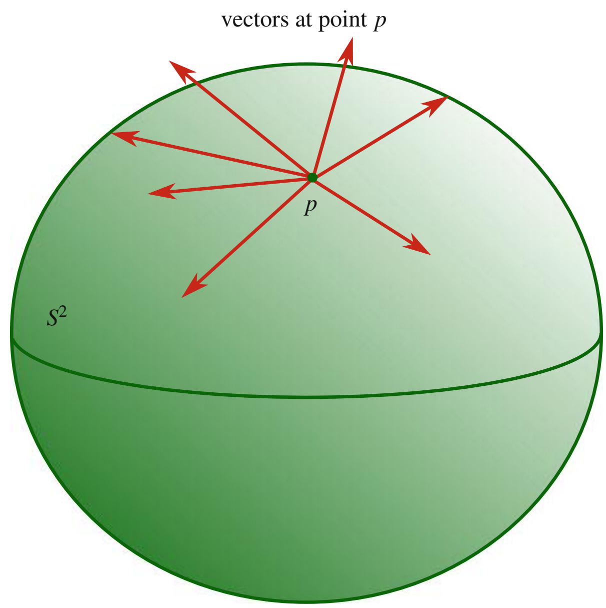


Fig. 10.9 The manifold S^2 along with a number of vectors at the point p tangent to S^2. However, these vectors are not actually in S^2. They are in fact in $\mathbb{R}^3$, the space in which we embed S^2 in order to draw it

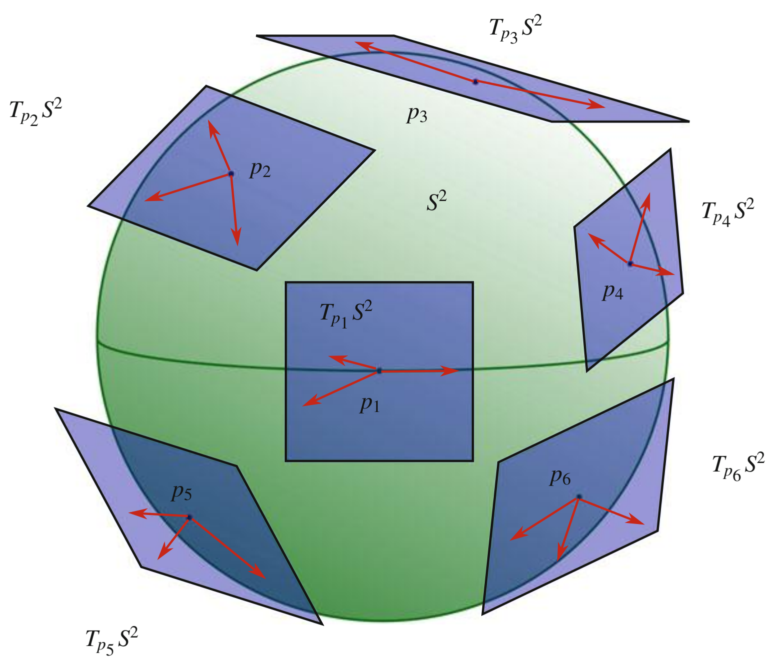


Fig. 10.10 When we introduced the concept of tangent spaces to the manifold S^2 we drew tangent planes to the manifold at various point on the manifold. Every point of S^2 actually has such a tangent plane

This is a perfectly good cartoon picture to help us visualize and understand tangent spaces, and we will continue using it, but it requires that our manifold be embedded in some $\mathbb{R}^n$ in which the tangent space is also embedded, just like both S^2 and the tangent spaces of S^2 are embedded in $\mathbb{R}^3$. According to a deep theorem in differential geometry called the Whitney embedding theorem any reasonably nice manifold can be embedded into $\mathbb{R}^n$ for some sufficiently large n, which means that we can actually continue to use our cartoon picture to help us visualize and think about tangent spaces. However, we essentially have to "step outside" of the manifold in order to visualize tangent spaces this way. A property that requires us to "step outside" of the manifold in order to think about it or define it is called an **extrinsic** property.

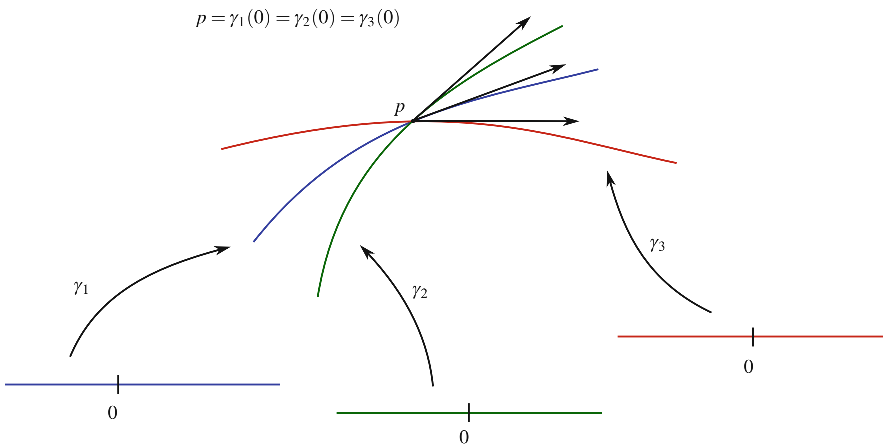


Fig. 10.11 Three curves $\gamma_i : (-\epsilon, \epsilon) \subset \mathbb{R} \longrightarrow M$, $i = 1, 2, 3$, such that $\gamma_i(0) = p$. Each curve has a tangent vector at the point p

We would like to be able to define and think about tangent spaces without "stepping outside" the manifold. There are many cases where being able to think about manifold properties while still "being inside" the manifold is useful. A property of a manifold that we can define or think about while still "being inside" the manifold is call an **intrinsic** property. An intrinsic property does not, in any way, require or rely on the manifold being embedded into some $\mathbb{R}^n$. Even though we have been thinking of tangent vectors and tangent spaces extrinsically up to now they are, in fact, intrinsic properties. We can define them and think about them while still "being inside" the manifold. When we do this the idea of arrows emanating from a point is no longer appropriate.

We will approach the intrinsic definition of tangent vectors by first considering all **smooth curves**

$$\gamma : (-\epsilon, \epsilon) \subset \mathbb{R} \longrightarrow M$$

such that $\gamma(0) = p \in M$ and ϵ is just some small positive number. In Fig. 10.11 we draw several smooth curves $\gamma_i : (-\epsilon, \epsilon) \to M$, where $\gamma_i(0) = p$ for all i. We will call the parameter that γ_i depends on time. Without getting into the technical meaning of a smooth curve just assume that it means what you intuitively think it means, that there are no sharp corners on the curve.

Each tangent vector v_p is actually tangent to some smooth curve that goes through the point $p \in M$. But while the tangent vector v_p may actually only exist in the space $\mathbb{R}^n$ in which M is embedded, the curves γ exist entirely in the manifold M. In essence each of these curves defines a different tangent vector at the point p. Thus the general idea is to identify tangent vectors to the manifold M at a point p with curves in the manifold that go through the point p, thereby eliminating our need to embed M into some larger space $\mathbb{R}^n$ in order to have tangent vectors.

It may seem possible that the tangent space of M is equivalent to the set of curves $\gamma : (-\epsilon, \epsilon) \to M$ such that $\gamma(0) = p$. This is almost true, but not quite. There are two issues we have to consider carefully. To address the first issue consider Fig. 10.12. Here we have three different curves $\gamma_1, \gamma_2, \gamma_3 : (-\epsilon, \epsilon) \to M$ with $\gamma_1(0) = \gamma_2(0) = \gamma_3(0) = p$ that are identical very close to p. If two curves are identical very close to p then can we in some sense consider them the same curve? At a first glance it would seem reasonable to do this.

To consider the second issue, consider the vectors v_p and $2v_p$ in T_pM. From Fig. 10.13 it looks like both v_p and $2v_p$ can be represented by the same curve γ. The question is, when we use curves how would we distinguish between the vectors v_p and $2v_p$ in T_pM? They are both pointing in the same direction, but the second vector is twice as long as the first vector.

The clue to how we would handle this second issue is that we called the parameter that γ depends on time. The derivative of γ at the time $t = 0$ gives a velocity at time 0. As we increase the rate at which we move along the curve γ then the velocity is greater and so the velocity vector becomes longer. So in order to distinguish between v_p and $2v_p$ the speed of the parametrization is increased so it is twice what it originally was at $t = 0$.

Let us consider a simple example. Let $M = \mathbb{R}^2_{xy}$ be the Euclidian plane parameterized by x and y. Consider the following three vectors at the origin of $\mathbb{R}^2_{xy}$:

$$u_0 = \begin{bmatrix} 1 \\ 2 \end{bmatrix}, \quad v_0 = 2u_0 = \begin{bmatrix} 2 \\ 4 \end{bmatrix}, \quad w_0 = -u_0 = \begin{bmatrix} -1 \\ -2 \end{bmatrix}.$$

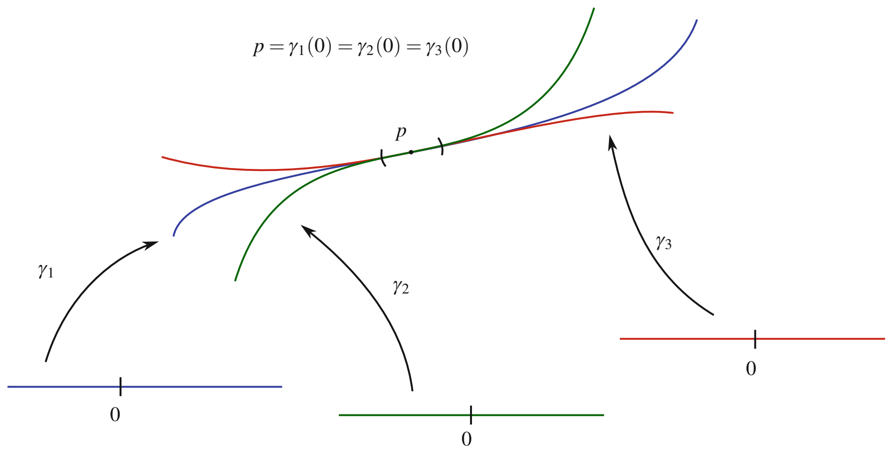


Fig. 10.12 As in Fig. 10.11 three curves $\gamma_i : (-\epsilon, \epsilon) \subset \mathbb{R} \longrightarrow M$, $i = 1, 2, 3$, such that $\gamma_i(0) = p$. However, here when we are very close to p these curves are identical

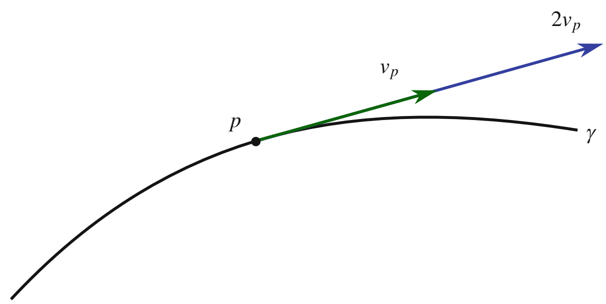


Fig. 10.13 Two vectors, v_p and $2v_p$, that both appear to be tangent to the same curve γ

See Fig. 10.14. These vectors are given by the following three curves $\alpha, \beta, \gamma : (-\epsilon, \epsilon) \to \mathbb{R}^2_{xy}$ where

$$\begin{aligned}
\alpha(t) &= (\alpha_1(t), \alpha_2(t)) = (t, 2t),\\
\beta(t) &= (\beta_1(t), \beta_2(t)) = (2t, 4t),\\
\gamma(t) &= (\gamma_1(t), \gamma_2(t)) = (-t, -2t).
\end{aligned}$$

Though drawing the curves α, β, and γ in $\mathbb{R}^2$ will give what looks like the same curve, technically they are all different curves because of their different parameterizations.

Now we will define the tangent space of M at point p, T_pM, in terms of curves on M, but we have to take the two above issues into consideration. Two curves that are identical very close to p, as in Fig. 10.12, have the same range close to p. If two curves have the same range close to p and have the same parametrization close to p they are called equivalent, which is denoted $\sim$. Suppose that γ_1, γ_2, and γ_3 all have the same range close to p and all have the same parametrization close to p then we would say $\gamma_1 \sim \gamma_2 \sim \gamma_3$. The set of all equivalent curves is called an **equivalence class**. The equivalence class of γ_1 is the set of all the curves equivalent to γ_1, is denoted by $[\gamma_1]$, and is defined by

$$[\gamma_1] \equiv \big\{\gamma \;\big|\; \gamma \sim \gamma_1\big\}.$$

Thus, if $\gamma_1 \sim \gamma_2 \sim \gamma_3$ we would have $[\gamma_1] = [\gamma_2] = [\gamma_3]$. This means that which member of the equivalence class you use to represent the equivalence class does not matter. Each equivalence class of curves at a point p is defined to be a tangent vector at p. We will write $[\gamma]_p$ when we want to indicate that the base point is p. The **tangent space** of M at p defined as the set of all tangent vectors, that is, equivalence classes of curves, at the point p,

$$T_pM = \Big\{\, [\gamma]_p \;\Big|\; \gamma : (-\epsilon, \epsilon) \to M \text{ and } \gamma(0) = p \Big\}.$$

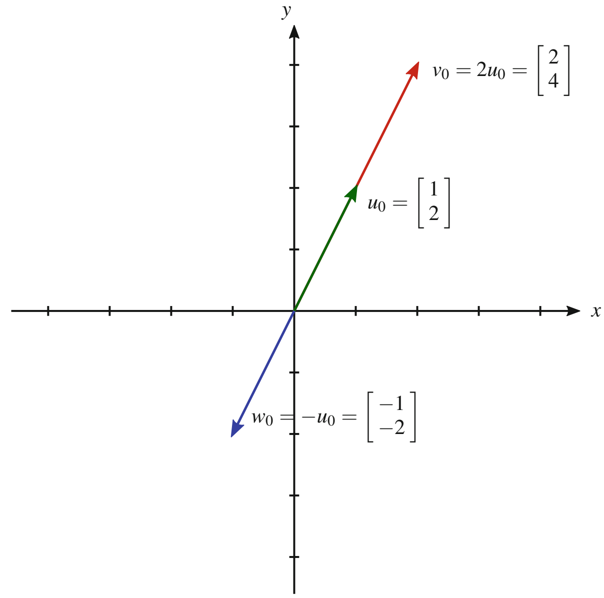


Fig. 10.14 The manifold $\mathbb{R}^2_{xy}$ with three vectors $u_0 = [1, 2]^T$, $v_0 = 2u_0 = [2, 4]^T$, and $w_0 = -u_0 = [-1, -2]^T$ at the origin. These three vectors are determined by three curves α, β, and γ with different parameterizations

Thus the tangent space to M at the point p is the set of all equivalence classes of curves that go through $p \in M$. Notice that this definition of the tangent space is intrinsic, it entirely relies on curves in the manifold M that go through the point p. Thus we do not need to "step outside" our manifold in order to visualize or think about the tangent space. This definition is independent of any bigger space that our manifold is embedded in.

Now we want to understand the relationship between our extrinsic and intrinsic views of vectors and tangent spaces. A vector v_p, which actually lies in the space our manifold is embedded in, is the velocity vector of some curve γ at the point p so we have made the identification $v_p = [\gamma]_p$ To obtain back the vector v_p from the curve γ we have to take the derivative of γ with respect to time and evaluate the derivative at time $t = 0$. Basically, we are finding the Jacobian matrix of the mapping $\gamma : (-\epsilon, \epsilon) \to M$.

Returning to our concrete example we have

$$u_{(0,0)} \equiv [\alpha]_{(0,0)} \qquad v_{(0,0)} \equiv [\beta]_{(0,0)}, \qquad w_{(0,0)} \equiv [\gamma]_{(0,0)}.$$

But notice, even though when we draw the curves on $\mathbb{R}^2$ and it turns out that they all look like "the same" curve in our picture, they are really different curves because they have different parameterizations and so are not part of the same equivalence class of curves. In other words, we have $\alpha \not\sim \beta$, $\alpha \not\sim \gamma$, and $\beta \not\sim \gamma$. Usually we simply use 0 to represent the base point $(0, 0)$ but since we want to emphasize the difference between the base point and evaluating the partial derivatives at time $t = 0$ we will be extra careful with our notation here. We can see how our two views of vectors correlate,

$$u_{(0,0)} \equiv \Big[\big(\alpha_1(t), \alpha_2(t)\big)\Big]_{(0,0)} = \big[(t, 2t)\big]_{(0,0)},$$

$$u_{(0,0)} = \begin{bmatrix} \left.\frac{\partial \alpha_1(t)}{\partial t}\right|_{t=0} \\ \left.\frac{\partial \alpha_2(t)}{\partial t}\right|_{t=0} \end{bmatrix}_{(0,0)} = \begin{bmatrix} \left.\frac{\partial (t)}{\partial t}\right|_{t=0} \\ \left.\frac{\partial (2t)}{\partial t}\right|_{t=0} \end{bmatrix}_{(0,0)} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}_{(0,0)}.$$

Similarly we have,

$$v_{(0,0)} \equiv \Big[\big(\beta_1(t), \beta_2(t)\big)\Big]_{(0,0)} = \big[(2t, 4t)\big]_{(0,0)},$$

$$v_{(0,0)} = \begin{bmatrix} \left.\frac{\partial \beta_1(t)}{\partial t}\right|_{t=0} \\ \left.\frac{\partial \beta_2(t)}{\partial t}\right|_{t=0} \end{bmatrix}_{(0,0)} = \begin{bmatrix} \left.\frac{\partial (2t)}{\partial t}\right|_{t=0} \\ \left.\frac{\partial (4t)}{\partial t}\right|_{t=0} \end{bmatrix}_{(0,0)} = \begin{bmatrix} 2 \\ 4 \end{bmatrix}_{(0,0)},$$

and

$$w_{(0,0)} \equiv \Big[\big(\gamma_1(t), \gamma_2(t)\big)\Big]_{(0,0)} = \big[(-t, -2t)\big]_{(0,0)},$$

$$w_{(0,0)} = \begin{bmatrix} \left.\frac{\partial \gamma_1(t)}{\partial t}\right|_{t=0} \\ \left.\frac{\partial \gamma_2(t)}{\partial t}\right|_{t=0} \end{bmatrix}_{(0,0)} = \begin{bmatrix} \left.\frac{\partial (-t)}{\partial t}\right|_{t=0} \\ \left.\frac{\partial (-2t)}{\partial t}\right|_{t=0} \end{bmatrix}_{(0,0)} = \begin{bmatrix} -1 \\ -2 \end{bmatrix}_{(0,0)}.$$

Below is a key for moving between the two ways we have defined vectors on a manifold. The equivalence class of curves definition is based on a curve that exists entirely inside our manifold M and does not require M to be embedded into a larger ambient space in order to "see" the vectors; our vectors are simply viewed as the curves that lie entirely in M. What I will call here the vector definition requires either embedding our manifold M into an ambient space $\mathbb{R}^n$ or drawing the tangent space to the manifold M at a point in order to "see" the vector. **In the equivalence class definition we view the curve as the vector and in the vector definition we view the derivative of the curve as the vector.** This may seem odd to you but it is simply thinking about vectors from two different perspectives. Hopefully this will not be too confusing.

Equivalence Class Definition	$\Longleftrightarrow$	Vector Definition
$v_p = [\gamma]_p, \quad p = \gamma(0)$		$v_p = [\gamma']_p = [\gamma]'_p$
$= \Big[\big(\gamma_1(t), \ldots, \gamma_n(t)\big)\Big]$		$= \Big[\big(\gamma'_1(t), \ldots, \gamma'_n(t)\big)\Big]_p$
		$= \Big[\big(\gamma_1(t), \ldots, \gamma_n(t)\big)\Big]'_p$
		$= \begin{bmatrix} \gamma'_1(t=0) \\ \vdots \\ \gamma'_n(t=0) \end{bmatrix}_p$
		$= \begin{bmatrix} \left.\frac{\partial \gamma_1(t)}{\partial t}\right\|_{t=0} \\ \vdots \\ \left.\frac{\partial \gamma_n(t)}{\partial t}\right\|_{t=0} \end{bmatrix}_p$

One more comment is in order here. As we know, integration is in a sense the opposite of differentiation. If we can view a vector v_p as the derivative of a curve γ at the point p, can we view the curve γ as the integral of the vector v_p? The problem here is that v_p exists at only one single point, and the idea of integration at a single point is ill-defined. What we need is a **smooth vector field** around a point in order to do integration. Without getting into the technical details of what a smooth vector field is, think of it as a vector field which changes smoothly as you vary the point. Once we have a smooth vector field in some neighborhood then we can talk of curves called **integral curves**. The curve γ that is used to represent the equivalence class $[\gamma]_p$ of some vector v_p need not be an integral curve, but if v_p is an element of a smooth vector field then usually people make the unstated assumption that γ is an integral curve. We will discuss integral curves in more detail in the contexts of electromagnetism and geometric mechanics in Chap. 12 and Appendix B.

Recall that in Sect. 2.3 we equated tangent vectors at a point p with directional derivatives. By doing this we identified the Euclidian basis vectors of the tangent space with partial derivative operators. In other words, we had found the following identifications,

$$e_{1_p} = \left.\frac{\partial}{\partial x}\right|_p, \quad e_{2_p} = \left.\frac{\partial}{\partial y}\right|_p, \quad e_{3_p} = \left.\frac{\partial}{\partial z}\right|_p.$$

Thus, for a vector $v_p \in T_p\mathbb{R}^3$,

$$v_p = \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix}_p$$

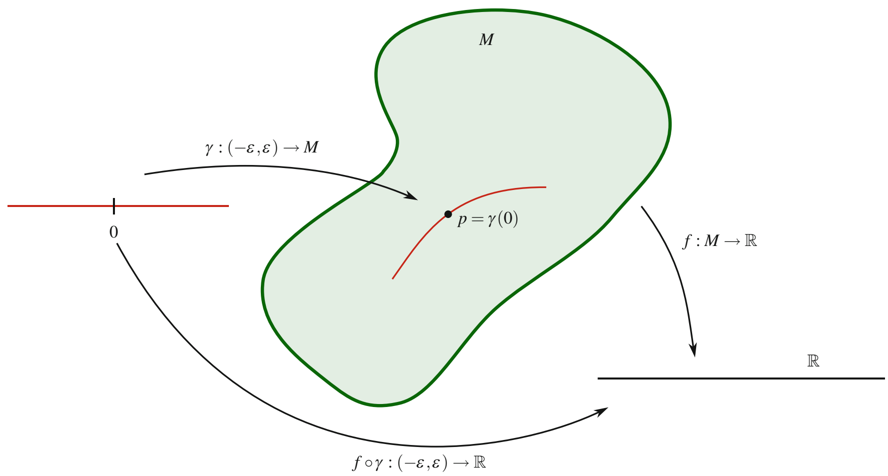


Fig. 10.15 The vector $[\gamma]_p$ given by $\gamma : (-\epsilon, \epsilon) \to M$, where $\gamma(0) = p$, and a function $f : M \to \mathbb{R}$. Then we have $f \circ \gamma : (-\epsilon, \epsilon) \subset \mathbb{R} \to \mathbb{R}$, which is the sort of mapping we know how to differentiate

$$
\begin{aligned}
&= v_1 e_{1_p} + v_2 e_{2_p} + v_3 e_{3_p} \\
&= v_1 \left.\frac{\partial}{\partial x}\right|_p + v_2 \left.\frac{\partial}{\partial y}\right|_p + v_3 \left.\frac{\partial}{\partial z}\right|_p .
\end{aligned}
$$

The directional derivative of the function $f : \mathbb{R}^3 \to \mathbb{R}$ in the direction v_p was written as the vector v_p operating on the function f,

$$
\begin{aligned}
v_p[f] &= \left(v_1 \left.\frac{\partial}{\partial x}\right|_p + v_2 \left.\frac{\partial}{\partial y}\right|_p + v_3 \left.\frac{\partial}{\partial z}\right|_p \right)[f] \\
&= v_1 \left.\frac{\partial f}{\partial x}\right|_p + v_2 \left.\frac{\partial f}{\partial y}\right|_p + v_3 \left.\frac{\partial f}{\partial z}\right|_p .
\end{aligned}
$$

We want to consider what this is in our more abstract intrinsic setting where the vector v_p is given by an equivalence class of curves $[\gamma]_p$, that is, when we have $v_p \equiv [\gamma]_p$. Before we used the notation $v_p[f]$ to denote $D_{v_p} f$, directional derivative of f in the direction of v_p. We used this notation to emphasise the fact that it is the vector v_p which is, in a sense, operating on the function f. We will continue to use this notation whenever we write the vector as v_p, but when we write the vector using our equivalence class of curves notation $[\gamma]_p$ then writing $[\gamma]_p[f]$ would be a little odd since the two sets of square brackets $[\cdot]$ have different meanings. Notation is sometimes as much the result of habit and convention as rational planning.

Consider $\gamma : (-\epsilon, \epsilon) \to M$, where $\gamma(0) = p$, and $f : M \to \mathbb{R}$. Then $f \circ \gamma : (-\epsilon, \epsilon) \subset \mathbb{R} \to \mathbb{R}$, as shown in Fig. 10.15. When our vector v_p is given by an equivalence class of curves $[\gamma]_p$ then the directional derivative of f in the v_p direction is given by

$$
\begin{aligned}
v_p[f] &= D_{[\gamma]_p} f \\
&= [f \circ \gamma]'_p \\
&= \left.\frac{\partial (f \circ \gamma)}{\partial t}\right|_{t=0} \\
&= \lim_{h \to 0} \frac{(f \circ \gamma)(h) - (f \circ \gamma)(0)}{h} .
\end{aligned}
$$

One often used notation is $D_{[\gamma]_p} f$, which mirrors the notation generally used in multi-variable calculus. Other notations for $D_{[\gamma]_p} f$ include $[(f \circ \gamma)']_p$ or $[f \circ \gamma]'_p$ or $[(f \circ \gamma)']_{t=0}$ or even $[f \circ \gamma]'_{t=0}$. There is no perfectly standard way of writing this and one can use all sorts of different permutations in the notation. The key of course is that if the vector is at point

$p = \gamma(0)$ that means that the derivative of γ has to be evaluated at $t = 0$. This is the pertinent point. Also, usually one just writes a 0. You need to pay attention to context to decide whether the 0 means to evaluate at $t = 0$ or the base point is the origin $(0, 0)$. With this warning made we will stop trying to be so precise.

Notice that the directional derivative of f at the point $p = \gamma(0)$ in the direction $[\gamma]_p$ is defined to be the same thing as the ordinary one-variable derivative of the function $f \circ \gamma$ at $t = 0$. It is simple enough to see that we arrive at the same answer with this notation. Assume $\gamma(t) = \big(x(t), y(t), z(t)\big)$, then we have

$$\begin{aligned} v_p &= \Big[\big(x(t), y(t), z(t)\big)\Big]'_p \\ &= \begin{bmatrix} x'(t=0) \\ y'(t=0) \\ z'(t=0) \end{bmatrix}_p . \end{aligned}$$

Using $f \circ \gamma(t) = f\big(x(t), y(t), z(t)\big)$ and we then get

$$\begin{aligned} v_p[f] &= D_{[\gamma]_p} f \\ &= \left.\frac{\partial (f \circ \gamma)}{\partial t}\right|_{t=0} \\ &= \left.\frac{\partial\ f\big(x(t), y(t), z(t)\big)}{\partial t}\right|_0 \\ &= \left.\frac{\partial f}{\partial x}\right|_{\gamma(0)} \cdot \left.\frac{\partial x}{\partial t}\right|_0 + \left.\frac{\partial f}{\partial y}\right|_{\gamma(0)} \cdot \left.\frac{\partial y}{\partial t}\right|_0 + \left.\frac{\partial f}{\partial z}\right|_{\gamma(0)} \cdot \left.\frac{\partial z}{\partial t}\right|_0 \qquad \text{chain rule} \\ &= \left.\frac{\partial f}{\partial x}\right|_p \underbrace{x'(0)}_{v_1} + \left.\frac{\partial f}{\partial y}\right|_p \underbrace{y'(0)}_{v_2} + \left.\frac{\partial f}{\partial z}\right|_p \underbrace{z'(0)}_{v_3}, \end{aligned}$$

which is exactly the same as what we had for $v_p[f]$ just above, with the base point p left off the notation.

Turning back to a concrete example, suppose we had the function $f : \mathbb{R}^2_{xy} \to \mathbb{R}$ given by $f(x, y) = x + \sin(y)$. We want to find $u_0[f]$, where $u_0 = \begin{bmatrix} 1 \\ 2 \end{bmatrix} = \big[(t, 2t)\big]_0$. (Note, here we have $0 = (0, 0)$.) Proceeding as we would have before we have

$$\begin{aligned} u_0[f] &= \left(1 \cdot \left.\frac{\partial}{\partial x}\right|_0 + 2 \cdot \left.\frac{\partial}{\partial y}\right|_0\right) f \\ &= 1 \cdot \left.\frac{\partial\big(x + \sin(y)\big)}{\partial x}\right|_0 + 2 \cdot \left.\frac{\partial\big(x + \sin(y)\big)}{\partial y}\right|_0 \\ &= 1 \cdot 1 + 2 \cdot \cos(0) \\ &= 3. \end{aligned}$$

However, with $u_0 = [\gamma]_0 = [(t, 2t)]_0$, we have

$$\begin{aligned} D_{u_0} f &= [f \circ \gamma]'_0 \\ &= \left.\frac{\partial (f \circ \gamma)}{\partial t}\right|_0 \\ &= \left.\frac{\partial\big(t + \sin(2t)\big)}{\partial t}\right|_0 \\ &= 1 + 2\cos(2t)\Big|_0 \\ &= 1 + 2 = 3. \end{aligned}$$

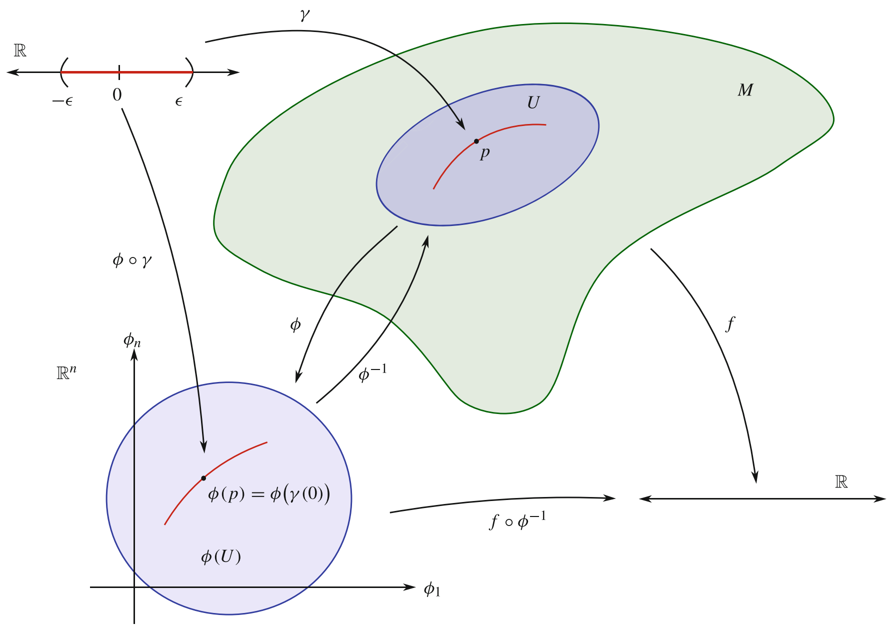


Fig. 10.16 Here the manifold M is shown with three things, (**a**) the curve γ that represents a tangent vector at the point p, (**b**) a real-valued function $f : M \to \mathbb{R}$ on the manifold M, and (**c**) a coordinate chart (U, ϕ) such that $p \in U$. It is obvious that $f \circ \gamma = f \circ \phi^{-1} \circ \phi \circ \gamma = (f \circ \phi^{-1}) \circ (\phi \circ \gamma)$. The mappings $\phi \circ \gamma : \mathbb{R} \to \mathbb{R}^n$ and $f \circ \phi^{-1} : \mathbb{R}^n \to \mathbb{R}$ are also shown

This all seems well and good, until you notice that in our example our manifold was $\mathbb{R}^2_{xy}$, that is, $\mathbb{R}^2$ with the Cartesian coordinate system "built in." Finding our composition $f \circ \gamma$ relied on knowing this coordinate system. How so? Since the function f was written explicitly in terms of the variables x and y which made writing down $f \circ \gamma$ easy. But how would we manage the directional derivatives on a manifold that does not have a nice standard coordinate system "built in"? In other words, what if it was not so straightforward to write down what $f \circ \gamma$ actually is?

If we want to find the derivative of $f : M \to \mathbb{R}$ at the point $p \in M$ in the direction of $v_p \equiv [\gamma]_p$ then we need to know how to compose f and γ to get the map $f \circ \gamma : (-\epsilon, \epsilon) \subset \mathbb{R} \to \mathbb{R}$. But for an arbitrary manifold M this may not actually be clear. The composition $f \circ \gamma$ actually relies on some coordinate chart ϕ. Thus, our composition $f \circ \gamma$ must be done via some coordinate chart so that we can deal with maps from $\mathbb{R}^n$ to $\mathbb{R}^m$, which we understand. Consider Fig. 10.16. Since differentiable manifolds come with coordinate charts (U, ϕ) we can make use of these. We have that

$$f \circ \gamma = f \circ \phi^{-1} \circ \phi \circ \gamma.$$

The map $\phi \circ \gamma : (-\epsilon, \epsilon) \subset \mathbb{R} \to \phi(U) \subset \mathbb{R}^n$ is the sort of map that we can actually do computations with and take derivatives of, and so is the map $f \circ \phi^{-1} : \phi(U) \subset \mathbb{R}^n \to \mathbb{R}$. So we actually have

$$\begin{aligned} v_p[f] &= D_{[\gamma]_p} f \\ &= [f \circ \gamma]'_0 \\ &= [f \circ \phi^{-1} \circ \phi \circ \gamma]'_0 \\ &= [f \circ \phi^{-1}]'_{\phi(p)} \cdot [\phi \circ \gamma]'_0 \\ &= \left. \frac{\partial (f \circ \phi^{-1})}{\partial \phi} \right|_{\phi(p)} \cdot \left. \frac{\partial (\phi \circ \gamma)}{\partial t} \right|_0 . \end{aligned}$$

Notice how we have

Chain Rule for curves	$[f \circ \phi^{-1} \circ \phi \circ \gamma]'_0 = [f \circ \phi^{-1}]'_{\phi \circ \gamma(0)} \cdot [\phi \circ \gamma]'_0 = [f \circ \phi^{-1}]'_{\phi(p)} \cdot [\phi \circ \gamma]'_0.$

This is really one way of writing the chain rule. It may not look anything like the chain rule you are used to, but this is what the chain rule looks like in this context, when we are differentiating while using the equivalence-class-of-curves definition of vectors.

To show this let us consider each of these terms separately. First we look at the mapping

$$\begin{aligned} \phi \circ \gamma : (-\epsilon, \epsilon) \subset \mathbb{R} &\longrightarrow \phi(U) \subset \mathbb{R}^n \\ t &\longmapsto \Big(\phi_1\big(\gamma(t)\big), \phi_2\big(\gamma(t)\big), \ldots, \phi_n\big(\gamma(t)\big)\Big) \\ &= \Big(\phi_1 \circ \gamma(t), \phi_2 \circ \gamma(t), \ldots, \phi_n \circ \gamma(t)\Big) \end{aligned}$$

so $\left.\frac{\partial(\phi\circ\gamma)}{\partial t}\right|_0$ is really the Jacobian matrix evaluated at $t = 0$,

$$\left.\frac{\partial(\phi \circ \gamma)}{\partial t}\right|_0 = \begin{bmatrix} \frac{\partial(\phi_1\circ\gamma)}{\partial t} \\ \frac{\partial(\phi_2\circ\gamma)}{\partial t} \\ \vdots \\ \frac{\partial(\phi_n\circ\gamma)}{\partial t} \end{bmatrix}_0 .$$

Next we look at the mapping

$$\begin{aligned} f \circ \phi^{-1} : \phi(U) \subset \mathbb{R}^n &\longrightarrow \mathbb{R} \\ (\phi_1, \phi_2, \ldots, \phi_n) &\longmapsto f\big(\phi^{-1}(\phi_1, \phi_2, \ldots, \phi_n)\big) \\ & \qquad f \circ \phi^{-1}(\phi_1, \phi_2, \ldots, \phi_n) \end{aligned}$$

so $\left.\frac{\partial(f\circ\phi^{-1})}{\partial\phi}\right|_{\phi(p)}$ is really the Jacobian matrix evaluated at $\phi(p)$,

$$\left.\frac{\partial(f \circ \phi^{-1})}{\partial \phi}\right|_{\phi(p)} = \left[\frac{\partial(f \circ \phi^{-1})}{\partial \phi_1}, \frac{\partial(f \circ \phi^{-1})}{\partial \phi_2}, \ldots, \frac{\partial(f \circ \phi^{-1})}{\partial \phi_n}\right]_{\phi(p)} .$$

Putting these two pieces together we have

$$\begin{aligned} D_{[\gamma]_p} f &= \left[\frac{\partial(f \circ \phi^{-1})}{\partial \phi_1}, \frac{\partial(f \circ \phi^{-1})}{\partial \phi_2}, \ldots, \frac{\partial(f \circ \phi^{-1})}{\partial \phi_n}\right]_{\phi(p)} \begin{bmatrix} \frac{\partial(\phi_1\circ\gamma)}{\partial t} \\ \frac{\partial(\phi_2\circ\gamma)}{\partial t} \\ \vdots \\ \frac{\partial(\phi_n\circ\gamma)}{\partial t} \end{bmatrix}_0 \\ &= \left.\frac{\partial(f \circ \phi^{-1})}{\partial \phi_1}\right|_{\phi(p)} \left.\frac{\partial(\phi_1 \circ \gamma)}{\partial t}\right|_0 + \cdots + \left.\frac{\partial(f \circ \phi^{-1})}{\partial \phi_n}\right|_{\phi(p)} \left.\frac{\partial(\phi_n \circ \gamma)}{\partial t}\right|_0 \\ &= \sum_{i=1}^{n} \left.\frac{\partial(f \circ \phi^{-1})}{\partial \phi_i}\right|_{\phi(p)} \left.\frac{\partial(\phi_i \circ \gamma)}{\partial t}\right|_0 . \end{aligned}$$

Also notice that this means we have

$$[f \circ \phi^{-1}]'_{\phi(p)} \cdot [\phi \circ \gamma]'_0 = \sum_{i=1}^{n} \left.\frac{\partial(f \circ \phi^{-1})}{\partial \phi_i}\right|_{\phi(p)} \left.\frac{\partial(\phi_i \circ \gamma)}{\partial t}\right|_0 ,$$

which makes it clear that $[f \circ \phi^{-1}]'_{\phi(p)} \cdot [\phi \circ \gamma]'_0$ is indeed simply another way of writing the chain rule.

Question 10.2 Suppose that (U, ϕ) and $(\tilde{U}, \tilde{\phi})$ are two charts on M, $p \in U \cap \tilde{U}$, $\gamma : (-\epsilon, \epsilon) \subset \mathbb{R} \to M$ with $\gamma(0) = p$, and f is a real-valued function on M. Show that

$$\left.\frac{\partial(f \circ \phi^{-1} \circ \phi \circ \gamma)}{\partial t}\right|_0 = \left.\frac{\partial(f \circ \tilde{\phi}^{-1} \circ \tilde{\phi} \circ \gamma)}{\partial t}\right|_0 .$$

10.3 Push-Forwards and Pull-Backs on Manifolds

We have already discussed push-forwards and pull-backs in Chap. 6. One of the examples we looked at closely was the change of variable formula $f : \mathbb{R}^2_{xy} \to \mathbb{R}^2_{uv}$ given by $f(x, y) = \big(u(x, y), v(x, y)\big) = (x + y, x - y)$. We had found the push-forward of f to be the mapping

$$T_p f \equiv f_* \equiv D_p f : T_p\mathbb{R}^2_{xy} \longrightarrow T_{f(p)}\mathbb{R}^2_{uv}$$

given by the Jacobian matrix

$$T_p f \equiv f_* \equiv D_p f = \begin{bmatrix} \frac{\partial u}{\partial x} & \frac{\partial u}{\partial y} \\ \frac{\partial v}{\partial x} & \frac{\partial v}{\partial y} \end{bmatrix}_p .$$

More generally, given a mapping between two manifolds $f : M \to N$ this mapping f induces another mapping $Tf : TM \to TN$ from the tangent bundle of M to the tangent bundle of N. At a particular point $p \in M$ we have $T_p f : T_p M \to T_p N$. This mapping pushes-forward vectors $v_p \in T_p M$ to vectors $T_p f \cdot v_p \in T_{f(p)} N$. See Fig. 10.17.

Now we want to look at this from our more abstract perspective. Suppose we had a map $\phi : M \to N$ and we had the vector v_p, which was determined by the curve γ. We want to know what $T_p\phi \cdot v_p$ is. From Fig. 10.18 it should be clear that the push-forward of vector $[\gamma]$ by the mapping ϕ is the curve $\phi \circ \gamma$. Thus we define the push-forward of $[\gamma]_p$ by $T_p\phi$ to be

Push-foraward of vector $[\gamma]_p$ by $T_p\phi$	$T_p\phi \cdot [\gamma]_p = [\phi \circ \gamma]_{\phi(p)}.$

In a lot of ways this is a much cleaner formulation of the push-forward. It also makes abstract computations much simpler, as we will see. Notationally there are several different ways to write this, all of which represent the same thing.

Ways to write push-foraward of vector $[\gamma]_p$by $T_p\phi$

$$\begin{aligned} T_p\phi \cdot [\gamma]_p &\equiv \phi_*(p)\big([\gamma]_p\big) \\ &\equiv \phi_*(p) \cdot [\gamma]_p \\ &\equiv \phi_*\big([\gamma]_p\big) \\ &\equiv [\phi \circ \gamma]_{\phi(p)}. \end{aligned}$$

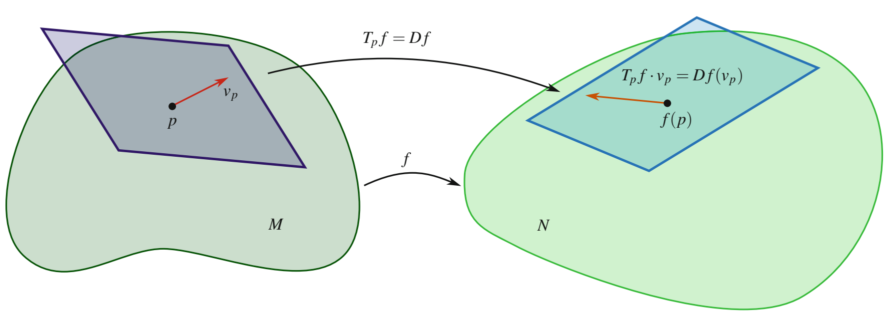


Fig. 10.17 Given a mapping $f : M \to N$ between manifolds then f induces the push-forward mapping $Tf : TM \to TN$. At a particular point we have $T_p f : T_p M \to T_p N$

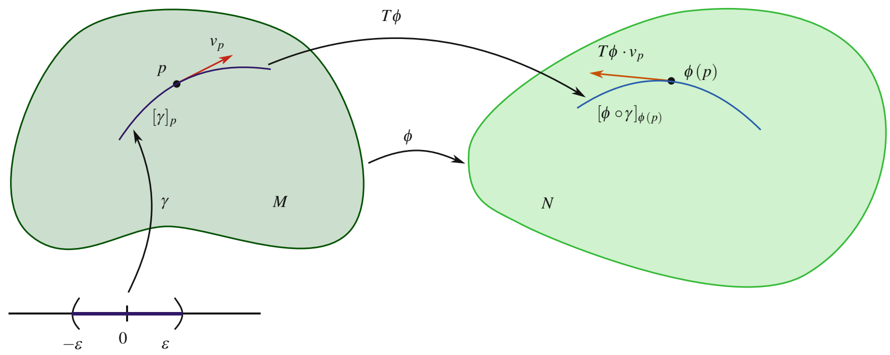


Fig. 10.18 Here we have a mapping $\phi : M \to N$ between manifolds and a vector $[\gamma]_p \in T_pM$, where $\gamma : (-\epsilon, \epsilon) \to M$ with $\gamma(0) = p$. What is the push-forward by $T_p\phi$ of this vector? It is simply the vector given by the curve $\phi \circ \gamma$, that is, by $[\phi \circ \gamma]_{\phi(p)}$

Of course, all of these notations can also be written leaving off the base point:

$$T\phi \cdot [\gamma] \equiv \phi_*\big([\gamma]\big) \equiv \phi_* \cdot [\gamma] = [\phi \circ \gamma].$$

Turning to our familiar example $\phi : \mathbb{R}^2_{xy} \to \mathbb{R}^2_{uv}$ given by $\phi(x, y) = \big(u(x, y), v(x, y)\big) = (x + y, x - y)$ we can see that these two ways of computing the push-forward of a vector give the same answer. Consider again the vector $u_0 = \begin{bmatrix} 1 \\ 2 \end{bmatrix} \equiv [\gamma]_0$ where

$$\begin{aligned} \gamma : (-\epsilon, \epsilon) &\longrightarrow \mathbb{R}^2_{xy} \\ t &\longmapsto (t, 2t). \end{aligned}$$

We wish to find the push-forward of u_0 by ϕ. First we do this by finding the tangent mapping $T\phi$ and then use that to find $T\phi \cdot u_0$.

$$T\phi = \phi_* = \begin{bmatrix} \frac{\partial u}{\partial x} & \frac{\partial u}{\partial y} \\ \frac{\partial v}{\partial x} & \frac{\partial v}{\partial y} \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}.$$

We do not bother with worrying about the base point for this matrix since clearly $T\phi$ does not depend on the base point it is located at. We then find the push-forward of u_0 by

$$T\phi \cdot u_0 = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \end{bmatrix}_{\phi(0)} = \begin{bmatrix} 3 \\ -1 \end{bmatrix}_0.$$

Now we compute the same thing using the more abstract equivalence class of curves definition of vectors,

$$T\phi \cdot [\gamma]_0 = [\phi \circ \gamma]_{\phi(0)} = [\phi \circ (t, 2t)]_0 = [(t + 2t, t - 2t)]_0 = [(3t, -t)]_0.$$

Thus the curve $\phi \circ \gamma = (3t, -t)$ is the push-forward vector we are looking for. That is, $T\phi \cdot u_0 \equiv [(3t, -t)]_0$. In order to check that our answer is correct and equivalent to what we found just above we have to take the derivative of this curve with respect to time,

$$[\phi \circ \gamma]'_{\phi(0)} = \begin{bmatrix} \left.\frac{\partial(\phi\circ\gamma)_1(t)}{\partial t}\right|_{t=0} \\ \left.\frac{\partial(\phi\circ\gamma)_2(t)}{\partial t}\right|_{t=0} \end{bmatrix}_{\phi(0)} = \begin{bmatrix} \left.\frac{\partial(3t)}{\partial t}\right|_{t=0} \\ \left.\frac{\partial(-t)}{\partial t}\right|_{t=0} \end{bmatrix}_{\phi(0)} = \begin{bmatrix} 3 \\ -1 \end{bmatrix}_0,$$

which is exactly what we wanted to show.

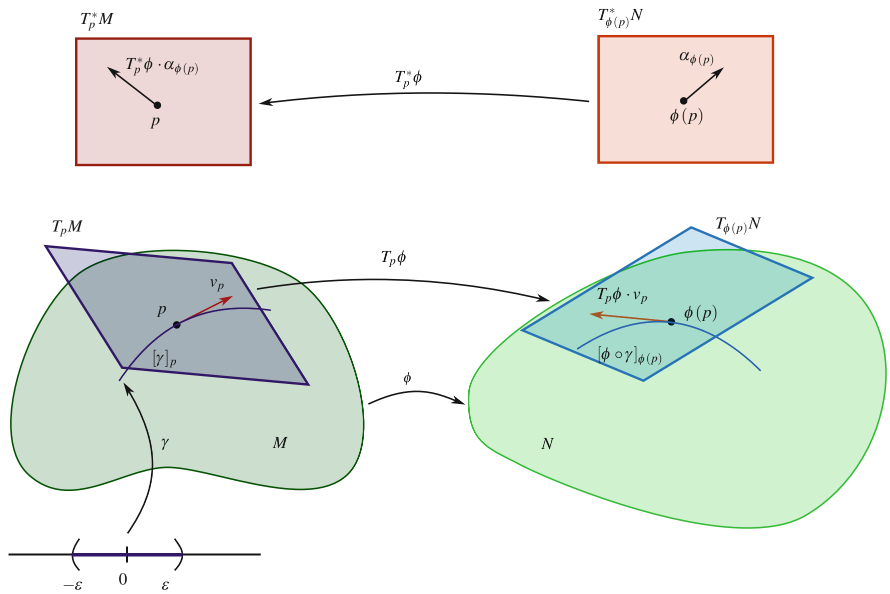


Fig. 10.19 Here the relationship between push-forward of vectors and pull-backs of one forms is illustrated. We picture the cotangent space T_p^*M and $T^*_{\phi(p)}N$ above the tangent space, even though both tangent space and cotangent space are "attached" to the manifold at the same point. Notice that the pull-back mapping goes in the opposite direction as the push-forward mapping

Question 10.3 For a mapping $\phi : M \to N$ and a function $f : N \to \mathbb{R}$, show that $v_p[f \circ \phi] = \big(T_p\phi \cdot v_p\big)_{\phi(p)}[f]$.

Now that we have addressed push-forwards of vectors we want to look at pull-backs of differential forms. Conceptually, what is happening is not very difficult to understand. The relevant mappings are defined as

$$\begin{aligned} \phi : M &\longrightarrow N \\ p &\longmapsto \phi(p) \end{aligned}$$

$$\begin{aligned} T_p\phi : T_pM &\longrightarrow T_{\phi(p)}N \\ v_p &\longmapsto T_p\phi \cdot v_p \\ [\gamma]_p &\longmapsto [\phi \circ \gamma]_{\phi(p)} \end{aligned}$$

$$\begin{aligned} T_p^*M &\longleftarrow T_{\phi(p)}N : T_p^*\phi \\ T_p^*\phi \cdot \alpha_{\phi(p)} &\longleftarrow\!\shortmid\ \alpha_{\phi(p)} \end{aligned}$$

and are shown in Fig. 10.19. The pull-back of a one-form $\alpha_{\phi(p)} \in T_{\phi(p)}N$ is $T_p^*\phi \cdot \alpha_{\phi(p)} \in T_p^*M$ which therefore eats vectors $v_p \in T_pM$. Thus we defined the pull-back of a one-form by

$$\Big(T_p^*\phi \cdot \alpha_{\phi(p)}\Big)\big(v_p\big) \equiv \alpha_{\phi(p)}\Big(T_p\phi \cdot v_p\Big).$$

Also recall how the point that indexes the pull-back mapping is the point p from the range, not the domain. This is a rather unusual notational convention, but given the above definition it makes the connection between the pull-back and the push-forward maps obvious. Using the equivalence class of curves definition of vectors that we introduced in this section we have the pull-back given by

$$\boxed{\begin{array}{ll}\text{Pull-back of one-form } \alpha_{\phi(p)} \text{ by } T_p^*\phi & \begin{aligned}\Big(T_p^*\phi \cdot \alpha_{\phi(p)}\Big)\big([\gamma]_p\big) &\equiv \alpha_{\phi(p)}\Big(T_p\phi \cdot [\gamma]_p\Big) \\ &= \alpha_{\phi(p)}\Big(\cdot [\phi \circ \gamma]_{\phi(p)}\Big)\end{aligned}\end{array}}$$

This equation actually looks really simple in more traditional notation. Here we write the pull-back without the base point so you can see how simple it really is,

$$\boxed{\phi^*\alpha\big([\gamma]\big) = \alpha\big([\phi \circ \gamma]\big).}$$

Similarly, for a k-form $\omega_{\phi(p)} \in \bigwedge^k_{\phi(p)}(N)$ we get the pull-back mapping

$$\begin{aligned}\bigwedge\nolimits^k_p(M) &\longleftarrow \bigwedge\nolimits^k_{\phi(p)}(N) : T_p^*\phi \\ T_p^*\phi \cdot \omega_{\phi(p)} &\longleftarrow\!\shortmid \omega_{\phi(p)},\end{aligned}$$

which is defined as

$$\Big(T_p^*\phi \cdot \omega_{\phi(p)}\Big)\big(v_{p_1}, v_{p_2}, \ldots, v_{p_k}\big) \equiv \omega_{\phi(p)}\Big(T_p\phi \cdot v_{p_1}, T_p\phi \cdot v_{p_2}, \ldots, T_p\phi \cdot v_{p_k}\Big).$$

In terms of the equivalence class of curves definition of vectors we have the pull-back of a p-form ω given by

$$\boxed{\begin{array}{ll}\text{Pull-back of } p\text{-form } \omega_{\phi(p)} \text{ by } T_p^*\phi & \begin{aligned}\Big(T_p^*\phi \cdot \omega_{\phi(p)}\Big)\big([\gamma_1]_p, [\gamma_2]_p, \ldots, [\gamma_k]_p\big) &\equiv \omega_{\phi(p)}\Big(T_p\phi \cdot [\gamma_1]_p, T_p\phi \cdot [\gamma_2]_p, \ldots, T_p\phi \cdot [\gamma_k]_p\Big) \\ &= \omega_{\phi(p)}\Big([\phi \circ \gamma_1]_{\phi(p)}, [\phi \circ \gamma_2]_{\phi(p)}, \ldots, [\phi \circ \gamma_k]_{\phi(p)}\Big).\end{aligned}\end{array}}$$

We also give the definition of the pull-back of a zero-form. Recall that a zero-form on the manifold N is just a real-valued function $f : N \to \mathbb{R}$. The pull-back of a function is exactly what you would expect it to be,

$$\boxed{\begin{array}{ll}\text{Pull-back of zero-form } f \text{ by } T_p^*\phi & T_p^*\phi \cdot f \equiv \big(f \circ \phi\big)(p) = f\big(\phi(p)\big).\end{array}}$$

This definition looks very simple in traditional notation and without the base point included,

$$\boxed{\begin{array}{ll}\text{Pull-back of zero-form } f \text{ by } \phi^* & \phi^* \cdot f \equiv f \circ \phi.\end{array}}$$

In fact, we have already seen and used this in Sect. 6.7 in the proof of the third identity $\phi^* d = d\phi^*$.

10.4 Calculus on Manifolds

There is no reason we couldn't classify all the calculus you learned in your calculus and vector calculus classes as "calculus on a manifold." In these cases the manifold would simply be the Euclidian spaces $\mathbb{R}^2$ or $\mathbb{R}^3$. However, when one hears the phrase *calculus on manifolds* or perhaps alternative phrases like, *integration on manifolds* or even *analysis on manifolds* then the general assumption is that the manifold is more complicated than Euclidian space, that it is the sort of manifold we have introduced in this chapter, one with an atlas made up of coordinate charts (U_i, ϕ_i).

10.4.1 Differentiation on Manifolds

The word calculus of course implies both differentiation and integration. In this section we will make a few big-picture comments about differentiation on manifolds. The next subsection, as well as the next chapter, will focus on integration on manifolds. There are **three distinct notions** of differentiation on manifolds that one often encounters. All three of these notions work a little bit differently.

1. Exterior derivative: This is the kind of differentiation we are studying in this book and have already looked extensively at exterior differentiation in Chap. 4. Exterior differentiation acts on differential forms. When the differential form is a zero-form, or in other words simply a function, exterior differentiation is identical to the directional derivative from calculus. As we saw in Sect. 4.5, in a very real sense exterior differentiation is a generalization of the idea of directional derivatives from our general calculus classes. It is also this version of differentiation that gives us the generalized Stokes' theorem that is proved in Chap. 11. This generalizes Stokes' Theorem is an important component of integration on manifolds. Finally, in Sect. 4.3 we showed the local existence and uniqueness of the exterior differential operator. Now that we have the necessary concepts we will show global existence and uniqueness of the exterior differential operator below.
2. Lie derivative: Lie derivatives act on objects called tensors. Differential forms are in fact a subset of tensors and so it also makes perfect sense to refer to the Lie derivative of a differential form. Also, Lie differentiation has a very nice geometric meaning that is easier to visualize and understand than than exterior differentiation. We will introduce tensors in Appendix A and discuss the Lie derivative in Sect. A.7. There we derive several formulas that relate exterior differentiation with Lie differentiation. In fact, one of these formulas is used to prove, by induction, the global exterior derivative formula that was discussed from a somewhat geometric standpoint in Sect. 4.4.2.
3. Covariant derivative: Covariant derivatives will not be discussed at all in this book, but they rely on the concept of a connection that was discussed very briefly at the start of Sect. 10.1. As was mentioned, connections are additional structures on a manifold that "connect" tangent spaces in order to tell us what the parallel transport of a vector from one tangent space to another is. The essential idea is that knowing what constitute parallel vectors in different tangent spaces allows us to measure how tensors and forms change as their input vectors are parallel transported. Covariant derivatives are a second way to generalize directional derivatives from our general calculus classes.

You should recognize that all three of these notions of derivative are equally valid when discussing $\mathbb{R}^2$ or $\mathbb{R}^3$. It is just that in your introductory calculus classes you were only really interested in functions and not differential forms or tensors. On top of that there was a very natural, and intuitively understandable, Euclidian connection on $\mathbb{R}^2$ or $\mathbb{R}^3$ built into our manifold. Because of these considerations there was no need to actually ever discuss the various concepts of differentiation in a general calculus class, old-fashioned directional derivatives were enough. By the time you got to vector calculus, you actually did study exterior differentiation, but as you leaned in Chap. 9 exterior differentiation was given the special names gradient, curl, and divergence. In other words, the nature of exterior differentiation was masked by a different vocabulary and an attempt to avoid introducing forms. That is why everything was kept as a vector.

But in the more general and abstract setting we need to pay attention to these different concepts of differentiation. There are a large number of identities that relate these different ideas of derivative to each other. All of this makes the topic of calculus on a manifold quite a bit more complicated than calculus on the Euclidian spaces $\mathbb{R}^2$ or $\mathbb{R}^3$. Calculus on a manifold is really starting to get into the field of differential geometry, and all of the above mentioned ideas would really constitute the introductory topics for a first course in differential geometry.

With these comments made, we now turn our attention to some of the basics that are important in the context of differential forms. We begin by returning briefly to where we left off in Sect. 4.3. In that section we took four properties that we wanted exterior differentiation to have, called them axioms, and then derived a general formula for exterior differentiation. The fourth axiom basically stated that in the case of a zero-form that exterior differentiation had to be the same as the directional derivative, thereby ensuring that exterior differentiation was in fact a generalization of the directional derivative. We then used these axioms to derive a general formula for the exterior derivative. Once we had done that and had a single formula for the exterior derivative of a general differential form we knew that the exterior derivative operator d had to both exist and be unique.

Take a moment to review Sect. 4.3 carefully. Notice two things. First, that axiom four stated that in local coordinates, for each function f, $df = \sum \frac{\partial f}{\partial x_i} dx_i$. Second, we had supposed that we had the n-form

$$\alpha = \sum \alpha_{i_1 \cdots i_n} \, dx_{i_1} \wedge \cdots \wedge dx_{i_n}.$$

In both of these cases we were implicitly relying on a coordinate system, and in the context of that section we were implicitly relying on the Cartesian coordinate system. But a general manifold does not have a single coordinate system, instead it has an atlas $\{(U_i, \phi_i)\}$ of coordinate patches (U_i, ϕ_i). What was done in Sect. 4.3 only applies to a single coordinate patch (U_i, ϕ_i). Of course the same argument and computation can be done on each of the many coordinate patches, but all that means is that on each of the many coordinate patches we found a formula that gives the exterior derivative of a differential form, as long as that form is written in the coordinates of that particular coordinate patch.

Now we show that the exterior derivative operator d exists and is unique globally. That means that d exists and is unique over the whole manifold $M = \bigcup U_i$. On the coordinate patch (U_i, ϕ_i) we showed that the exterior derivative operator, which we will now label d_{U_i}, exists and is unique. We define the global differential operator d as $d = d_{U_i}$ on this coordinate patch. Similarly, on the coordinate patch (U_j, ϕ_j) we showed that the exterior derivative operator, which we will now label d_{U_j}, exists and is unique. Again, we define the global differential operator d as $d = d_{U_j}$ on this coordinate patch. The question is, what happens if $U_i \cap U_j \neq \varnothing$? Since d_{U_i} exists and is unique on U_i and d_{U_j} exists and is unique on U_j, then on $U_i \cap U_j$ we must have $d_{U_i} = d_{U_j} = d$. Since this is true on the intersection of all coordinate patches then we have d existing and unique globally, that is, over $M = \bigcup U_i$.

10.4.2 Integration on Manifolds

Now we turn our attention to integration on a manifold. Since there are several different ideas for differentiation on a manifold you should not be surprised that there are also different ideas of integration on a manifold. Here we will restrict ourselves to discussing integration of differential forms on a manifold.

In order for everything to turn out nicely we will assume the manifold is **oriented**. There are ways to do integration when a manifold is not oriented, but we will not discuss them here. An oriented n-dimensional manifold M is a manifold that has an n-form that is not zero at any point of the manifold. In other words, if $f dx_1 \wedge \cdots \wedge dx_n$ is an n-form on M then there is no $p \in M$ such that $f(p) = 0$. That implies that for all points $p \in M$ either $f(p) > 0$ or $f(p) < 0$. A form that has this property is called a **nowhere-zero n-form**. Thus the set of all nowhere-zero n-forms splits into two equivalence classes, one equivalence class that consists of all the everywhere positive nowhere-zero n-forms and an a second equivalence class that consists of all the everywhere negative nowhere-zero n-forms. By choosing one of these two equivalence classes we are specifying what is called an **orientation** of the manifold. We can think of an orientation as allowing us to find volumes on the manifold in a consistent way. Alternatively, we can think of an orientation as allowing us to orient all tangent spaces in a consistent way. The standard orientation of the manifold $\mathbb{R}^n$ is the equivalence class that contains the nowhere-zero n-form $dx_1 \wedge \cdots \wedge dx_n$. This orientation is often described as the orientation induced by $dx_1 \wedge \cdots \wedge dx_n$.

But notice what we did here. Saying a manifold M has an orientation induced by the n-form $dx_1 \wedge \cdots \wedge dx_n$ is all well and good as long as we have a manifold whose atlas consists of a single chart. But what if we do not? Consider Fig. 10.20 where we show a manifold with two charts, (U_i, ϕ_i) and (U_j, ϕ_j), where $U_i \cap U_j \neq \varnothing$. What would a volume form on U_i look like? Let us consider the standard volume form $dx_1 \wedge \cdots \wedge dx_n$, which is defined on all of $\mathbb{R}^n$ and hence is clearly defined on $\phi_i(U_i) \subset \mathbb{R}^n$. We can pull-back this standard volume form onto $U_i \subset M$. Thus $T^*\phi_i \cdot \big(dx_1 \wedge \cdots \wedge dx_n\big)$ is, in a sense, a "standard" volume form on $U_i \subset M$.

Similarly, what would a volume form on U_j look like? Like before we will consider the standard volume form on $\phi_j(U_i) \subset \mathbb{R}^n$, which we will call $dy_1 \wedge \cdots \wedge dy_n$ so as not to confuse it with $dx_1 \wedge \cdots \wedge dx_n$. We can pull-back this standard volume form onto $U_j \subset M$ to get the pull-back $T^*\phi_j \cdot \big(dy_1 \wedge \cdots \wedge dy_n\big)$ as a "standard" volume form on $U_j \subset M$. How do these two volume forms relate to each other on $U_i \cap U_j$ where they are both defined? In general they will not be the same. But remember what we are actually after. What we really want is to define a consistent volume form ω on all of M so that we can use it to define an orientation on M.

What does finding a "consistent volume form" ω on all of M actually mean? Consider Fig. 10.21. Suppose we have the standard volume form $dy_1 \wedge \cdots \wedge dy_n$ on $\phi_j(U_i)$. If there is a volume form $f(x_1, \ldots, x_n) dx_1 \wedge \cdots \wedge dx_n$ on $\phi_i(U_i)$ such that on $U_i \cap U_j$ we have

$$T^*\phi_i \Big(f(x_1, \ldots, x_n) dx_1 \wedge \cdots \wedge dx_n \Big) = T^*\phi_j \Big(dy_1 \wedge \cdots \wedge dy_n \Big)$$

then we could use these two volume forms, one on U_j and one on U_i to "stitch together" a volume form that is defined on $U_i \cup U_j$. By doing this repeatedly we can "stitch together" a volume form ω that is defined on all of M. But first we need to find the volume form $f(x_1, \ldots, x_n) dx_1 \wedge \cdots \wedge dx_n$ on $\phi_i(U_i)$. Looking at Fig. 10.21 it is clear that the equality we want,

$$T^*\phi_i \Big(f(x_1, \ldots, x_n) dx_1 \wedge \cdots \wedge dx_n \Big) = T^*\phi_j \Big(dy_1 \wedge \cdots \wedge dy_n \Big),$$

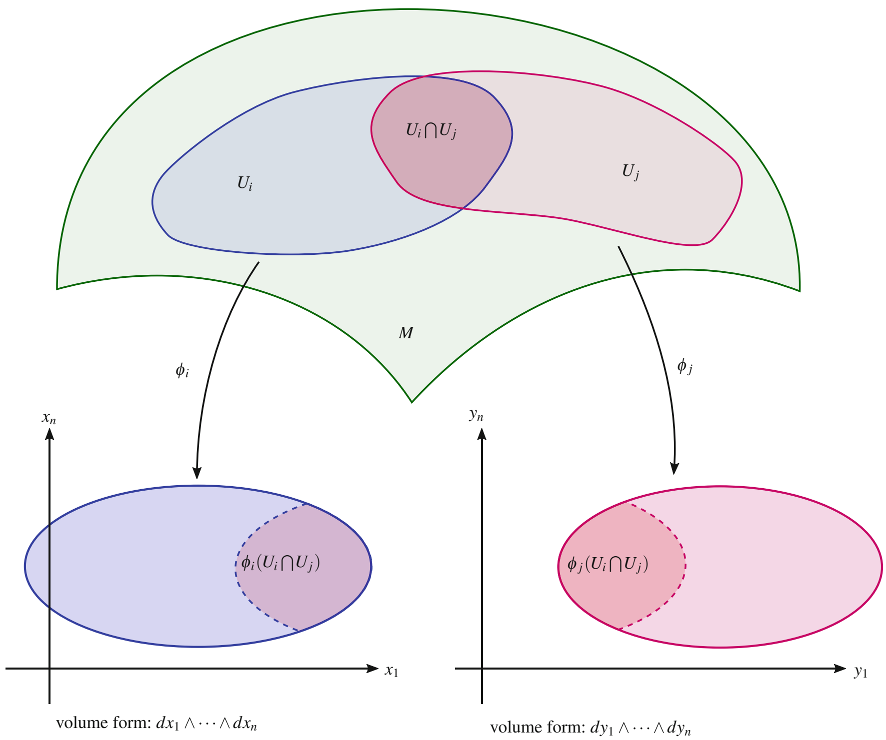


Fig. 10.20 The manifold M with two charts (U_i, ϕ_i) and (U_j, ϕ_j), where $U_i \cap U_j \neq \varnothing$. The standard volume for $dx_1 \wedge \cdots \wedge dx_n$ for $\mathbb{R}^n$ exists on both $\phi_i(U_i)$ and on $\phi_j(U_j)$

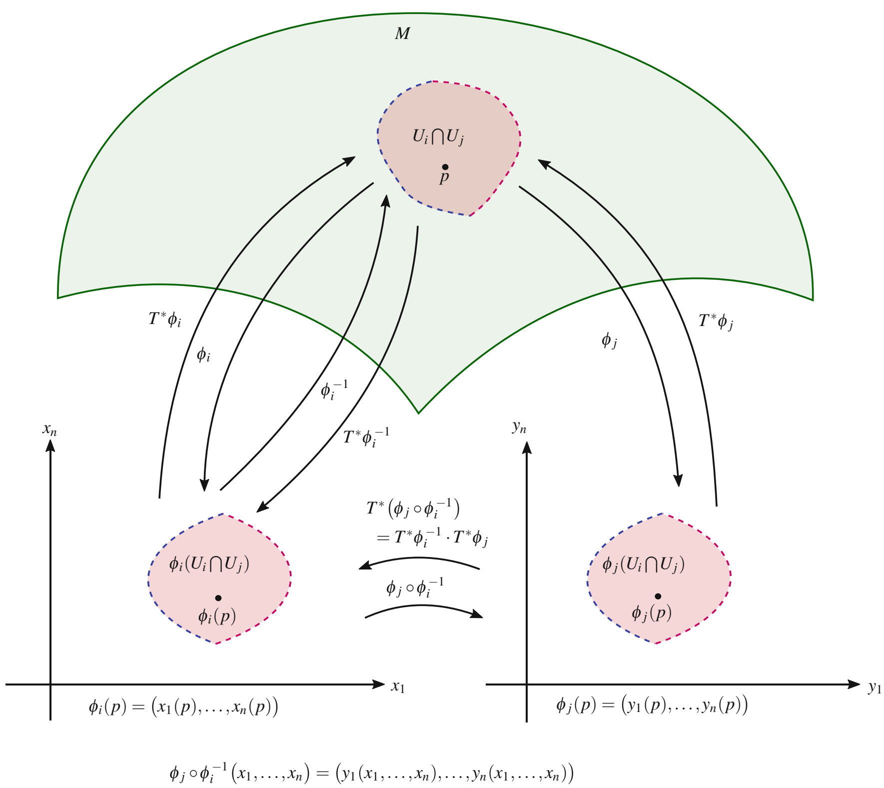


Fig. 10.21 Here we focus on $U_i \cap U_j$ from Fig. 10.20. All the necessary mappings to find $T^*\left(\phi_j \circ \phi_i^{-1}\right)\left(dy_1 \wedge \cdots \wedge dy_n\right)$

is equivalent to

$$\begin{aligned} f(x_1,\dots,x_n)dx_1\wedge\cdots\wedge dx_n &= T^*\phi_i^{-1}\cdot T^*\phi_j\Big(dy_1\wedge\cdots\wedge dy_n\Big)\\ &= T^*\Big(\phi_j\circ\phi_i^{-1}\Big)\Big(dy_1\wedge\cdots\wedge dy_n\Big). \end{aligned}$$

The way we actually go about finding this is to rely on the mapping

$$\phi_j\circ\phi_i^{-1}(x_1,\dots,x_n)=\Big(y_1(x_1,\dots,x_n),\dots,y_n(x_1,\dots,x_n)\Big)$$

to write the volume form $dy_1\wedge\cdots\wedge dy_n$ in terms of $dx_1,\dots,dx_n$,

$$\begin{aligned} dy_1\wedge\cdots\wedge dy_n &= dy_1(x_1,\dots,x_n)\wedge\cdots\wedge dy_n(x_1,\dots,x_n)\\ &= \left(\frac{\partial y_1}{\partial x_1}dx_1+\cdots+\frac{\partial y_1}{\partial x_n}dx_n\right)\wedge\cdots\wedge\left(\frac{\partial y_n}{\partial x_1}dx_1+\cdots+\frac{\partial y_n}{\partial x_n}dx_n\right)\\ &= \underbrace{\begin{vmatrix} \frac{\partial y_1}{\partial x_1} & \cdots & \frac{\partial y_1}{\partial x_n}\\ \vdots & \ddots & \vdots\\ \frac{\partial y_n}{\partial x_1} & \cdots & \frac{\partial y_n}{\partial x_1}\end{vmatrix}}_{\substack{\text{determinant of}\\ \text{Jacobian of } \phi_j\circ\phi_i^{-1}}} dx_1\wedge\cdots\wedge dx_n. \end{aligned}$$

Thus, we have

$$T^*\Big(\phi_j\circ\phi_i^{-1}\Big)\Big(dy_1\wedge\cdots\wedge dy_n\Big)=\begin{vmatrix} \frac{\partial y_1}{\partial x_1} & \cdots & \frac{\partial y_1}{\partial x_n}\\ \vdots & \ddots & \vdots\\ \frac{\partial y_n}{\partial x_1} & \cdots & \frac{\partial y_n}{\partial x_1}\end{vmatrix} dx_1\wedge\cdots\wedge dx_n.$$

In general the determinant of the Jacobian matrix of $\phi_j\circ\phi_j^{-1}$ could either be positive or negative, and usually we would not care. The pull-back of $dy_1\wedge\cdots\wedge dy_n$ simply is what it is. However, right now we are interested in a consistent volume form ω on all of M which means, in $U_i\cap U_j$, we have

$$T^*\phi_i\Big(f(x_1,\dots,x_n)dx_1\wedge\cdots\wedge dx_n\Big)=T^*\phi_j\Big(dy_1\wedge\cdots\wedge dy_n\Big).$$

Since $dx_1\wedge\cdots\wedge dx_n$ and $dy_1\wedge\cdots\wedge dy_n$ have the same orientation, in order for the right and the left sides of the above equation to have same orientation we would need to have the Jacobian of $\phi_j\circ\phi_j$ to be positive. Thus, as long as the mappings ϕ_i and ϕ_j are such that the transition function $\phi_j\circ\phi_i^{-1}$ has a Jacobian with a positive determinant, we can define a consistent volume form on $U_i\cup U_j$.

Extending this to all of M, we can find a consistent volume form on M as long as M has an atlas $\big\{(U_i,\phi_i)\big\}$ such that every transition function $\phi_j\circ\phi_i^{-1}$ has a Jacobian with a positive determinant. Another way to phrase this would be to say that a manifold M is orientable as long as it **admits** an atlas whose transition functions have a positive Jacobian determinant.

Now we turn to the next idea that we need to handle integration on manifolds. Suppose we have a manifold M with an atlas $\big\{(U_i,\phi_i)\big\}$. Since $\big\{(U_i,\phi_i)\big\}$ is an atlas of M we have $M=\bigcup U_i$. The set of coordinate patches $\big\{U_i\big\}$ is called a **covering** of M. For each coordinate patch U_i suppose we had a real-valued function $\varphi_i: M\to\mathbb{R}$ that was non-zero only for some subset V of U_i. In other words, if $p\notin V\subset U_i$ then $\varphi_i(p)=0$. The function φ_i is said to be **subordinate** to the coordinate patch U_i. A *partition of unity* for M subordinate to the covering $\{U_i\}$ is a collection of of functions $\varphi_i: M\to\mathbb{R}$ such that

(a) for each coordinate patch U_i in the cover of M we have a function φ_i, which is subordinate to the coordinate patch U_i,
(b) such that $0\le\varphi_i(p)\le 1$ for all $p\in U_i$, and
(c) for every $p\in M$ we have $\sum_i\varphi_i(p)=1$.

If the manifold M is orientable it is generally assumed that the atlas used to make up the covering of M used in finding the partition of unity is in fact the atlas whose transition functions have a positive Jacobian determinant, that is, the atlas used to define the volume form ω on M.

The partition of unity allows us to "break up" any differential form α into a set collection of differential forms which are subordinate to the covering $\{U_i\}$,

$$\alpha = \sum_i \varphi_i \alpha.$$

It is clear that $\varphi_i \alpha$ is only non-zero on the coordinate patch U_i. We can only actually integrate α over regions that cover multiple coordinate patches by breaking it up into forms that are integrable on each coordinate patch. Thus for example, if we wanted to integrate α over M we would have

$$\begin{aligned}\int_M \alpha &= \int_M \sum_i \varphi_i \alpha \\ &= \sum_i \int_M \varphi_i \alpha \\ &= \sum_i \int_{U_i} \varphi_i \alpha\end{aligned}$$

where the last equality follows since $M = \bigcup_i U_i$ and $\varphi_i \alpha = 0$ outside of U_i. If we were integrating over a region $R \subset M$ then we have

$$\int_R \alpha = \sum_i \int_{U_i \cap R} \varphi_i \alpha.$$

So, being able to integrate a form on any region of M, or even on all of M, boils down to being able to integrate on a single coordinate patch. So now we will turn our attention to integrating a form α on a single coordinate patch. Notice, we change notation and simply use α instead of $\varphi_i \alpha$.

We begin by assuming α is an n-form on the n-dimensional manifold M and we want to find $\int_{U_i} \alpha$. All we have to do is pull-back α to $\phi_i(U_i)$ and integrate there, in other words,

$$\int_{U_i} \alpha = \int_{\phi_i(U_i)} T^*\phi_i^{-1} \cdot \alpha.$$

If you were asked to integrate the real-valued function $f : M \to \mathbb{R}$ then what this really means is that we want to integrate the n-form $f\omega$, where ω is the volume form on M. Thus we use

$$\begin{aligned}\int_{U_i} f\omega &= \int_{\phi_i(U_i)} T^*\phi_i^{-1} \cdot (f\omega) \\ &= \int_{\phi_i(U_i)} f \circ \phi_i^{-1}\, T^*\phi_i^{-1} \cdot \omega.\end{aligned}$$

Question 10.4 Show the second equality in the above equation.

In order to integrate a k-form β on an n-dimensional manifold M, where $k < n$, then we need need a k-dimensional submanifold to integrate over. Without getting into the technical definition of what a submanifold is, you can think of it essentially as a parameterized surface $\Sigma \subset M$ that is given by a mapping $\Phi : \mathbb{R}^k \to M$. This is exactly what was done in Sect. 7.4 where we integrated over one- and two-dimensional parameterized surfaces. We can then integrate β on the submanifold Σ by

$$\int_{\Sigma} \beta = \int_{\Phi^{-1}(\Sigma)} T^*\Phi \cdot \beta.$$

So, even though the discussion in this section has been quite abstract and theoretical, it turns out that we have already done a number of concrete examples for what we are discussing.

This actually covers only the very basics of integration on manifolds. In Chap. 11 we will expand upon this topic by proving a very important theorem regarding integration on manifolds, a theorem that is central to a great deal of mathematics, called the generalized Stokes' theorem, or more usually, simply called Stokes' theorem.

10.5 Summary, References, and Problems

10.5.1 Summary

An n-dimensional manifold is a space M that can be completely covered by a collection of local coordinate neighborhoods U_i with one-to-one mappings $\phi_i : U_i \to \mathbb{R}^n$, which are called a coordinate maps. Together U_i and ϕ_i are called a coordinate patch or a chart, which is generally denoted as (U_i, ϕ_i). The set of all the charts together, $\big\{(U_i, \phi_i)\big\}$, is called a coordinate system or an atlas of M. Since the U_i cover all of M we write that $M = \bigcup U_i$. Also, since ϕ_i is one-to-one it is invertible, so ϕ_i^{-1} exists and is well defined. If two charts have a non-empty intersection, $U_i \cap U_j \neq \varnothing$, then the functions $\phi_j \circ \phi_i^{-1} : \mathbb{R}^n \to \mathbb{R}^n$ are called transition functions.

$$\begin{aligned} U_i \quad &: \quad \text{coordinate neighborhood} \\ \phi_i : U_i \to \mathbb{R}^n \quad &: \quad \text{coordinate map} \\ (U_i, \phi_i) \quad &: \quad \text{coordinate patch/chart} \\ \big\{(U_i, \phi_i)\big\} \quad &: \quad \text{coordinate system/atlas} \\ \phi_j \circ \phi_i^{-1} : \mathbb{R}^n \to \mathbb{R}^n \quad &: \quad \text{transition function} \end{aligned}$$

Suppose $r \in U_i \cap U_i \neq \emptyset$, then $\phi_i(r) \in \mathbb{R}^n$ and $\phi_j(r) \in \mathbb{R}^n$. Furthermore, $\phi_j \circ \phi_i^{-1}$ sends $\phi_i(U_i \cap U_j) \subset \mathbb{R}^n$ to $\phi_j(U_i \cap U_j) \subset \mathbb{R}^n$. That is, $\phi_j \circ \phi_i^{-1}$ is a map of a subset of $\mathbb{R}^n$ to another subset of $\mathbb{R}^n$, and so the mapping $\phi_j \circ \phi_i^{-1} : \mathbb{R}^n \to \mathbb{R}^n$ with domain $\phi_i(U_i \cap U_j)$ and range $\phi_j(U_i \cap U_j)$ is the sort of mapping that we know how to differentiate from multivariable calculus. A differentiable manifold is a set M, together with a collection of charts (U_i, ϕ_i), where $M = \bigcup U_i$, such that every mapping $\phi_j \circ \phi_i^{-1}$, where $U_i \cap U_j \neq \emptyset$, is differentiable.

We define the tangent space of M at point p, T_pM, in terms of curves on M. If two curves have the same range close to p and have the same parametrization close to p they are called equivalent, which is denoted $\sim$. The set of all equivalent curves is called an equivalence class and is defined by

$$[\gamma_1] \equiv \big\{\gamma \;\big|\; \gamma \sim \gamma_1\big\}.$$

Each equivalence class of curves at a point p is defined to be a tangent vector at p. The tangent space of M at p is defined as the set of all tangent vectors, that is, equivalence classes of curves, at the point p,

$$T_pM = \Big\{ [\gamma]_p \;\Big|\; \gamma : (-\epsilon, \epsilon) \to M \text{ and } \gamma(0) = p \Big\}.$$

The relation between the equivalence class (intrinsic) definition and the vector (extrinsic) definitions of the tangent space is summarized,

Equivalence Class Definition	$\Longleftrightarrow$	Vector Definition
$v_p = [\gamma]_p, \quad p = \gamma(0)$		$v_p = [\gamma']_p = [\gamma]'_p$
$= \Big[\big(\gamma_1(t), \ldots, \gamma_n(t)\big)\Big]$		$= \Big[\big(\gamma'_1(t), \ldots, \gamma'_n(t)\big)\Big]_p$
		$= \Big[\big(\gamma_1(t), \ldots, \gamma_n(t)\big)\Big]'_p$
		$= \begin{bmatrix} \gamma'_1(t=0) \\ \vdots \\ \gamma'_n(t=0) \end{bmatrix}_p$
		$= \begin{bmatrix} \left.\frac{\partial \gamma_1(t)}{\partial t}\right\vert_{t=0} \\ \vdots \\ \left.\frac{\partial \gamma_n(t)}{\partial t}\right\vert_{t=0} \end{bmatrix}_p .$

The equivalence class of curves definition gives us another way to think about the chain rule,

Chain Rule for curves
$$[f \circ \phi^{-1} \circ \phi \circ \gamma]'_0 = [f \circ \phi^{-1}]'_{\phi\circ\gamma(0)} \cdot [\phi \circ \gamma]'_0 = [f \circ \phi^{-1}]'_{\phi(p)} \cdot [\phi \circ \gamma]'_0.$$

For the equivalence class of curves definition of vectors the push-forward of $[\gamma]_p$ by $T_p\phi$ is

Push-foraward of vector $[\gamma]_p$ by $T_p\phi$
$$T_p\phi \cdot [\gamma]_p = [\phi \circ \gamma]_{\phi(p)}.$$

Notationally there are several different ways to write this, all of which represent the same thing,

Ways to write push-foraward of vector $[\gamma]_p$by $T_p\phi$
$$\begin{aligned} T_p\phi \cdot [\gamma]_p &\equiv \phi_*(p)\big([\gamma]_p\big) \\ &\equiv \phi_*(p) \cdot [\gamma]_p \\ &\equiv \phi_*\big([\gamma]_p\big) \\ &\equiv [\phi \circ \gamma]_{\phi(p)}. \end{aligned}$$

In the equivalence class of curves definition of vectors the pull-back of a p-form ω is given by

Pull-back of p-form $\omega_{\phi(p)}$ by $T_p^*\phi$
$$\begin{aligned} \Big(T_p^*\phi \cdot \omega_{\phi(p)}\Big)\big([\gamma_1]_p, [\gamma_2]_p, \ldots, [\gamma_k]_p\big) &\equiv \omega_{\phi(p)}\Big(T_p\phi \cdot [\gamma_1]_p, T_p\phi \cdot [\gamma_2]_p, \ldots, T_p\phi \cdot [\gamma_k]_p\Big) \\ &= \omega_{\phi(p)}\Big([\phi \circ \gamma_1]_{\phi(p)}, [\phi \circ \gamma_2]_{\phi(p)}, \ldots, [\phi \circ \gamma_k]_{\phi(p)}\Big) \end{aligned}$$

and the pull-back of a zero form (function) $f : N \to \mathbb{R}$ is

Pull-back of zero-form f by $T_p^*\phi$
$$T_p^*\phi \cdot f \equiv \big(f \circ \phi\big)(p) = f\big(\phi(p)\big).$$

This definition looks very simple in traditional notation and without the base point included,

Pull-back of zero-form f by ϕ^*
$$\phi^* \cdot f \equiv f \circ \phi.$$

While this has been far from a complete introduction to manifolds, we have attempted to provide the basic ideas associated with a general manifold M. In particular, the idea of charts and atlases are the essential ingredients in the definition of a general manifold. We have also tried to show, particularly in Sect. 10.4, how important being careful with regards to charts is and how it does add an extra layer of complexity to the study of calculus on manifolds. It is usually not difficult to deal with, but attention does need to be paid to this extra layer of complexity. Since this is primarily a book on differential forms we have tried to avoid this extra layer of complexity throughout most of the book by sticking to the Euclidian manifolds $\mathbb{R}^n$, which all have an atlas consisting of one chart, the Euclidian chart that we are very familiar with.

10.5.2 References and Further Reading

There are literally hundreds of books on manifold theory and everything presented here is quite standard. If anything we have sacrificed mathematical rigor for readability and comprehensibility given this introduction to manifolds is so short. This chapter is meant to be simply a very basic introduction to some of the fundamentals, enough to get the reader started in more advanced books. The following five references are all good and are given in order of increasing levels of difficulty; Munkres [35], Walschap [47], Renteln [37], Martin [33], and Conlon [10]. In addition, the somewhat olders book by Bishop and Crittenden [5] and do Carmo [14] are very nice, if one can get hold of them.

10.5.3 Problems

Question 10.5 Write the following vectors on the manifold $\mathbb{R}^2$ as an equivalence class of curves in three different ways. (It may be useful to consider using trigonometric or exponential functions.)

$a)\ \begin{bmatrix} 2 \\ 3 \end{bmatrix}_{(2,1)} \quad b)\ \begin{bmatrix} -4 \\ 5 \end{bmatrix}_{(3,-2)} \quad c)\ \begin{bmatrix} 7 \\ -3 \end{bmatrix}_{(-3,-1)} \quad d)\ \begin{bmatrix} -3 \\ -7 \end{bmatrix}_{(-4,8)} \quad e)\ \begin{bmatrix} 3 \\ 1 \end{bmatrix}_{(3,-2)}$

Question 10.6 Using the vectors v_p from Question 10.5 written as the equivalence class of a curve, find $v_p[f_i]$ for the below functions f_i. Verify that you obtain the same answer regardless of which of the three representations of the curve you use.

$a)\ f_1(x, y) = xy^2 \quad b)\ f_2(x, y) = y + \cos(x) \quad c)\ f_3(x, y) = e^x + 3y - 4$
$d)\ f_4(x, y) = \sqrt{xy} + xy \quad e)\ f_5(x, y) = (x + 2y)^2$

Question 10.7 Write the following vectors on the manifold $\mathbb{R}^3$ as an equivalence class of curves in three different ways. (It may be useful to consider using trigonometric or exponential functions.)

$a)\ \begin{bmatrix} 4 \\ -2 \\ -8 \end{bmatrix}_{(5,-2,4)} \quad b)\ \begin{bmatrix} 3 \\ -6 \\ 5 \end{bmatrix}_{(3,-4,5)} \quad c)\ \begin{bmatrix} -4 \\ 7 \\ -1 \end{bmatrix}_{(2,4,-3)} \quad d)\ \begin{bmatrix} 5 \\ 9 \\ -9 \end{bmatrix}_{(6,-4,2)} \quad e)\ \begin{bmatrix} -5 \\ -2 \\ -3 \end{bmatrix}_{(7,3,-4)}$

Question 10.8 Using the vectors v_p from Question 10.7 written as the equivalence class of a curve, find $v_p[f_i]$ for the below functions f_i. Verify that you obtain the same answer regardless of which of the three representations of the curve you use.

$a)\ f_1(x, y, z) = xy^2z \quad b)\ f_2(x, y, z) = y + \cos(x) + \sin(z) \quad c)\ f_3(x, y, z) = e^x + ye^z$
$d)\ f_4(x, y, z) = \sqrt{xyz} + xyz \quad e)\ f_5(x, y, z) = (x + 2y - 3z)^2$

Question 10.9 Let $\phi : \mathbb{R}^2 \to \mathbb{R}^2$ be given by $\phi(x, y) = \big(2y, 3x\big)$. Using the equivalence class of curves version of vectors v_p in Question 10.5 find $T\phi \cdot v_p$. Verify your answer does not depend on which representative curve you use.

Question 10.10 Let $\phi : \mathbb{R}^2 \to \mathbb{R}^2$ be given by $\phi(x, y) = \big(3x + 2y, 3x - 2y\big)$. Using the equivalence class of curves version of vectors v_p in Question 10.5 find $T\phi \cdot v_p$. Verify your answer does not depend on which representative curve you use.

Question 10.11 Let $\phi : \mathbb{R}^2 \to \mathbb{R}^2$ be given by $\phi(x, y) = \big(xy, xy^3\big)$. Using the equivalence class of curves version of vectors v_p in Question 10.5 find $T\phi \cdot v_p$. Verify your answer does not depend on which representative curve you use.

Question 10.12 Let $\phi : \mathbb{R}^3 \to \mathbb{R}^3$ be given by $\phi(x, y, z) = \big(-2z, 3x, -4y\big)$. Using the equivalence class of curves version of vectors v_p in Question 10.7 find $T\phi \cdot v_p$. Verify your answer does not depend on which representative curve you use.

Question 10.13 Let $\phi : \mathbb{R}^3 \to \mathbb{R}^3$ be given by $\phi(x, y, z) = \big(3x + 2y + z, 3x - 2y + z, 5z + 7\big)$. Using the equivalence class of curves version of vectors v_p in Question 10.7 find $T\phi \cdot v_p$. Verify your answer does not depend on which representative curve you use.

Question 10.14 Let $\phi : \mathbb{R}^3 \to \mathbb{R}^3$ be given by $\phi(x, y, z) = \big(xy, yz, zx\big)$. Using the equivalence class of curves version of vectors v_p in Question 10.7 find $T\phi \cdot v_p$. Verify your answer does not depend on which representative curve you use.

Question 10.15 Let $\phi : \mathbb{R}^2 \to \mathbb{R}^2$ be given by $\phi(x, y) = \big(u(x, y), v(x, y)\big) = \big(2y, 3x\big)$ and let $\alpha = 3v\ du + 2u\ dv$. Evaluate $T^*\phi \cdot \alpha$ at each of the vectors in Question 10.5. Verify your answer does not depend on which representative curve you use.

Question 10.16 Let $\phi : \mathbb{R}^2 \to \mathbb{R}^2$ be given by $\phi(x, y) = \big(u(x, y), v(x, y)\big) = \big(3x + 2y, 3x - 2y\big)$ and let $\alpha = \cos(u)\ du + \sin(v)\ dv$. Evaluate $T^*\phi \cdot \alpha$ at each of the vectors in Question 10.5. Verify your answer does not depend on which representative curve you use.

Question 10.17 Let $\phi : \mathbb{R}^2 \to \mathbb{R}^2$ be given by $\phi(x, y) = \big(u(x, y), v(x, y)\big) = \big(xy, xy^3\big)$ and let $\alpha = (2u+v)\ du + u^2v^3\ dv$. Evaluate $T^*\phi \cdot \alpha$ at each of the vectors in Question 10.5. Verify your answer does not depend on which representative curve you use.

Question 10.18 Let $\phi : \mathbb{R}^3 \to \mathbb{R}^3$ be given by $\phi(x, y, z) = \big(u(x, y, z), v(x, y, z), w(x, y, z)\big) = \big(-2z, 3x, -4y\big)$ and let $\alpha = 2w\ du + 3v\ dv + 4u\ dw$. Evaluate $T^*\phi \cdot \alpha$ at each of the vectors in Question 10.7. Verify your answer does not depend on which representative curve you use.

Question 10.19 Let $\phi : \mathbb{R}^3 \to \mathbb{R}^3$ be given by $\phi(x, y, z) = \big(u(x, y, z), v(x, y, z), w(x, y, z)\big) = \big(3x + 2y + z, 3x - 2y + z, 5z + 7\big)$ and let $\alpha = \sin(v)\ du + \cos(u)\ dv + 3w\ dw$. Evaluate $T^*\phi \cdot \alpha$ at each of the vectors in Question 10.7. Verify your answer does not depend on which representative curve you use.

Question 10.20 Let $\phi : \mathbb{R}^3 \to \mathbb{R}^3$ be given by $\phi(x, y, z) = \big(u(x, y, z), v(x, y, z), w(x, y, z)\big) = \big(xy, yz, zx\big)$ and let $\alpha = (u + v)\ du + (v + w)\ dv + (w + u)\ dw$. Evaluate $T^*\phi \cdot \alpha$ at each of the vectors in Question 10.7. Verify your answer does not depend on which representative curve you use.

Chapter 11
Generalized Stokes' Theorem

The generalized version of Stokes' theorem, henceforth simply called Stokes' theorem, is an extraordinarily powerful and useful tool in mathematics. We have already encountered it in Sect. 9.5 where we found a common way of writing the fundamental theorem of line integrals, the vector calculus version of Stokes' theorem, and the divergence theorem as $\int_M d\alpha = \int_{\partial M} \alpha$. More precisely Stokes' theorem can be stated as follows.

Theorem 11.1 (Stokes' Theorem) *Let M be a smooth oriented n-dimensional manifold and let α be an $(n-1)$-form on M. Then*

$$\int_M d\alpha = \int_{\partial M} \alpha,$$

where ∂M is given the induced orientation.

Its proof appears in pretty much all the standard texts of differential geometry and is done with various levels of abstraction and rigor. As with much of this book we will try to strike a balance between understandability, rigor, and abstraction. In particular, the proof we give works for "nice" manifolds M. Hopefully at the end of this chapter you will understand the big picture well enough that you could follow a more mathematically rigorous version of the proof.

The general strategy that the proof follows is that it is proved for a very specific and straight-forward case, and we use the simple case to prove harder and harder cases. In essence we "bootstrap" our way up to the general version. Each section in this chapter, except the last section, is essentially one of the "bootstrap" steps. Understanding the general ideas is not so difficult, but getting a totally rigorous and airtight proof can sometimes feel like an exercise in minutia, trivialities, and nitpicking. The final section of the chapter uses the visualization techniques developed in Chap. 5 to visualize what Stokes' theorem is saying in three dimensions, at least insofar as these visualization techniques work.

Finally, proofs of Stokes' theorem are either based on k-dimensional unit cubes, which is the route we will follow, or sometimes on k-dimensional simplices, usually just called k-simplices. (This is one of those rare English words that has two plurals is use, so sometimes you will see the word simplexes as well.) Applications of Stoke's theorem to homology and cohomology make it useful to use k-simplicies. While it is possible to move between the unit k-cubes and k-simplices this does add an almost overwhelmingly tedious layer to the bootstrapping process, so we will not do that step here. In reality this step adds virtually nothing to understanding the big picture, though it is something you should be aware of and on the lookout for when you read other proofs of Stokes' theorem.

11.1 The Unit Cube I^k

The very first thing we need to do is discuss the orientations of unit k-cubes in $\mathbb{R}^n$, where $n \geq k$. Unit k-cubes are generally denoted by I^k. After looking at the orientations of k-cubes I^k we will then look at the boundary of I^k and the orientations of the boundary pieces. Once we have a firm handle on this we can then prove Stokes' theorem for the case of the unit k-cube I^k.

J. P. Fortney, *A Visual Introduction to Differential Forms and Calculus on Manifolds*,
https://doi.org/10.1007/978-3-319-96992-3_11

The 0-Cube I^0

First we consider the 0-cube. The 0-cube actually isn't very interesting, it is just a point. For convenience sake we will place our cube at the origin. So we can consider the 0-cube in $\mathbb{R}^n$ to simply be a point which we will denote as $\{0\}$. Being degenerate the 0-cube does not have an orientation in the same way that the other k-cubes do, though as we will see we can assign it either a positive or a negative orientation as necessary.

The 1-Cube I^1

Next we will look at the unit 1-cube $I^1 = \{x_1 \in \mathbb{R} \mid 0 \leq x_1 \leq 1\}$, which is Fig. 11.1. We will say that the orientation of $I^1 \subset \mathbb{R}^1$ is determined by the volume form dx_1. How does that work? Consider the unit vector $[1]$ based at the origin pointing in the direction of the 1-cube. We have $dx_1([1]) = 1$ which is positive, so this gives us the positive orientation on I^1. The opposite direction $-[1]$ gives $dx_1(-[1]) = -1$, which clearly gives us the negative orientation, see Fig. 11.2.

The 2-Cube I^2

Now let us consider the 2-cube $I^2 = \{(x_1, x_2) \mid 1 \leq x_i \leq 1,\ i = 1, 2\}$. Again, the volume form $dx_1 \wedge dx_2$ determines (often we say it induces) an orientation on I^2. Consider the two unit vectors

$$e^1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \text{ and } e^2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}.$$

The 2-cube is exactly the parallelepiped spanned by these two unit vectors as shown in Fig. 11.3. Using the two-dimensional volume form $dx_1 \wedge dx_2$ we have

$$dx_1 \wedge dx_2 \left(\begin{bmatrix} 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \end{bmatrix} \right) = \begin{vmatrix} 1 & 0 \\ 0 & 1 \end{vmatrix} = 1 > 0,$$

$$dx_1 \wedge dx_2 \left(\begin{bmatrix} 0 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \end{bmatrix} \right) = \begin{vmatrix} 0 & 1 \\ 1 & 0 \end{vmatrix} = -1 < 0.$$

Fig. 11.1 The unit cube in one dimension shown on the x_1-axis

Fig. 11.2 The two orientations of the one-dimensional unit cube. The orientations are determined using the one-dimensional volume form dx_1 and are generally called positive (left) and negative (right)

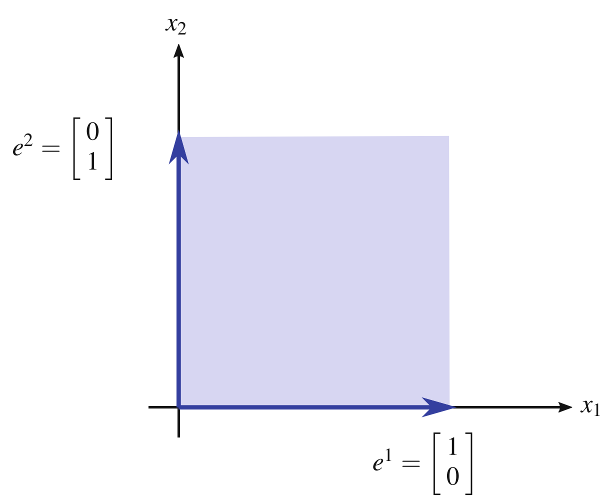


Fig. 11.3 The unit cube in two-dimensions is the parallelepiped spanned by the unit vectors e^1 and e^2

Thus the ordering of the unit vectors determines the orientation of the 2-cube. The ordering

$$\left\{ \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \end{bmatrix} \right\}$$

denotes the positive orientation and the orientation

$$\left\{ \begin{bmatrix} 0 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \end{bmatrix} \right\}$$

denotes the negative orientation. The positive orientation orientation is usually shown graphically as in Fig. 11.4 while the negative orientation is usually show graphically as in Fig. 11.5.

The 3-Cube I^3

Now we will take a look at the 3-cube $I^3 = \{(x_1, x_2, x_3) \mid 0 \leq x_i \leq 1, \ i = 1, 2, 3 \ \}$, shown in Fig. 11.6. The 3-cube is exactly the parallelepiped spanned by the vectors

$$e^1 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \ e^2 = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, e^3 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}.$$

The volume form $dx_1 \wedge dx_2 \wedge dx_2$ determines, or induces, an orientation on I^3. Clearly we have

$$dx_1 \wedge dx_2 \wedge dx_3 \left(\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \right) = \begin{vmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{vmatrix} = 1 > 0.$$

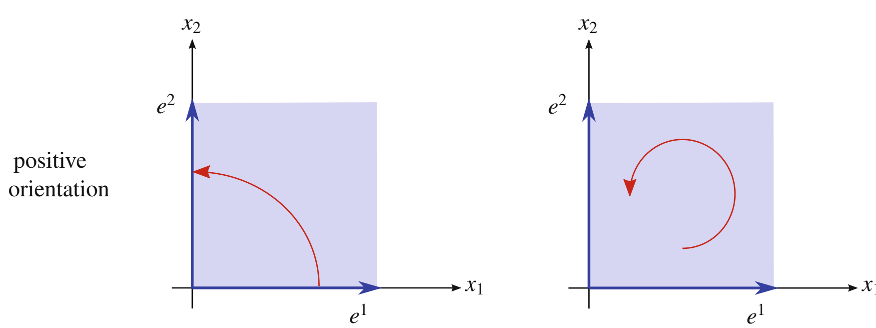


Fig. 11.4 Using the volume form $dx_1 \wedge dx_2$ we see that the ordering $\{e^1, e^2\}$, that is, e^1 followed by e^2, gives a positive orientation. Shown are two ways we can indicate a positive orientation

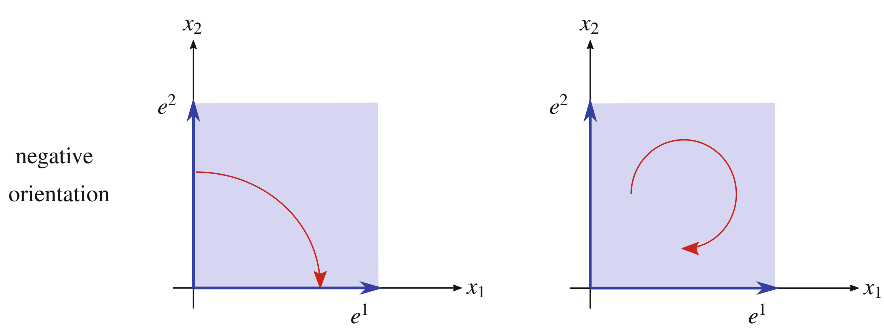


Fig. 11.5 Using the volume form $dx_1 \wedge dx_2$ we see that the ordering $\{e^2, e^1\}$, that is, e^2 followed by e^1, gives a negative orientation. Shown are two ways we can indicate a negative orientation

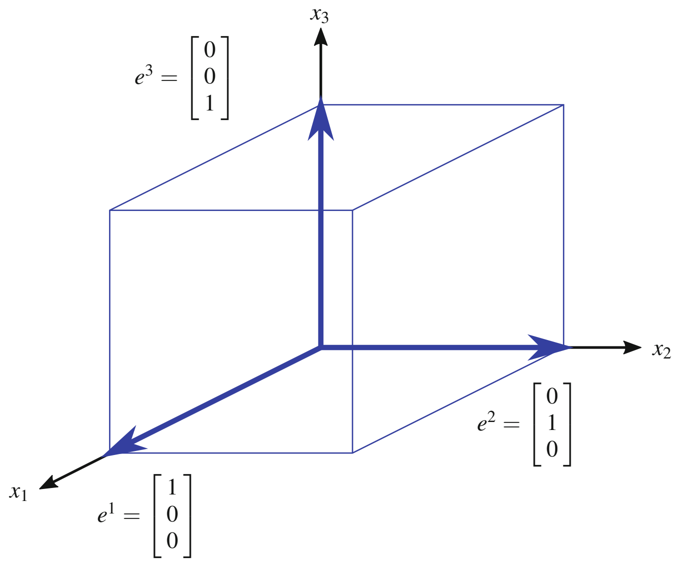


Fig. 11.6 The unit cube in three dimensions is the parallelepiped spanned by the three unit vectors e^1, e^2, and e^3

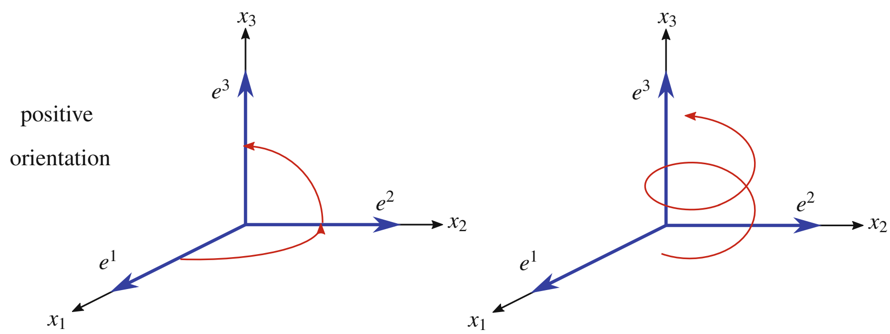


Fig. 11.7 Using the volume form $dx_1 \wedge dx_2 \wedge dx_3$ we see that the ordering $\{e^1, e^2, e^3\}$, that is, e^1 followed by e^2 followed by e^3, gives a positive orientation. Shown are two ways we can indicate a positive orientation. Positive orientation in three dimensions is generally called the "right-hand rule"

Thus the vector ordering

$$\left\{ \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \right\} = \{e^1, e^2, e^3\}$$

is considered a positive orientation. The positive orientation is usually show graphically in Fig. 11.7. This is exactly the so called "right-hand rule." But notice, we also have $dx_1 \wedge dx_2 \wedge dx_3(e_2, e_3, e_1) = 1$ and so $\{e_2, e_3, e_1\}$ also is a positive orientation. Similarly, $dx_1 \wedge dx_2 \wedge dx_3(e_3, e_1, e_2) = 1$ so $\{e_3, e_1, e_2\}$ is yet another positive orientation. So we see that there are three different orderings that give a positive orientation of I^3. Next notice that $dx_1 \wedge dx_2 \wedge dx_3(e_2, e_1, e_3) = -1 < 0$ so $\{e^2, e^1, e^3\}$ gives us a negative orientation, which you could consider a "left-hand rule." See Fig. 11.8.

Question 11.1 What are the three orderings that give a negative orientation to I^3?

So we can see that choosing an order of our basis elements is not a very efficient way to label or determine k-cube orientations. That is why we use the standard volume form $dx_1 \wedge dx_2 \wedge dx_3$, that is the volume form with the indices in numeric order. Of course, this actually also boils down to choosing an ordering for our vector space basis elements, but we have to start somewhere. The standard volume form induces positive and negative orientations on each unit cube, depending on how we order the unit vectors that are used to generate the cube.

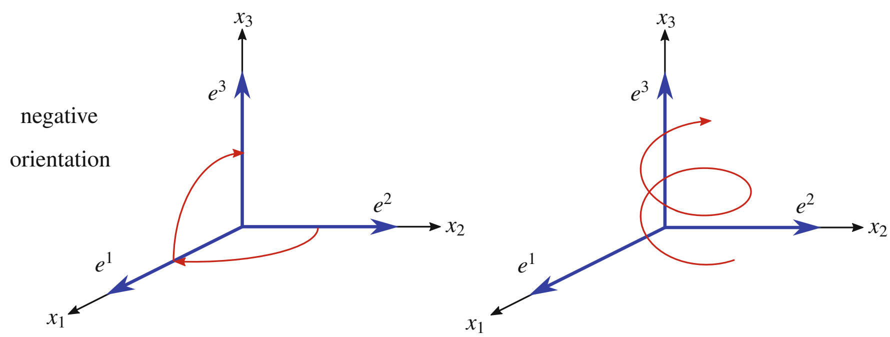


Fig. 11.8 Using the volume form $dx_1 \wedge dx_2 \wedge dx_3$ we see that the ordering $\{e^2, e^1, e^3\}$, that is, e^2 followed by e^1 followed by e^3, gives a negative orientation. Shown are two ways we can indicate a negative orientation. Negative orientation in three dimensions is generally called the "left-hand rule"

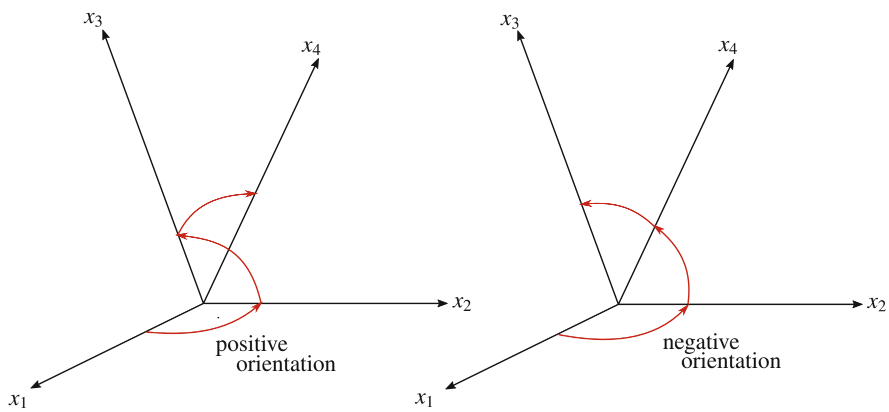


Fig. 11.9 An attempt to draw positive (left) and negative (right) orientations for a four dimensional space. Since we have four axes they can no longer be drawn as all perpendicular to the three others, but it should be understood that each of the depicted axis is in fact perpendicular to the remaining three

The k-Cube I^k

For I^k, where $k > 3$, it becomes impossible to draw either the k-cube or orientations, but the idea is intuitive enough. The k-cube is the parallelepiped spanned by the unit vectors $e^1, e^2, \ldots, e^k$. And given a volume form $d_1 \wedge dx_2 \wedge \cdots \wedge dx_k$ the volume form induces an orientation. For example, given the volume form $dx_1 \wedge dx_2 \wedge dx_3 \wedge dx_4$ any ordering of e_1, e_2, e_3, e_4 that results in a positive volume is called a positive orientation of I^4 and any ordering that results in a negative volume is called a negative orientation. We attempt to "draw" a positive and negative orientation for I^4 in Fig. 11.9.

Question 11.2 Find all the positive and negative orientations of e_1, e_2, e_3, e_4.

Boundaries of k-Cubes, ∂I^k

Now that we have some idea about the orientations of unit cubes we want to figure out what the boundaries of a unit cube are. Of course you have some intuitive idea of what boundaries are, but our goal is to make these ideas mathematically precise. We will also consider the idea of the orientations of the boundaries, which may be a little confusing at first.

Since the 0-cube is just a point, it does not have a boundary. We did not originally consider the orientation of a lone point I^0, but in the context of the boundary of I^1 we will assign an orientation to a point. First we will give the general definition of boundary of a k-cube, and then we will see some concrete examples of boundaries.

Given a k-cube I^k

$$I^k = \left\{ (x_1, \ldots, x_k) \;\middle|\; 0 \le x_i \le 1,\ i = 1, \ldots, k \right\}$$

we define the $(i, 0)$-face by letting $x_i = 0$, so we have

$$\begin{aligned} I^k_{(i,0)} &= I^k(x_1, \ldots, x_{i-1}, \underbrace{0}_{x_i=0}, x_{i+1}, \ldots, x_k) \\ &= \Big\{ (x_1, \ldots, x_{i-1}, 0, x_{i+1}, \ldots, x_k) \mid (x_1, \ldots, x_{i-1}, 0, x_{i+1}, \ldots, x_k) \in I^k \Big\} \end{aligned}$$

and the $(i, 1)$-face by letting $x_i = 1$, so we have

$$\begin{aligned} I^k_{(i,1)} &= I^k(x_1, \ldots, x_{i-1}, \underbrace{1}_{x_i=1}, x_{i+1}, \ldots, x_k) \\ &= \Big\{ (x_1, \ldots, x_{i-1}, 1, x_{i+1}, \ldots, x_k) \mid (x_1, \ldots, x_{i-1}, 1, x_{i+1}, \ldots, x_k) \in I^k \Big\}. \end{aligned}$$

Each face $I^k_{(i,a)}$, where $a = 0, 1$, will be given the orientation determined by $(-1)^{i+a}$. If $(-1)^{i+a} = 1$ the orientation is positive and if $(-1)^{i+a} = -1$ then the orientation is negative. We will see, through the use of examples, why this is the appropriate way to define the face orientation. The boundary of I^k, denoted as ∂I^k, is the collection of all the faces of I^k along with the determined orientation. Abstractly we will write

$$\partial I^k = \sum_{i=1}^{k} \sum_{a=0}^{1} (-1)^{i+a} I^k_{(i,a)},$$

though keep in mind that we are of course not adding these faces in the usual sense. This is simply our way of denoting the collection of faces, along with their respective orientations, that make up the boundary of I^k.

Finding ∂I^1

Let us see how this works for I^1. We have

$$\begin{aligned} I^1_{(1,0)} &= \big\{(0)\big\}, \\ I^1_{(1,1)} &= \big\{(1)\big\}. \end{aligned}$$

Next we see that $I^1_{(1,0)}$ has a negative orientation since $(-1)^{1+0} = -1$ and $I^1_{(1,1)}$ has a positive orientation since $(-1)^{1+1} = 1$. This is shown in Fig. 11.10. Finally, we would denote the boundary of I^1 by

$$\begin{aligned} \partial I^k &= \sum_{i=1}^{1} \sum_{a=0}^{1} (-1)^{i+a} I^i_{(i,a)} \\ &= I^1_{(1,1)} - I^1_{(1,0)}. \end{aligned}$$

Finding ∂I^2

Next we turn our attention to I^2. The 2-cube has four faces given by

$$\begin{aligned} I^2_{(1,0)} &= \Big\{ (0, x_2) \mid (0, x_2) \in I^2 \Big\}, \\ I^2_{(1,1)} &= \Big\{ (1, x_2) \mid (1, x_2) \in I^2 \Big\}, \end{aligned}$$

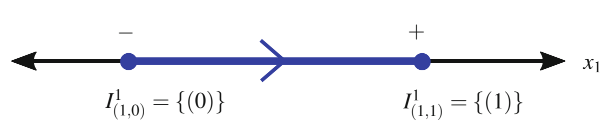


Fig. 11.10 The positively oriented one-cube I^1 along with the faces $I^1_{(1,0)}$ with a negative orientation, and the face $I^1_{(1,1)}$ with a positive orientation

$$I^2_{(2,0)} = \left\{ (x_1, 0) \mid (x_1, 0) \in I^2 \right\},$$
$$I^2_{(2,1)} = \left\{ (x_1, 1) \mid (x_1, 1) \in I^2 \right\}.$$

The picture of the 2-cube with boundary is given in Fig. 11.11. We now give the faces of the 2-cube the following orientations:

$$\begin{aligned}
I^2_{(1,0)} &\text{ has orientation determined by } (-1)^{1+0} = -1 \;\Rightarrow\; \text{negative},\\
I^2_{(1,1)} &\text{ has orientation determined by } (-1)^{1+1} = 1 \;\Rightarrow\; \text{positive},\\
I^2_{(2,0)} &\text{ has orientation determined by } (-1)^{2+0} = 1 \;\Rightarrow\; \text{positive},\\
I^2_{(2,1)} &\text{ has orientation determined by } (-1)^{2+1} = -1 \;\Rightarrow\; \text{negative},
\end{aligned}$$

which gives us the picture of the 2-cube with oriented boundary as in Fig. 11.12. Notice that by giving the faces $I^2_{(i,a)}$ the orientations determined by $(-1)^{i+a}$ our boundary orientations are somehow consistent, they "go around" the 2-cube in a consistent way. We denote the boundary of the two-cube by

$$\begin{aligned}
\partial I^2 &= \sum_{i=1}^{2} \sum_{a=0}^{1} (-1)^{i+a} I^2_{(i,a)}\\
&= -I^2_{(1,0)} + I^2_{(1,1)} + I^2_{(2,0)} - I^2_{(2,1)}.
\end{aligned}$$

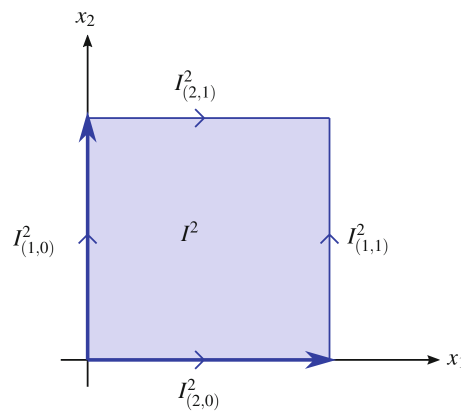


Fig. 11.11 The two-cube I^2 along with its boundary, which is given by the four faces $I^2_{(1,0)}$, $I^2_{(1,1)}$, $I^2_{(2,0)}$, and $I^2_{(2,1)}$. The four faces are shown with positive orientation

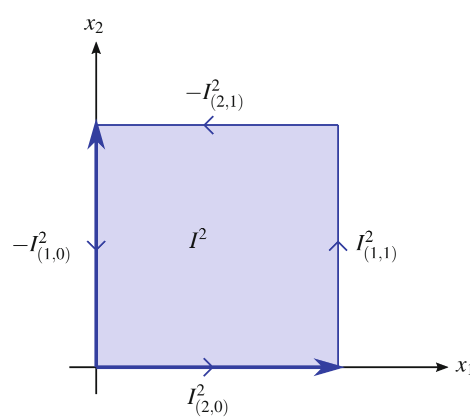


Fig. 11.12 The two-cube I^2 along with its four boundary faces, here shown with the orientations determined by the formula. For example, face $I^2_{(i,a)}$ has orientation determined by whether $(-1)^{i+a}$ is positive or negative

Question 11.3 The 2-cube in Fig. 11.11 is drawn in such a way that the four boundaries are show with positive orientation by the arrows. Explain why the arrows are the direction they are.

Finding $\partial\partial I^1$

Taking a step back we can see that that $\partial\partial I^1 = 0$ as well. We had $\partial I^1 = \{(1)\} - \{(0)\}$, where (0) and (1) are points on $\mathbb{R}^1$. Since a point does not have a boundary we say $\partial\{(0)\} = 0$ and $\partial\{(1)\} = 0$ so we get

$$\partial\partial I^1 = \partial\{(1)\} - \partial\{(0)\} = 0 - 0 = 0.$$

Finding $\partial\partial I^2$

Now, we will go one step further and find the boundaries of the boundary elements. That is, we will find $\partial\left(-I^2_{(1,0)}\right)$, $\partial I^2_{(1,1)}$, $\partial I^2_{(2,0)}$, and $\partial\left(-I^2_{(2,1)}\right)$. We begin with $\partial I^2_{(2,0)}$. We see that

$$\left(I^2_{(2,0)}\right)^1_{(1,0)} = \big\{(0,0)\big\} \text{ which has orientation } (-1)^{1+0} = -1 \;\Rightarrow\; \text{negative,}$$
$$\left(I^2_{(2,0)}\right)^1_{(1,1)} = \big\{(1,0)\big\} \text{ which has orientation } (-1)^{1+1} = 1 \;\Rightarrow\; \text{positive.}$$

We show $\partial I^2_{(2,0)}$ in Fig. 11.13.

In order to find $\partial I^2_{(1,1)}$ we need to pay special attention to the subscripts. Recall that

$$I^2_{(1,1)} = \left\{ (\underbrace{1}_{x_1=1}, x_2) \;\middle|\; (1, x_2) \in I^2 \right\}.$$

Notice that $I^2_{(1,1)}$ is now in fact a one-cube, that is, $\left(I^2_{(1,1)}\right)^1$. So when we want to take the $(1,0)$ boundary of this one-cube we have that the variable this 1 refers to is the first variable in the one-cube, which is actually still being labeled as x_2,

$$\partial\left(I^2_{(1,1)}\right)^1_{(1,0)} = \left\{ (\underbrace{1}_{x_1=1}, \underbrace{x_2}_{x_1=0}) \right\} = \big\{(1,0)\big\}.$$

Similarly,

$$\partial\left(I^2_{(1,1)}\right)^1_{(1,1)} = \left\{ (\underbrace{1}_{x_1=1}, \underbrace{x_2}_{x_1=1}) \right\} = \big\{(1,1)\big\}.$$

For the orientations,

$$\left(I^2_{(1,1)}\right)^1_{(1,0)} = \big\{(1,0)\big\} \text{ which has orientation } (-1)^{1+0} = -1 \Rightarrow \text{ negative,}$$
$$\left(I^2_{(1,1)}\right)^1_{(1,1)} = \big\{(1,1)\big\} \text{ which has orientation } (-1)^{1+1} = 1 \Rightarrow \text{ positive.}$$

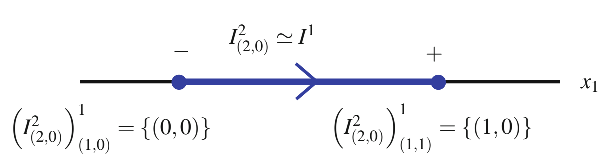


Fig. 11.13 The boundary of $I^2_{(2,0)}$ is given by $\left(I^2_{(2,0)}\right)^1_{(1,0)}$ with negative orientation and $\left(I^2_{(2,0)}\right)^1_{(1,1)}$ with positive orientation

We can then draw $\partial I^2_{(1,1)}$ as in Fig. 11.14. Now for $\partial\left(-I^2_{(2,1)}\right)$ we see that

$$\left(-I^2_{(2,1)}\right)^1_{(1,0)} = \big\{(0,1)\big\} \text{ which has orientation } -(-1)^{1+0} = 1 \Rightarrow \text{ positive},$$

$$\left(-I^2_{(2,1)}\right)^1_{(1,1)} = \big\{(1,1)\big\} \text{ which has orientation } -(-1)^{1+1} = -1 \Rightarrow \text{ negative},$$

thereby allowing us to draw $\partial\left(-I^2_{(2,1)}\right)$ as in Fig. 11.15. Finally, for $\partial\left(-I^2_{(1,0)}\right)$ we see that

$$\left(-I^2_{(1,0)}\right)^1_{(1,0)} = \big\{(0,0)\big\} \text{ which has orientation } -(-1)^{1+0} = 1 \Rightarrow \text{ positive},$$

$$\left(-I^2_{(1,0)}\right)^1_{(1,1)} = \big\{(0,1)\big\} \text{ which has orientation } -(-1)^{1+1} = -1 \Rightarrow \text{ negative},$$

thereby allowing us to draw $\partial\left(-I^2_{(1,0)}\right)$ as in Fig. 11.16.

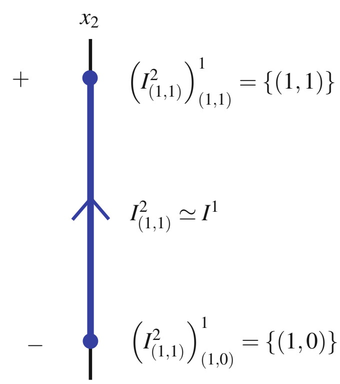


Fig. 11.14 The boundary of $I^2_{(1,1)}$ is given by $\left(I^2_{(1,1)}\right)^1_{(1,0)}$ with negative orientation and $\left(I^2_{(1,1)}\right)^1_{(1,1)}$ with positive orientation

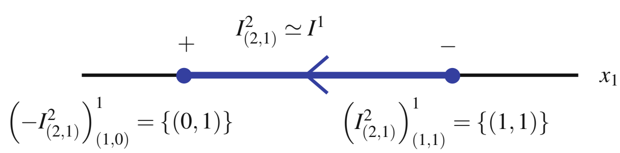


Fig. 11.15 The boundary of $-I^2_{(2,1)}$ is given by $\left(-I^2_{(2,1)}\right)^1_{(1,0)}$ with positive orientation and $\left(-I^2_{(2,1)}\right)^1_{(1,1)}$ with negative orientation

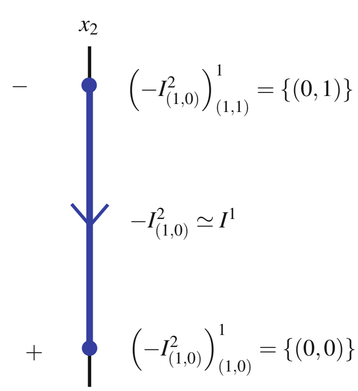


Fig. 11.16 The boundary of $-I^2_{(1,0)}$ is given by $\left(-I^2_{(1,0)}\right)^1_{(1,0)}$ with positive orientation and $\left(-I^2_{(1,0)}\right)^1_{(1,1)}$ with negative orientation

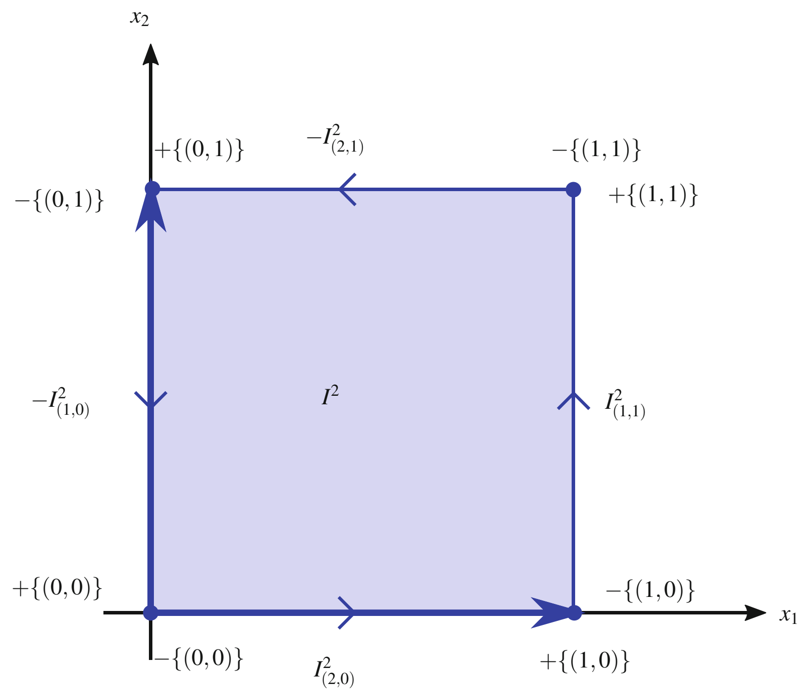


Fig. 11.17 To get a picture for $\partial\partial I^2$ we put together Figs. 11.13, 11.14, 11.15, and 11.16

Putting this all together, for the $\partial\partial I^2$ we have the picture in Fig. 11.17, which gives us

$$\begin{aligned}\partial\partial I^2 &= \partial\left(-I^2_{(1,0)}\right) + \partial I^2_{(1,1)} + \partial I^2_{(2,0)} + \partial\left(-I^2_{(2,1)}\right)\\&= \big\{(0,0)\big\} - \big\{(0,1)\big\} - \big\{(1,0)\big\} + \big\{(1,1)\big\} - \big\{(0,0)\big\} + \big\{(1,0)\big\} + \big\{(0,1)\big\} - \big\{(1,1)\big\}\\&= 0.\end{aligned}$$

Thus we see that the boundary elements and their orientations are defined in such a way that the boundary of the boundary is zero, which makes intuitive sense. After all, a closed curve does not have a boundary, which is exactly what the boundary of the two-cube is.

Finding ∂I^3

Now we will consider I^3. First we will find the faces of I^3 and then we will find the orientations of the faces, which we will use to find ∂I^3. I^3 clearly has six faces, given by

$$\begin{aligned}I^3_{(1,0)} &= \{(0, x_2, x_3) \mid 0 \le x_2, x_3 \le 1\} = \text{ back face},\\I^3_{(1,1)} &= \{(1, x_2, x_3) \mid 0 \le x_2, x_3 \le 1\} = \text{ front face},\\I^3_{(2,0)} &= \{(x_1, 0, x_3) \mid 0 \le x_1, x_3 \le 1\} = \text{ left face},\\I^3_{(2,1)} &= \{(x_1, 1, x_3) \mid 0 \le x_1, x_3 \le 1\} = \text{ right face},\\I^3_{(3,0)} &= \{(x_1, x_2, 0) \mid 0 \le x_1, x_2 \le 1\} = \text{ bottom face},\\I^3_{(3,1)} &= \{(x_1, x_2, 1) \mid 0 \le x_1, x_2 \le 1\} = \text{ top face},\end{aligned}$$

and shown in Fig. 11.18.

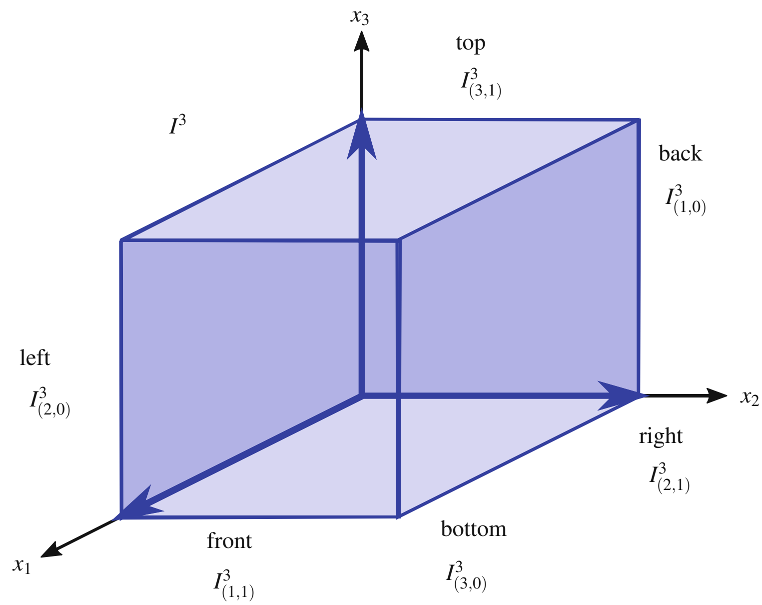


Fig. 11.18 The three-cube I^3 along with its six faces

We first show the positive orientations of all six faces of I^3. The volume form that determines the orientation of the front and back faces is $dx_2 \wedge dx_3$ since the front and back faces are in the (x_2, x_3)-plane. Consider the two vectors

$$\begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} \text{ and } \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$$

at the point $p = (0, 0, 0)$ for the back face and the point $q = (1, 0, 0)$ for the front face. We have

$$dx_2 \wedge dx_3(p) \left(\begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}_p, \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}_p \right) = \begin{vmatrix} 1 & 0 \\ 0 & 1 \end{vmatrix} = 1,$$

$$dx_2 \wedge dx_3(q) \left(\begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}_q, \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}_q \right) = \begin{vmatrix} 1 & 0 \\ 0 & 1 \end{vmatrix} = 1$$

for both points. The positive orientations of the front and back faces are shown in Fig. 11.19.

The volume form that determines positive orientation for the left and right is $dx_1 \wedge dx_3$ since the left and right faces are in the (x_1, x_3)-plane. Exactly as before consider the two vectors

$$\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \text{ and } \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$$

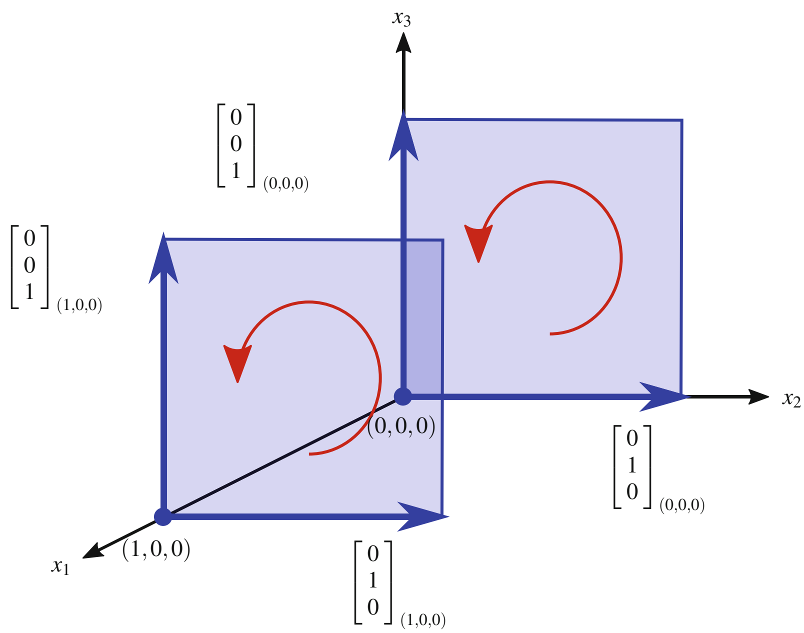


Fig. 11.19 The front and back face of I^3 shown with positive orientation

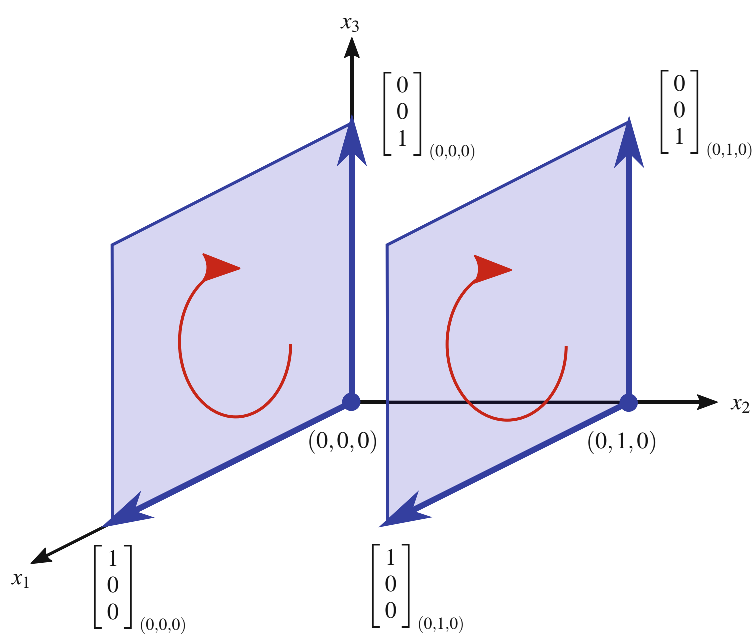


Fig. 11.20 The left and right face of I^3 shown with positive orientation

at the point $p = (0, 0, 0)$ for the left face and the point $q = (0, 1, 0)$ for the right face. We have

$$dx_1 \wedge dx_3(p)\left(\begin{bmatrix}1\\0\\0\end{bmatrix}_p, \begin{bmatrix}0\\0\\1\end{bmatrix}_p\right) = \begin{vmatrix}1&0\\0&1\end{vmatrix} = 1,$$

$$dx_1 \wedge dx_3(q)\left(\begin{bmatrix}1\\0\\0\end{bmatrix}_q, \begin{bmatrix}0\\0\\1\end{bmatrix}_q\right) = \begin{vmatrix}1&0\\0&1\end{vmatrix} = 1$$

for both points The positive orientations of the left and right faces are shown in Fig. 11.20.

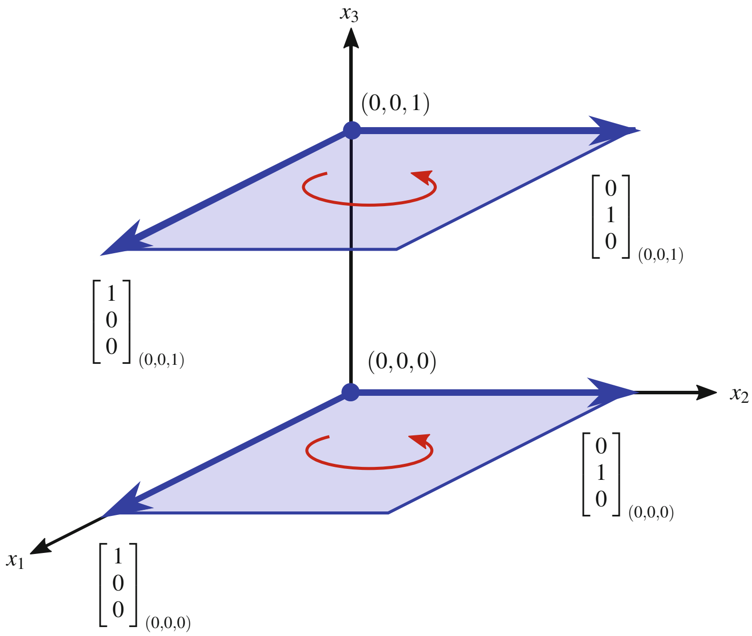


Fig. 11.21 The top and bottom face of I^3 shown with positive orientation

Question 11.4 Do this analysis for the top and bottom faces to get the positive orientations of those faces shown in Fig. 11.21.

Now we figure out the orientations of the six faces.

$$
\begin{aligned}
I^3_{(1,0)} &\text{ has orientation determined by } (-1)^{1+0} = -1 \;\Rightarrow\; \text{negative},\\
I^3_{(1,1)} &\text{ has orientation determined by } (-1)^{1+1} = +1 \;\Rightarrow\; \text{positive},\\
I^3_{(2,0)} &\text{ has orientation determined by } (-1)^{2+0} = +1 \;\Rightarrow\; \text{positive},\\
I^3_{(2,1)} &\text{ has orientation determined by } (-1)^{2+1} = -1 \;\Rightarrow\; \text{negative},\\
I^3_{(3,0)} &\text{ has orientation determined by } (-1)^{3+0} = -1 \;\Rightarrow\; \text{negative},\\
I^3_{(3,1)} &\text{ has orientation determined by } (-1)^{3+1} = +1 \;\Rightarrow\; \text{positive}.
\end{aligned}
$$

Putting everything together we can write down the boundary of I^3 as

$$
\begin{aligned}
\partial I^3 &= \sum_{i=1}^{3}\sum_{a=0}^{1}(-1)^{i+a} I^3_{(i,a)}\\
&= -I^3_{(1,0)} + I^3_{(1,1)} + I^3_{(2,0)} - I^3_{(2,1)} - I^3_{(3,0)} + I^3_{(3,1)}.
\end{aligned}
$$

Question 11.5 Sketch the unit cube I^3 including the appropriate orientations of the various faces. Do the orientations seem to fit together?

Finding $\partial\partial I^3$

Now we want to find the boundary of ∂I^3. Recall, ∂I^3 consists of six different two-cubes. We will begin by finding the edges of the bottom face $I^3_{(3,0)}$

$$
\begin{aligned}
\left(I^3_{(3,0)}\right)^2_{(1,0)} &= \{\, (0, x_2, 0) \mid 0 \le x_2 \le 1 \,\} \text{ with orientation } (-1)^{3+0}(-1)^{1+0} = 1,\\
\left(I^3_{(3,0)}\right)^2_{(1,1)} &= \{\, (1, x_2, 0) \mid 0 \le x_2 \le 1 \,\} \text{ with orientation } (-1)^{3+0}(-1)^{1+1} = -1,
\end{aligned}
$$

$$\left(I^3_{(3,0)}\right)^2_{(2,0)} = \{\ (x_1, 0, 0) \mid 0 \le x_1 \le 1\ \} \text{ with orientation } (-1)^{3+0}(-1)^{2+0} = -1,$$

$$\left(I^3_{(3,0)}\right)^2_{(2,1)} = \{\ (x_1, 1, 0) \mid 0 \le x_1 \le 1\ \} \text{ with orientation } (-1)^{3+0}(-1)^{2+1} = 1.$$

Similarly the edges of the top face $I^3_{(3,1)}$ are

$$\left(I^3_{(3,1)}\right)^2_{(1,0)} = \{\ (0, x_2, 1) \mid 0 \le x_2 \le 1\ \} \text{ with orientation } (-1)^{3+1}(-1)^{1+0} = -1,$$

$$\left(I^3_{(3,1)}\right)^2_{(1,1)} = \{\ (1, x_2, 1) \mid 0 \le x_2 \le 1\ \} \text{ with orientation } (-1)^{3+1}(-1)^{1+1} = 1,$$

$$\left(I^3_{(3,1)}\right)^2_{(2,0)} = \{\ (x_1, 0, 1) \mid 0 \le x_1 \le 1\ \} \text{ with orientation } (-1)^{3+1}(-1)^{2+0} = 1,$$

$$\left(I^3_{(3,1)}\right)^2_{(2,1)} = \{\ (x_1, 1, 1) \mid 0 \le x_1 \le 1\ \} \text{ with orientation } (-1)^{3+1}(-1)^{2+1} = -1.$$

We show the bottom and the top faces of I^3 along with the orientation of their boundary edges in Fig. 11.22.

Make sure that you understand how the notation works above. The bottom/top case is the easiest case. For the back/front and the left/right there is something a little bit tricky going on with the notation. Let us first of all consider the back edge $I^3_{(1,0)} = \{(0, x_2, x_1) | 0 \le x_2, x_3 \le 1\}$, which is a 2-cube. Let us first consider $\left(I^3_{(1,0)}\right)^2_{(1,0)}$. For the 3-cube the variables were (x_1, x_2, x_3). The $(1, 0)$ in $I^3_{(1,0)}$ means that the first variable, x_1, is replaced by 0. But now let us consider the back face of the 3-cube, which is a 2-cube. $\left(I^3_{(1,0)}\right)^2$. For this 2-cube the variables are (x_2, x_3), so the second $(1, 0)$ in $\left(I^3_{(1,0)}\right)^2_{(1,0)}$ means that the first variable is replaced by a 0, but in this case the first variable is x_2. So we have

$$\left(I^3_{(1,0)}\right)^2_{(1,0)} = \{\ (0, 0, x_3) \mid 0 \le x_3 \le 1\ \}.$$

The orientation is $(-1)^{1+0}(-1)^{1+0} = 1$, which is positive. Similarly, we have

$$\left(I^3_{(1,0)}\right)^2_{(1,1)} = \{\ (0, 1, x_3) \mid 0 \le x_3 \le 1\ \} \text{ with orientation } (-1)^{1+0}(-1)^{1+1} = -1 \ \Rightarrow \ \text{negative},$$

$$\left(I^3_{(1,0)}\right)^2_{(2,0)} = \{\ (0, x_2, 0) \mid 0 \le x_2 \le 1\ \} \text{ with orientation } (-1)^{1+0}(-1)^{2+0} = -1 \ \Rightarrow \ \text{negative},$$

$$\left(I^3_{(1,0)}\right)^2_{(2,1)} = \{\ (0, x_2, 1) \mid 0 \le x_2 \le 1\ \} \text{ with orientation } (-1)^{1+0}(-1)^{2+1} = 1 \ \Rightarrow \ \text{positive}.$$

The back face of the 3-cube along with the boundary orientations is shown in Fig. 11.23.

Question 11.6 Find the four boundary edges of the front $I^3_{(1,1)} = \{\ (1, x_2, x_3) \mid 0 \le x_2, x_3 \le 1\ \}$ and find the orientations of these edges. Your answers should agree with Fig. 11.23.

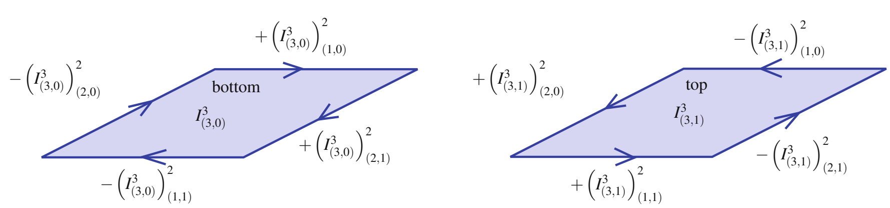


Fig. 11.22 The bottom and top faces of the three-cube shown along with the orientations of their boundaries

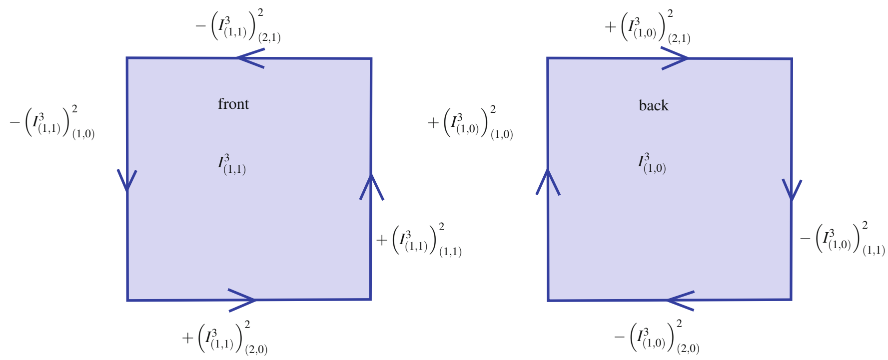


Fig. 11.23 The front and back faces of the three-cube I^3 shown along with the orientations of their boundaries

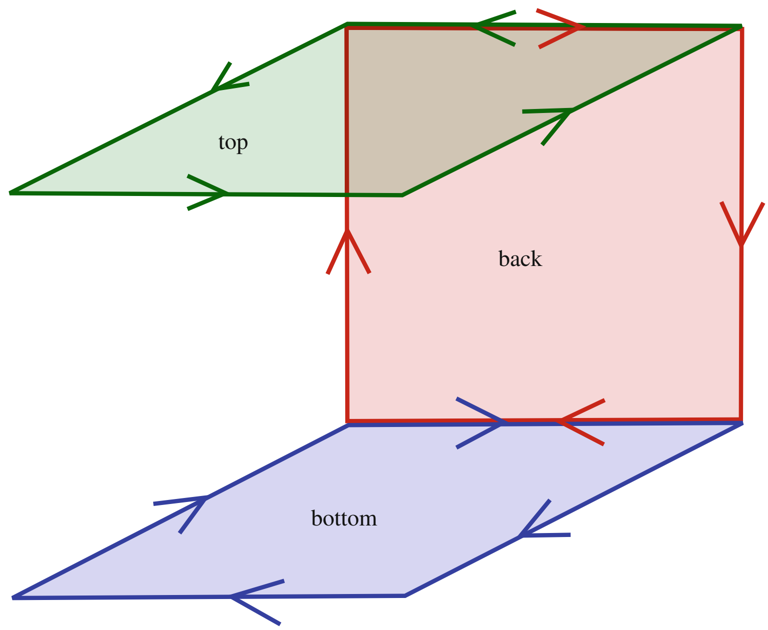


Fig. 11.24 The top, bottom, and back faces of the three-cube I^3 shown along with their boundaries. Notice where the boundaries of the three faces are the same, when the orientations are taken into account we have $\left(I^3_{(3,1)}\right)^2_{(1,0)} = -\left(I^3_{(1,0)}\right)^2_{(2,1)}$ and $\left(I^3_{(3,0)}\right)^2_{(1,0)} = -\left(I^3_{(1,0)}\right)^2_{(2,0)}$

Question 11.7 Show that when orientations are taken into account $\left(I^3_{(3,1)}\right)^2_{(1,0)} = -\left(I^3_{(1,0)}\right)^2_{(2,1)}$ and $\left(I^3_{(3,0)}\right)^2_{(1,0)} = -\left(I^3_{(1,0)}\right)^2_{(2,0)}$. See Fig. 11.24.

Question 11.8 Show that when orientations are taken into account $\left(I^3_{(3,1)}\right)^2_{(1,1)} = -\left(I^3_{(1,1)}\right)^2_{(2,1)}$ and $\left(I^3_{(3,0)}\right)^2_{(1,1)} = -\left(I^3_{(1,1)}\right)^2_{(2,0)}$. See Fig. 11.25.

Question 11.9 Find the four boundary edges of the left face $I^3_{(2,0)} = \{ (x_1, 0, x_3) \mid 0 \le x_1, x_3 \le 1 \}$ and the right face $I^3_{(2,1)} = \{ (x_1, 1, x_3) \mid 0 \le x_1, x_3 \le 1 \}$ and find the orientations of these eight boundary edges. Sketch the top, left, and bottom together and then sketch the top, right, and bottom together.

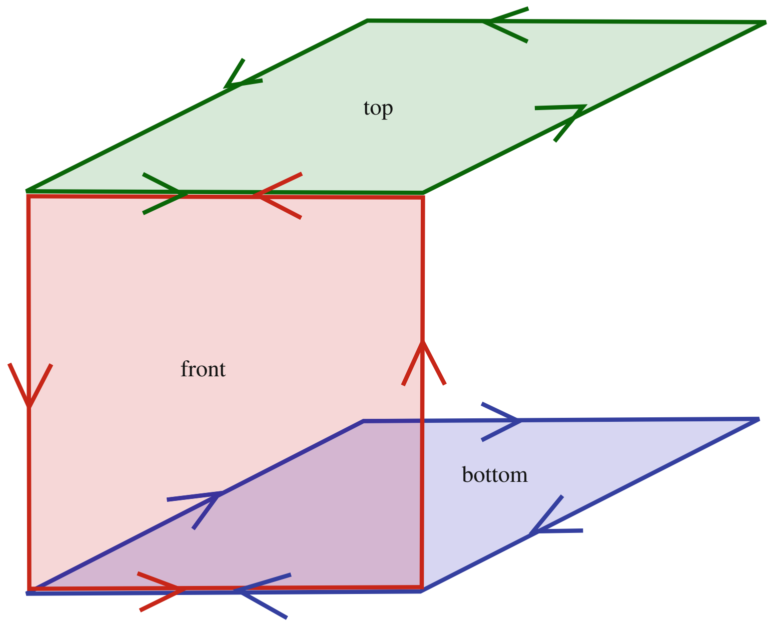


Fig. 11.25 The top, bottom, and back faces of the three-cube I^3 shown along with their boundaries. Notice where the boundaries of the three faces are the same, when the orientations are taken into account we have $\left(I^3_{(3,1)}\right)^2_{(1,1)} = -\left(I^3_{(1,1)}\right)^2_{(2,1)}$ and $\left(I^3_{(3,0)}\right)^2_{(1,1)} = -\left(I^3_{(1,1)}\right)^2_{(2,0)}$

You should be able to see, by inspecting the various sketches, that the following 1-cubes are identified

$$\left(I^3_{(2,0)}\right)^2_{(1,0)} = \left(I^3_{(1,0)}\right)^2_{(1,0)},$$
$$\left(I^3_{(2,1)}\right)^2_{(1,1)} = \left(I^3_{(1,1)}\right)^2_{(1,0)},$$
$$\left(I^3_{(2,0)}\right)^2_{(2,0)} = \left(I^3_{(3,0)}\right)^2_{(2,0)},$$
$$\left(I^3_{(2,0)}\right)^2_{(2,1)} = \left(I^3_{(3,1)}\right)^2_{(2,0)}.$$

Question 11.10 Put everything together to show that $\partial\partial I^3 = 0$.

Finding $\partial\partial I^k$
Consider the case where $i \le j \le k-1$ and $a, b = 0, 1$,

$$\begin{aligned}
\left(I^k_{(i,a)}\right)^{k-1}_{(j,b)} &= \left(\left\{(x_1, \dots, x_{i-1}, a, x_{i+1}, \dots, x_k)\right\}\right)^{k-1}_{(j,b)} \\
&= \left(\left\{(x_1, \dots, x_{i-1}, a, \underbrace{x_i, \dots, x_{k-1}}_{\text{relabled}})\right\}\right)^{k-1}_{(j,b)} \quad \begin{array}{c}\text{relabeled to}\\ (k-1)\text{-cube}\\ \text{variable names}\end{array} \\
&= \left\{(x_1, \dots, x_{i-1}, a, x_i, \dots, x_{j-1}, b, x_{j+1}, \dots, x_{k-1})\right\} \\
&= \left\{(x_1, \dots, x_{i-1}, a, \underbrace{x_{i+1}, \dots, x_j, b, x_{j+2}, \dots, x_k}_{\text{relabled}})\right\} \quad \begin{array}{c}\text{relabeled to}\\ k\text{-cube}\\ \text{variable names}\end{array}
\end{aligned}$$

so b is in the $j+1$ slot when we use the k-cube variable names. We also have

$$\begin{aligned}\left(I^k_{(j+1,b)}\right)^{k-1}_{(i,a)} &= \left(\left\{(x_1,\ldots,x_j,b,x_{j+2},\ldots,x_k)\right\}\right)^{k-1}_{(i,a)} \\ &= \left\{(x_1,\ldots,x_{i-1},a,x_{i+1},\ldots,x_j,b,x_{j+2},\ldots,x_k)\right\}. \quad \begin{array}{c}\text{no relabeling}\\ \text{necessary}\\ \text{since } i\le j\end{array}\end{aligned}$$

Hence for $i \le j \le k-1$ we have

$$\left(I^k_{(i,a)}\right)^{k-1}_{(j,b)} = \left(I^k_{(j+1,b)}\right)^{k-1}_{(i,a)}.$$

Question 11.11 Explain why the condition that $i \le j \le k-1$ is not restrictive. That is, that every boundary element of ∂I^k is included.

Question 11.12 Show that for general I^k we have $\partial\partial I^k = 0$.

11.2 The Base Case: Stokes' Theorem on I^k

With all of this necessary background material covered, the actual proof of Stokes' theorem is not actually so difficult. As we stated earlier, we will be "bootstrapping" ourselves up to the fully general case. The first step is to prove Stoke's theorem on the k-cube I^k. Given a $(k-1)$-form α defined on a neighborhood of the unit k-cube I^k, we want to show that

$$\int_{I^k} d\alpha = \int_{\partial I^k} \alpha.$$

First, α is a $(k-1)$-form defined on some neighborhood of $I^k \subset \mathbb{R}^k$, so alpha has the general form

$$\begin{aligned}\alpha &= \alpha_1(x_1,\ldots,x_k)\, dx_2 \wedge dx_3 \wedge \cdots \wedge dx_k \\ &\quad + \alpha_2(x_1,\ldots,x_k)\, dx_1 \wedge dx_3 \wedge \cdots \wedge dx_k \\ &\quad + \cdots \\ &\quad + \alpha_k(x_1,\ldots,x_k)\, dx_1 \wedge dx_2 \wedge \cdots \wedge dx_{k-1} \\ &= \sum_{i=1}^k \alpha_i(x_1,\ldots,x_k)\, dx_1 \wedge \cdots \wedge \widehat{dx_i} \wedge \cdots \wedge dx_k,\end{aligned}$$

where $\widehat{dx_i}$ means that the dx_i is omitted and $\alpha_i : \mathbb{R}^k \to \mathbb{R}$, $0 \le i \le k$, are real-valued functions defined in a neighborhood of I^k. We know that exterior differentiation d is linear, which means that

$$d\left(\sum_{i=1}^k \alpha_i\, dx_1 \wedge \cdots \wedge \widehat{dx_i} \wedge \cdots \wedge dx_k\right) = \sum_{i=1}^k d\left(\alpha_i\, dx_1 \wedge \cdots \wedge \widehat{dx_i} \wedge \cdots \wedge dx_k\right).$$

We also know that integration is linear, so we have

$$\int_{I^k} \sum_{i=1}^k d\left(\alpha_i\, dx_1 \wedge \cdots \wedge \widehat{dx_i} \wedge \cdots \wedge dx_k\right) = \sum_{i=1}^k \int_{I^k} d\left(\alpha_i\, dx_1 \wedge \cdots \wedge \widehat{dx_i} \wedge \cdots \wedge dx_k\right).$$

Now let us consider the first term of the $(k-1)$-form α, which is $\alpha_1\ dx_2 \wedge \ldots \wedge dx_k$. Then we have

$$
\begin{aligned}
&\int_{I^k} d\left(\alpha_1\ dx_2 \wedge \cdots \wedge dx_k\right) \\
&= \int_{I^k} \frac{\partial \alpha_1}{\partial x_1}\ dx_1 \wedge dx_2 \wedge \cdots \wedge dx_k \\
&= \int_{x_k=0}^{1} \cdots \int_{x_2=0}^{1} \int_{x_1=0}^{1} \frac{\partial \alpha_1}{\partial x_1}\ dx_1\ dx_2\ \cdots\ dx_k \quad \text{traditional integration (orientation does not matter)} \\
&= \int_{x_k=0}^{1} \cdots \int_{x_2=0}^{1} \Big[\alpha_1(x_1, x_2, \ldots, x_k)\Big]_0^1\ dx_2\ \cdots\ dx_k \quad \text{Fundamental Theorem of Calculus} \\
&= \int_{x_k=0}^{1} \cdots \int_{x_2=0}^{1} \Big[\alpha_1(1, x_2, \ldots, x_k) - \alpha_1(0, x_2, \ldots, x_k)\Big]\ dx_2\ \cdots\ dx_k \\
&= \underbrace{\int_{x_k=0}^{1} \cdots \int_{x_2=0}^{1} \alpha_1(1, x_2, \ldots, x_k)\ dx_2\ \cdots\ dx_k}_{\int \text{ of } \alpha_1 \text{ restricted to face } I^k_{(1,1)}} \\
&\qquad - \underbrace{\int_{x_k=0}^{1} \cdots \int_{x_2=0}^{1} \alpha_1(0, x_2, \ldots, x_k)\ dx_2\ \cdots\ dx_k}_{\int \text{ of } \alpha_1 \text{ restricted to face } I^k_{(1,0)}} \\
&= \int_{I^k_{(1,1)}} \alpha_1\ dx_2 \wedge \cdots \wedge dx_k - \int_{I^k_{(1,0)}} \alpha_1\ dx_2 \wedge \cdots \wedge dx_k.
\end{aligned}
$$

As you notice, we switched notation from $dx_1 \wedge dx_2 \wedge \cdots \wedge dx_k$ in the second line to $dx_1\ dx_2\ \cdots\ dx_k$ in the third line, and then switched back to $dx_2 \wedge \cdots \wedge dx_k$ in the final line. As you hopefully remember from an earlier chapter, our traditional integration that you learned in calculus class differs from the integration of differential forms in only one way, it does not keep track of the orientation of the surface. We make this switch in order to use the fundamental theorem of calculus, which we only know in terms of our traditional notation. Nothing that we have done while we use the traditional notation affects the orientation of the cube over which we are integrating. We recognize this because the ordering when we switch from the forms notation to the traditional notation and then when we switch from our traditional notation back to the forms notation is consistent. To see this compare the second line where we have the volume form $dx_1 \wedge dx_2 \wedge \cdots \wedge dx_k$ with the last line where we have the volume form $dx_2 \wedge \cdots \wedge dx_k$. The volume form in the last line is one dimension lower than that in the first line, it is missing the dx_1 term because we integrated with respect to the dx_1 term, but other than that one omission the ordering of the other $k-1$ terms is unchanged. Because the ordering of the remaining terms is unchanged when we are ready to convert back to forms notation everything is consistent. For convenience sake we will rewrite this equality that we have just found denoting the first term of α as α_1 instead of $\alpha_1\ dx_2 \wedge \ldots \wedge dx_k$, so we have

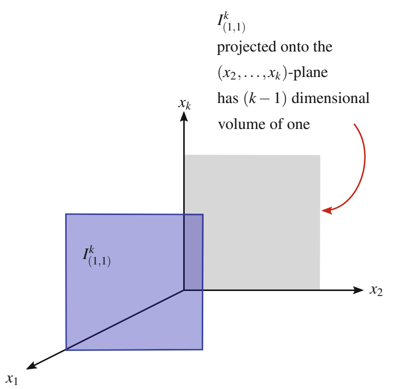
$$
\int_{I^k} d\alpha_1 = \int_{I^k_{(1,1)}} \alpha_1 - \int_{I^k_{(1,0)}} \alpha_1.
$$

This is admittedly a slight abuse of notation, but it does simplify things considerably.

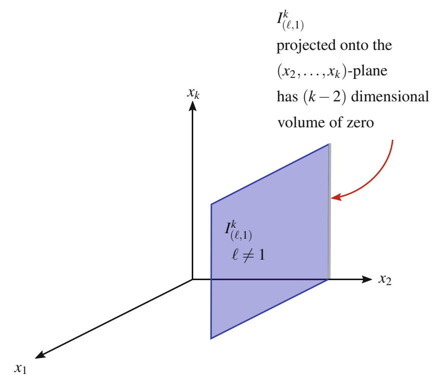
Now we take a moment to show that the integral of the $(k-1)$-form $\alpha_1 dx^2 \wedge \cdots \wedge dx_k$ is zero on all of the other faces of ∂I^k and that the only contribution we would expect from integrating α_1 over the boundary of I^k is that which is given by the right hand side of the above equation. All the other faces of ∂I^k are of the form $I^k_{(i,a)}$ for $2 \le i \le k$ and $a = 0, 1$.

Now recall what $dx_2 \wedge dx_3 \wedge \cdots \wedge dx_k$ does, it projects parallelepipeds to the $(x_2 \cdots x_k)$-plane and finds the volume of that projection. We are interested in taking integrals of α_1 on the various faces of I^k. Consider $I^k_{(1,1)}$, which is a $(k-1)$-dimensional parallelepiped. Projecting $I^k_{(1,1)}$ onto the $(x_2 \cdots x_k)$-plane and finding the volume of this projection gives a volume of one. Similarly, projecting $I^k_{(1,0)}$ onto the $(x_2 \cdots x_k)$-plane also gives us a volume of one. See Fig. 11.26. Thus we are able to take integrals of α_1 over over this face.

But now consider $I^k_{(2,1)}$. When this face is projected onto the $(x_2 \cdots x_k)$-plane the $(k-1)$-dimensional volume is zero, since the projection of $I^k_{(2,1)}$ onto the $(x_2 \cdots x_k)$-plane is a $(k-2)$-dimensional set, which has zero $(k-1)$-dimensional



The volume form $dx_2 \wedge dx_3 \wedge \cdots \wedge dx_k$ finds the volume of the projection of $I^k_{(1,1)}$ onto the $(x_2 \cdots x_k)$-plane. The case of $I^k_{(1,0)}$ is similar. Both have a $(k-1)$-dimensional volume of one.



The volume form $dx_2 \wedge dx_3 \wedge \cdots \wedge dx_k$ finds the volume of the projection of $I^k_{(2,1)}$ onto the $(x_2 \cdots x_k)$-plane. The $(k-1)$-dimensional volume is zero. Similarly, the $(k-1)$ dimensional volumes of the projections of $I^i_{(j,a)}$ for $j = 2,\ldots,k$ and $a = 0,1$ onto the $(x_2 \cdots x_k)$-plane are zero.

Fig. 11.26 Finding the volumes of the faces of I^k when they are projected to the $(x_2 \cdots x_k)$-plane

volume. This $\int_{I^k_{(2,1)}} \alpha_1 = 0$. Similarly we find that $\int_{I^k_{(j,a)}} \alpha_1 = 0$ for all $j = 2,\ldots,k$ and $a = 0,1$, see Fig. 11.26. Thus, the only integral on the boundary of I^k that can contribute anything are those we already have, $\int_{I^k_{(1,1)}} \alpha_1$ and $\int_{I^k_{(1,0)}} \alpha_1$.

Also, notice what we found, that

$$\int_{I^k} d\alpha_1 = \int_{I^k_{(1,1)}} \alpha_1 - \int_{I^k_{(1,0)}} \alpha_1.$$

The integral over $I^k_{(1,1)}$ is positive, which is the orientation of the face $I^k_{(1,1)}$ in the boundary of I^k, since $(-1)^{1+1} = 1$. Similarly, the integral over $I^k_{(1,0)}$ is negative, which is the orientation of the face $I^k_{(1,0)}$ in the boundary of I^k, since $(-1)^{1+0} = -1$. Thus we could, and in fact do, make the following definition

$$\int_{I^k_{(1,1)}} \alpha_1 - \int_{I^k_{(1,0)}} \alpha_1 \equiv \int_{I^k_{(1,1)} - I^k_{(1,0)}} \alpha_1,$$

which allows us to write

$$\int_{I^k} d\alpha_1 = \int_{I^k_{(1,1)} - I^k_{(1,0)}} \alpha_1.$$

Now we will look at the second term of α. We have

$$\begin{aligned}
&\int_{I^k} d\left(\alpha_2\, dx_1 \wedge dx_3 \wedge \cdots \wedge dx_k\right) \\
&= \int_{I^k} \frac{\partial \alpha_2}{\partial x_2}\, dx_2 \wedge dx_1 \wedge dx_3 \wedge \cdots dx_k \\
&= (-1) \int_{I^k} \frac{\partial \alpha_2}{\partial x_2}\, dx_1 \wedge dx_2 \wedge dx_3 \wedge \cdots dx_k \\
&= (-1) \int_{x_k=0}^{1} \cdots \int_{x_2=0}^{1} \int_{x_1=0}^{1} \frac{\partial \alpha_2}{\partial x_2}\, dx_1\, dx_2 \cdots dx_k \qquad \text{traditional integration (orientation not matter)} \\
&= (-1) \int_{x_k=0}^{1} \cdots \int_{x_1=0}^{1} \int_{x_2=0}^{1} \frac{\partial \alpha_2}{\partial x_2}\, dx_2\, dx_1 \cdots dx_k \qquad \text{Fubini's Theorem}
\end{aligned}$$

$$= (-1)\int_{x_k=0}^{1} \cdots \int_{x_1=0}^{1} \Big[\alpha_2(x_1, x_2, x_3 \ldots, x_k)\Big]_0^1 \, dx_1 \, dx_3 \cdots \, dx_k \quad \begin{matrix}\text{Fundamental}\\ \text{Theorem}\\ \text{of calculus}\end{matrix}$$

$$= (-1)\int_{x_k=0}^{1} \cdots \int_{x_3=0}^{1} \int_{x_1=0}^{1} \Big[\alpha_2(x_1, 1, x_3 \ldots, x_k) - \alpha_1(x_1, 0, x_3, \ldots, x_k)\Big] \, dx_1 \, dx_3 \, \cdots \, dx_k$$

$$= (-1)\underbrace{\int_{x_k=0}^{1} \cdots \int_{x_3=0}^{1} \int_{x_1=0}^{1} \alpha_2(x_1, 1, x_3, \ldots, x_k) \, dx_1 \, dx_3 \, \cdots \, dx_k}_{\int \text{ of } \alpha_2 \text{ restricted to face } I^k_{(2,1)}}$$

$$- (-1)\underbrace{\int_{x_k=0}^{1} \cdots \int_{x_3=0}^{1} \int_{x_1=0}^{1} \alpha_2(x_1, 0, x_3, \ldots, x_k) \, dx_1 \, dx_3 \, \cdots \, dx_k}_{\int \text{ of } \alpha_2 \text{ restricted to face } I^k_{(2,0)}}$$

$$= -\int_{I^k_{(2,1)}} \alpha_2 \, dx_1 \wedge dx_3 \wedge \cdots \wedge dx_k + \int_{I^k_{(2,0)}} \alpha_2 \, dx_1 \wedge dx_3 \wedge \cdots \wedge dx_k$$

$$= \int_{-I^k_{(2,1)} + I^k_{(2,0)}} \alpha_2 \, dx_1 \wedge dx_3 \wedge \cdots \wedge dx_k.$$

As you notice, we switched notation from $dx_1 \wedge dx_2 \wedge \cdots \wedge dx_k$ in the third line to $dx_1 \; dx_2 \; \cdots \; dx_k$ in the fourth line, and then switched back to $dx_2 \wedge \cdots \wedge dx_k$ in the next to last line. We make this switch in order to use both the Fubini's Theorem and the fundamental theorem of calculus, which we only know in terms of our traditional notation. You may think that by switching the order of integration when applying Fubini's Theorem is changing our volume-orientation, but since traditional integration does not keep track of volume-orientation sign this change does not change the accompanying volume-orientation sign. This means that what matters is that the orientations of our volume forms when we switch to and from our traditional notation are consistent. For example, the volume form in the third line is $dx_1 \wedge dx_2 \wedge dx_3 \wedge \cdots \wedge dx_k$ and the orientation in the next to last line is $dx_1 \wedge dx_3 \wedge \cdots \wedge dx_k$. Though we are missing dx_2 in the next to last line because we integrated with respect to that term, the ordering of the remaining terms is unchanged and thus consistent.

Notice how the signs again match the orientations of the faces in the boundary of I^k. This allows us to rewrite the end result of this calculation. If we additionally use the same abuse of notation that we used earlier, writing α_2 for $\alpha_2 \, dx_1 \wedge dx_3 \wedge \cdots \wedge dx_k$, we have the equality

$$\int_{I^k} d\alpha_2 = \int_{-I^k_{(2,1)} + I^k_{(2,0)}} \alpha_2.$$

Now that we have done this for the first two terms of the $(k-1)$-form you have some idea how we will proceed. Now we will do it for the full $\alpha = \sum \alpha_i dx_1 \wedge \cdots \wedge \widehat{dx_i} \wedge \cdots \wedge dx_k$, where $\widehat{dx_i}$ means that the dx_ith term is omitted. First we find $d\alpha$,

$$\begin{aligned}
& d\left(\sum_{i=1}^{k} \alpha_i dx_1 \wedge \cdots \wedge \widehat{dx_i} \wedge \cdots \wedge dx_k\right) \\
&= \sum_{i=1}^{k} d\left(\alpha_i dx_1 \wedge \cdots \wedge \widehat{dx_i} \wedge \cdots \wedge dx_k\right) \\
&= \sum_{i=1}^{k} \frac{\partial \alpha_i}{\partial x_i} \, dx_i \wedge dx_1 \wedge \cdots \wedge \widehat{dx_i} \wedge \cdots \wedge dx_k \\
&= \sum_{i=1}^{k} (-1)^{i-1} \, \frac{\partial \alpha_i}{\partial x_i} \, \underbrace{dx_1 \wedge \cdots \wedge dx_i \wedge \cdots \wedge dx_k}_{i-1 \text{ transpositions gives } (-1)^{i-1}} \\
&= \left(\sum_{i=1}^{k} (-1)^{i-1} \, \frac{\partial \alpha_i}{\partial x_i}\right) \, dx_1 \wedge \cdots \wedge dx_i \wedge \cdots \wedge dx_k
\end{aligned}$$

Now that we have $d\alpha$ we proceed with the integration,

$$
\begin{aligned}
\int_{I^k} d\alpha &= \int_{I^k} \left(\sum_{i=1}^{k} (-1)^{i-1} \frac{\partial \alpha_i}{\partial x_i} \right) dx_1 \wedge \cdots \wedge dx_k \\
&= \sum_{i=1}^{k} (-1)^{i-1} \int_{I^k} \frac{\partial \alpha_i}{\partial x_i} \, dx_1 \wedge \cdots \wedge dx_k \\
&= \sum_{i=1}^{k} (-1)^{i-1} \int_0^1 \cdots \int_0^1 \frac{\partial \alpha_i}{\partial x_i} \, dx_i dx_1 \cdots \widehat{dx_i} \cdots dx_k \qquad \begin{array}{c}\text{traditional integration}\\ \text{(orientation not matter)}\end{array} \\
&= \sum_{i=1}^{k} (-1)^{i-1} \int_0^1 \cdots \int_0^1 \left[\alpha_i(x_1, \ldots, x_k) \right]_0^1 \, dx_1 \cdots \widehat{dx_i} \cdots dx_k \qquad \begin{array}{c}\text{Fundamental}\\ \text{Theorem}\\ \text{of calculus}\end{array} \\
&= \sum_{i=1}^{k} (-1)^{i-1} \int_0^1 \cdots \int_0^1 \Big[\alpha_i(x_1, \ldots, \underbrace{1}_{i^{th}}, \ldots, x_k) - \alpha_i(x_1, \ldots, \underbrace{0}_{i^{th}}, \ldots, x_k) \Big] \, dx_1 \cdots \widehat{dx_i} \cdots dx_k \\
&= \sum_{i=1}^{k} (-1)^{i-1} \Bigg[\int_{I^k_{(i,1)}} \alpha_i \, dx_1 \wedge \cdots \wedge \widehat{dx_i} \wedge \cdots \wedge dx_k \\
&\qquad\qquad - \int_{I^k_{(i,0)}} \alpha_i \, dx_1 \wedge \cdots \wedge \widehat{dx_i} \wedge \cdots \wedge dx_k \Bigg] \\
&= \sum_{i=1}^{k} \Bigg[\underbrace{(-1)^{i-1} \int_{I^k_{(i,1)}} \alpha_i \, dx_1 \wedge \cdots \wedge \widehat{dx_i} \wedge \cdots \wedge dx_k}_{\substack{I^k_{(i,1)} \text{ has orientation } (-1)^{i+1} \text{ but } (-1)^{i-1} = (-1)^{i+1} \\ \text{so sign of integral matches orientation of } I^k_{(i,1)} \text{ in } \partial I^k}} \\
&\qquad\qquad + \underbrace{(-1)^{i} \int_{I^k_{(i,0)}} \alpha_i \, dx_1 \wedge \cdots \wedge \widehat{dx_i} \wedge \cdots \wedge dx_k}_{\substack{I^k_{(i,0)} \text{ has orientation } (-1)^{i+0} \text{ but } (-1)^{i+0} = (-1)^{i} \\ \text{so sign of integral matches orientation of } I^k_{(i,0)} \text{ in } \partial I^k}} \Bigg] \\
&= \sum_{i=1}^{k} \Bigg[\int_{(-1)^{i-1} I^k_{(i,1)}} \alpha_i \, dx_1 \wedge \cdots \wedge \widehat{dx_i} \wedge \cdots \wedge dx_k \\
&\qquad\qquad + \int_{(-1)^{i+0} I^k_{(i,0)}} \alpha_i \, dx_1 \wedge \cdots \wedge \widehat{dx_i} \wedge \cdots \wedge dx_k \Bigg] \\
&= \sum_{i=1}^{k} \int_{(-1)^{i-1} I^k_{(i,1)} + (-1)^{i+0} I^k_{(i,0)}} \alpha_i \, dx_1 \wedge \cdots \wedge \widehat{dx_i} \wedge \cdots \wedge dx_k \\
&= \int_{\sum_{i=1}^{k} \left((-1)^{i-1} I^k_{(i,1)} + (-1)^{i+0} I^k_{(i,0)} \right)} \sum_{i=1}^{k} \alpha_i \, dx_1 \wedge \cdots \wedge \widehat{dx_i} \wedge \cdots \wedge dx_k \qquad \begin{array}{c}\text{Only } \alpha_i \text{ term of } \alpha\\ \text{contributes to faces}\\ I^i_{(i,1)} \text{ and } I^i_{(i,0)}\\ \text{so we can add this.}\end{array} \\
&= \int_{\partial I^k} \alpha.
\end{aligned}
$$

Thus we have shown for a $(k-1)$-form α defined in a neighborhood of the unit k-cube that

$$\boxed{\int_{I^k} d\alpha = \int_{\partial I^k} \alpha.}$$

This is exactly Stokes' theorem for the case of the unit k-cube. Now that we have this all-important base case the other cases follow fairly quickly and easily.

11.3 Manifolds Parameterized by I^k

Now suppose we have a manifold with boundary M that is parameterized by the unit cube. That is, we have a mapping $\phi : \mathbb{R}^k \to \mathbb{R}^n$ where $\phi(I^k) = M$. See Fig. 11.27. Actually, ϕ need not even be one-to-one. The mapping ϕ is often called a **singular k-cube** on M. The word singular indicates that the mapping ϕ need not be one-to-one. Requiring ϕ to be non-singular is more restrictive and not necessary, though we will require that the ϕ be differentiable.

Whenever looking at the proofs of Stokes' theorem in other books recall that $T^*\phi$ is our somewhat nonstandard notation for ϕ^*. We will not be terribly rigorous at this point, but we we will also require that ϕ respect boundaries, and by that we mean if $M = \phi(I^k)$ then $\partial M = \phi(\partial I^k)$. In other words, we require that

$$\partial\phi(I^k) = \phi(\partial I^k).$$

This is sufficiently general and "obvious" that is does not cause a problem. We also need to recall the following identity regarding the integration of pull-backed forms,

$$\int_{\phi(R)} \alpha = \int_R T^*\phi \cdot \alpha$$

for some region R. In the previous chapter we would have written this with $\phi^{-1}(R)$, but if ϕ were singular, that is, not one-to-one, then ϕ would not be invertible so we avoid that notation here. Finally, recall that the exterior derivative commutes with pull-back, that is, that

$$d \circ T^*\phi = T^*\phi \circ d$$

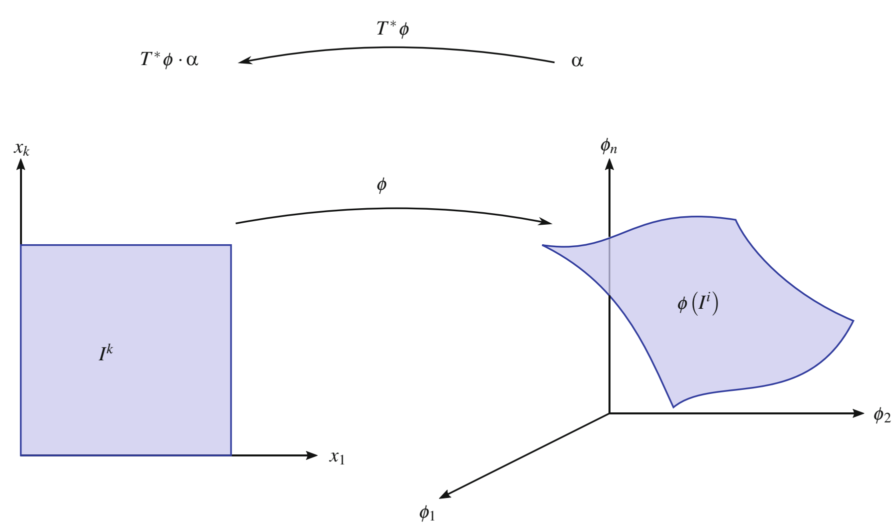


Fig. 11.27 The mapping $\phi : \mathbb{R}^k \to \mathbb{R}^n$ where $\phi(I^k) = M$. A differential form α defined on $\phi(I^k)$ can be pulled back by $T^*\phi$ to the differential form $T^*\phi \cdot \alpha$ defined on I^k

or $d \circ \phi^* = \phi^* \circ d$ in most differential geometry textbooks. Using all of this, the final step is almost anti-climactic,

$$\begin{aligned}
\int_M d\alpha &= \int_{\phi(I^i)} d\alpha && \text{notation} \\
&= \int_{I^k} T^*\phi \cdot d\alpha && \text{integration of pulled-back forms} \\
&= \int_{I^k} d(T^*\phi \cdot \alpha) && \text{exterior derivative commutes with pull-back} \\
&= \int_{\partial I^k} T^*\phi \cdot \alpha && \text{base case} \\
&= \int_{\phi(\partial I^k)} \alpha. && \text{integration of pulled-back forms} \\
&= \int_{\partial\phi(I^k)} \alpha && \phi \text{ respects boundaries} \\
&= \int_{\partial M} \alpha && \text{notation}
\end{aligned}$$

Thus, for a singular k-cube on M we have

$$\boxed{\int_M d\alpha = \int_{\partial M} \alpha.}$$

11.4 Stokes' Theorem on Chains

Now suppose that our manifold with boundary, M, is not nicely parameterized by a single map ϕ, but can be broken into cuboid regions, each of which can be parameterized by a map $\phi_i : I^k \to \phi_i(I^k)$, $i = 1, \ldots, r$. We will assume a finite number of maps ϕ_i are sufficient. See Fig. 11.28. But a warning, here we are using the notation ϕ_i differently than we did in past chapters. Here each ϕ_i is a mapping $\phi_i : \mathbb{R}^k \to \mathbb{R}^n$ whereas before we had that a mapping $\phi : \mathbb{R}^k \to \mathbb{R}^n$ that was defined as $\phi = (\phi_1, \phi_2, \ldots, \phi_k)$ where each ϕ_i was a real-valued function $\phi_i : \mathbb{R}^k \to \mathbb{R}$.

We choose the maps that are used to parameterize the manifold M such that the interiors of $\phi_i(I^k)$ are non-overlapping and the boundaries of $\phi_i(I^k)$ match up. We won't go into excessive technical details trusting that the above picture will give you the right intuitive idea. A **singular k-chain** C on M is defined to be

$$C = \phi_1 + \phi_2 + \cdots + \phi_r$$

where each ϕ_i, $i = 1, \ldots, r$, is a singular k-cube on M. In other words, we have

$$\begin{aligned}
M &= \bigcup_{i=1}^{r} \phi_i(I^k) \\
&\equiv \phi_1(I^k) + \phi_2(I^k) + \cdots + \phi_r(I^k) \\
&\equiv C(I^k).
\end{aligned}$$

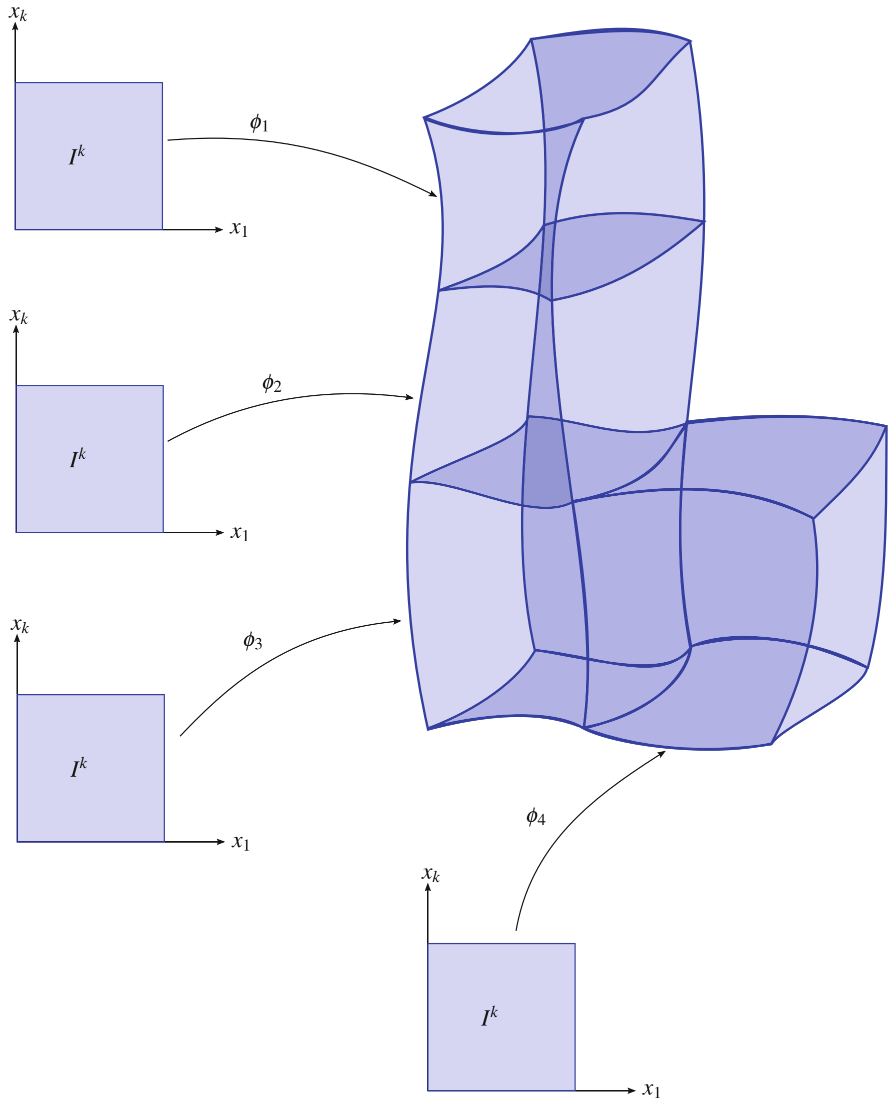


Fig. 11.28 The manifold M can be broken up into a finite number of cuboid regions each of which is parameterized by a map $\phi_i : I^k \to \phi(I^k) \subset M$. If the interiors of the $\phi_i(I^k)$ are non-overlapping, the boundaries of $\phi_i(I^k)$ match up, and $M = \bigcup_{i=1}^r \phi_i(I^k)$ then $\phi_1 + \phi_2 + \cdots + \phi_n$ is called a singular k-chain on M

We will also assume that all the interior boundaries of the singular k-chain C match up and have the opposite orientations as Fig. 11.29 shows. The orientation of $\phi_i(I^k_{(l,1)})$ is $(-1)^{l+1}$ and the orientation of $\phi_j(I^k_{(l,0)})$ is $(-1)^{l+0}$ so the contributions to the integral of α on $\phi_i(I^k_{(l,1)})$ is canceled by the contribution of the integral of α on $\phi_j(I^k_{(l,0)})$.

We have

$$
\begin{aligned}
\partial(M) &= \partial\big(C(I^k)\big) \\
&= \partial\big(\phi_1(I^k) + \phi_2(I^k) + \cdots + \phi_r(I^k)\big) \\
&= \partial\big(\phi_1(I^k)) + \partial(\phi_2(I^k)) + \cdots + \partial(\phi_r(I^k)\big) \\
&= \phi_1\big(\partial(I^k)\big) + \phi_2\big(\partial(I^k)\big) + \cdots + \phi_r\big(\partial(I^k)\big)
\end{aligned}
$$

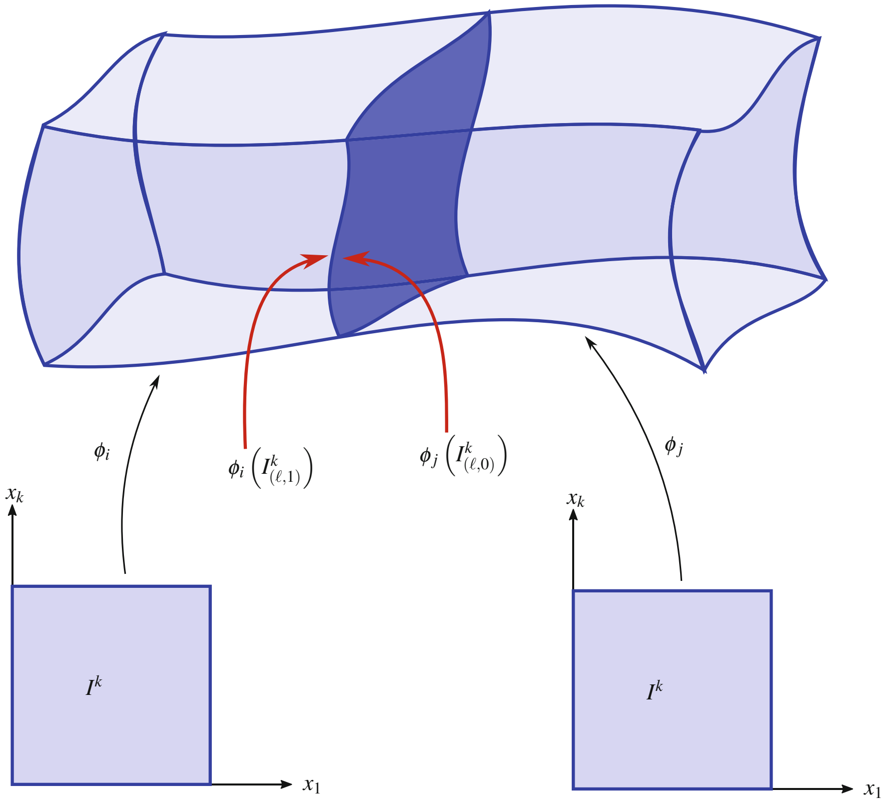


Fig. 11.29 Here the boundaries of two cuboid regions match up and have opposite orientations. That is, $\phi_i(I^k_{(l,1)}) = -\phi_j(I^k_{(l,0)})$

So proving Stokes' theorem on chains is easy,

$$\begin{aligned}
\int_M d\alpha &= \int_{C(I^k)} d\alpha \\
&= \int_{\phi_1(I^k)+\cdots+\phi_r(I^k)} d\alpha \\
&= \int_{\phi_1(I^k)} d\alpha + \cdots + \int_{\phi_r(I^k)} d\alpha \\
&= \int_{I^k} T^*\phi_1 \cdot d\alpha + \cdots + \int_{I^k} T^*\phi_r \cdot d\alpha \\
&= \int_{I^k} d(T^*\phi_1 \cdot \alpha) + \cdots + \int_{I^k} d(T^*\phi_r \cdot \alpha) \\
&= \int_{\partial I^k} T^*\phi_1 \cdot \alpha + \cdots + \int_{\partial I^k} T^*\phi_r \cdot \alpha \\
&= \int_{\phi_1(\partial I^k)} \alpha + \cdots + \int_{\phi_r(\partial I^k)} \alpha \\
&= \int_{\phi_1(\partial I^k)+\cdots+\phi_r(\partial I^k)} \alpha \\
&= \int_{\partial\phi_1(I^k)+\cdots+\partial\phi_r(I^k)} \alpha \\
&= \int_{\partial(M)} \alpha.
\end{aligned}$$

So we have

$$\boxed{\int_M d\alpha = \int_{\partial M} \alpha}$$

for singular k-chains on M. Sometimes Stokes' theorem is stated in terms of chains.

Theorem 11.2 (Stokes' Theorem (Chain Version)) *For any n-form α and $(n+1)$-chain c then*

$$\int_c d\alpha = \int_{\partial c} \alpha.$$

Question 11.13 Explain the rationale in each step of the above chain of equalities.

11.5 Extending the Parameterizations

Finally, while it may look like the cubical domain is special, it really isn't. We will only sketch this last step in the proof of Stokes' theorem, but it should be enough to convince you that it is true. We will leave it to you, in the exercises, to fill in more of the details. The most important thing is to realize that virtually any domain can be either reparameterized into unit cubes or broken up into pieces which can then be reparameterized into unit cubes. Reverting to two-dimensions consider the domain

$$D = \left\{ (u, v) \in \mathbb{R}^2 \;\middle|\; a \le u \le b,\ f(u) \le v \le g(u) \right\}$$

where we know $\phi : D \to \mathbb{R}^n$ is a nice parametrization of the manifold $M = \phi D$. See Fig. 11.30. Our intent is to show that having a parameterizations of our manifold M by domains like D is enough for Stokes' theorem. Consider the mapping

$$(s, t) \in I^2 \overset{\psi}{\longmapsto} \Big((1-s)a + sb,\ (1-t) f\big((1-s)a + sb\big) + t g\big((1-s)a + sb\big) \Big).$$

Question 11.14 Show that $\psi : I^2 \to D$ is one-to-one and onto.

Question 11.15 Show that $\phi \circ \psi : I^2 \to M$ is a surface parameterized by I^2 and hence the proof of Stokes' theorem applies to M.

A similar construction can be made for dimensions greater than two. Thus we have argued that Stokes' theorem

$$\boxed{\int_M d\alpha = \int_{\partial M} \alpha}$$

applies to any manifold whose parameterizations can reparameterized into unit cubes or broken up into pieces which can then be reparameterized into unit cubes. Without getting into the technical details this applies to essentially all manifolds.

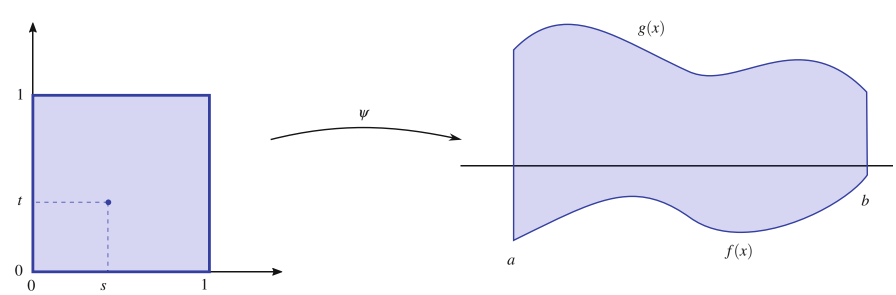


Fig. 11.30 Almost any domain can be reparameterized into a unit cube or broken into pieces that can be reparameterized into unit cubes

11.6 Visualizing Stokes' Theorem

In this section we will try to provide a cartoon image of what the generalized Stokes' theorem means, at least in three dimensions, based on the material in Chap. 5. This presentation is more in line with the way that physicists sometimes visualize differential forms than with the perspective that mathematicians generally take. This visualization technique is actually not general, there are manifolds and forms for which it breaks down and does not work, however, when appropriate it does provide a very nice geometrically intuitive way of looking at Stoke's theorem. In particular we will rely on the visualizations developed in Sect. 5.5. With the aid of Stokes' theorem we will be able to gain a slightly more nuanced view of these visualizations, at least in the cases where these visualizations are possible. We will tie together all of these ideas in this section.

First we recall that in $\mathbb{R}^3$ we can visualize one-forms as sheets filling $\mathbb{R}^3$, as shown in Fig. 5.36, two-forms as tubes filling $\mathbb{R}^3$, as shown in Fig. 5.40, and three-forms as cubes filling $\mathbb{R}^3$, as shown in Fig. 5.30. We will simply simply continue to view zero-forms f on $\mathbb{R}^3$ as we have always viewed functions, as a value $f(p)$ attached to each point $p \in \mathbb{R}^3$. We also will recall a differential form ω is called closed if $d\omega = 0$.

We will begin by considering a zero-form f. Clearly df is a one-form which we can imagine as sheets filling space. A one-dimensional manifold is a curve C, so in the case of the zero-form f Stokes' theorem gives us

$$\int_C df = \int_{\partial C} f.$$

From Sect. 9.5.2 we of course recognize this as simply the fundamental theorem of line integrals. As we move along C the integral of df counts the number of sheets of the one-form df the curve C goes through. See Fig. 11.31.

Question 11.16 Based on what you know about how one-forms can be written as sheets and how line integrals are calculated, argue that indeed $\int_C df$ counts the number of sheets of df that the curve C passes through.

Question 11.17 Suppose f is closed, that is, $df = 0$. What does this say about the function f? Use Fig. 11.31 as a guide.

Question 11.18 Based on the last two questions, how can we use f to help us draw the sheets of the one-form df?

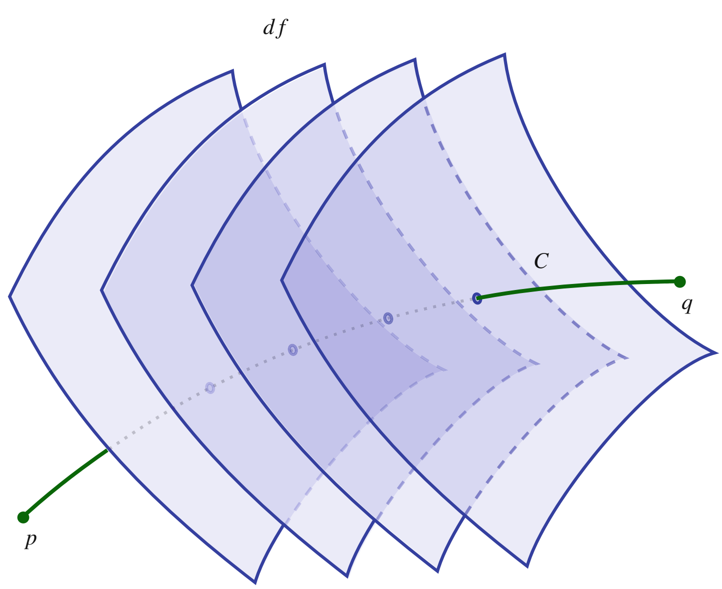


Fig. 11.31 A curve C passing through the sheets of df. In this figure $\int_C df = 4$. Thus, by Stokes' theorem we know $\int_{\partial C} f = f(q) - f(p) = 4$. We can think of the number of sheets as being given by the change in the value of the function between points p and q

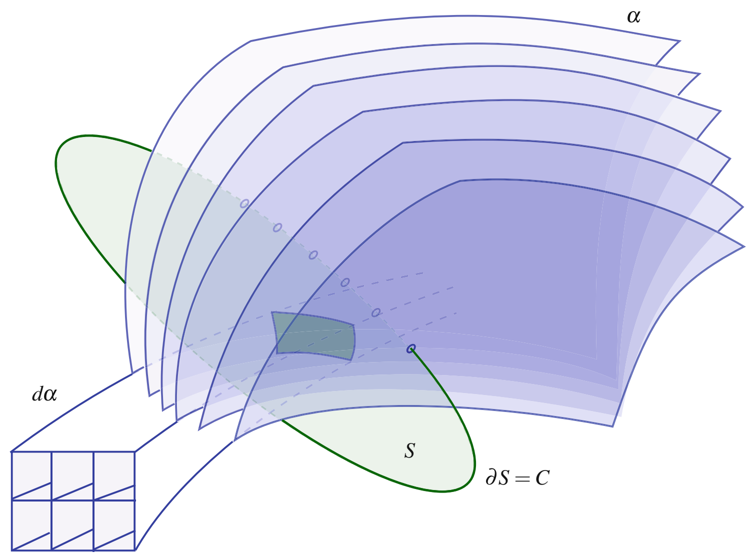


Fig. 11.32 A surface S with boundary $\partial S = C$. In the picture the one-form α is depicted as sheets and the two-form $d\alpha$ is depicted as tubes. Here $\int_{\partial S} \alpha$ counts the number of sheets that the curve C goes through and $\int_S d\alpha$ counts the number of tubes that go through the surface S. In this picture we have both $\int_S d\alpha = 6$ and $\int_{\partial S} \alpha = 6$. We can think of the sheets of α as emanating from the tubes of $d\alpha$, one sheet for each tube

Next we consider a one-form α. Since α is a one-form then we can visualize it as $\mathbb{R}^3$ filled with sheets whereas its exterior derivative $d\alpha$ is a two-form, which can be visualized as $\mathbb{R}^3$ filled with tubes. How exactly do these two pictures relate? A two-dimensional manifold is a surface S with boundary being a curve, $\partial S = C$. We will turn to Stokes' theorem

$$\int_S d\alpha = \int_{\partial S} \alpha.$$

From Sect. 9.5.3 we know this is simply the vector calculus version of Stokes' theorem. $\int_S d\alpha$ counts the number of tubes of $d\alpha$ that go through the surface S and, as before, $\int_{\partial S} \alpha$ counts the number of surfaces of α. From Stokes' theorem these are the same. See Fig. 11.32. Based on Stokes' theorem the number of tubes of $d\alpha$ going through S is the same as the number of sheets of α going through ∂S, so in a sense each tube of $d\alpha$ gives rise to additional sheets of α.

Question 11.19 Based on what you know about how two-forms are depicted as tubes and how surface integrals are calculated, argue that indeed $\int_S d\alpha$ counts the number of tubes of $d\alpha$ that pass through S.

Question 11.20 Suppose α is closed. What does that say about the number of tubes passing through S or the number of sheets passing through $\partial S = C$? Is it possible for α to still have sheets even though $d\alpha = 0$? Sketch a picture of this situation.

If α is a two-form then it can be visualized as $\mathbb{R}^3$ filled with tubes whereas the exterior derivative $d\alpha$ is a three-form, which can be visualized as $\mathbb{R}^3$ filled with small cubes. A three-dimensional manifold is a volume V with boundary being a surface, $\partial V = S$. Stokes' theorem becomes

$$\int_V d\alpha = \int_{\partial V} \alpha.$$

From Sect. 9.5.4 we know this is simply the divergence theorem from vector calculus. The integral $\int_V d\alpha$ counts the number of cubes of $d\alpha$ inside the volume V while $\int_{\partial V} \alpha$ counts the number of tubes of α that penetrate the surface S. This leads to a picture very similar to the last two, see Fig. 11.33. Since these are equal by Stokes' theorem then we can imagine that the tubes of α are emanating from the cubes of $d\alpha$, one tube for each cube.

Question 11.21 Based on what you know about how three-forms are depicted as cubes and how volume integrals are calculated, argue that $\int_V d\alpha$ indeed counts the number of cubes of $d\alpha$ inside V.

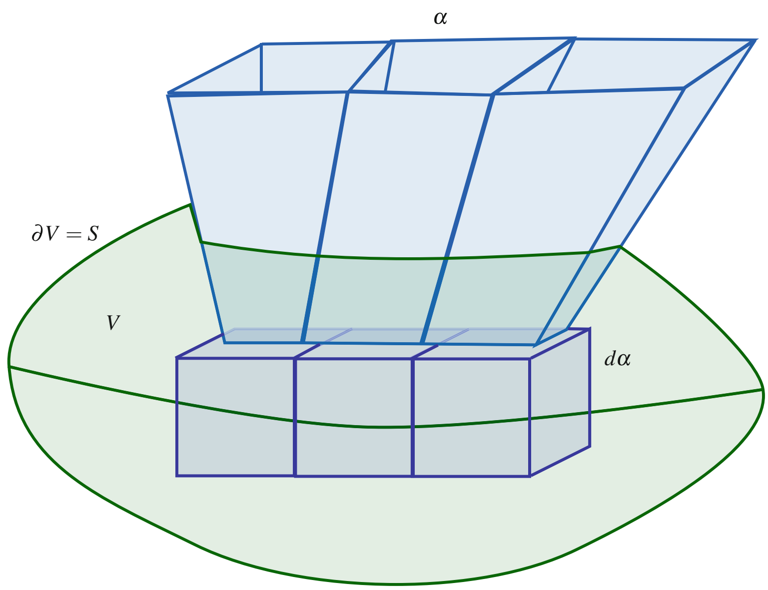


Fig. 11.33 A volume V with boundary $\partial V = S$. In the picture the two-form α is depicted as tubes and the three-form $d\alpha$ is depicted as cubes. Here $\int_{\partial V} \alpha$ counts the number of tubes that go through the surface S and $\int_V d\alpha$ counts the number of cubes that are contained in the volume V. In this picture we have both $\int_V d\alpha = 3$ and $\int_{\partial V} \alpha = 3$. We can think of the tubes of α as emanating from the cubes of $d\alpha$, one tube for each cube

Question 11.22 Suppose α is closed. What does this say about the number of cubes inside V or the number of tubes that pass though $\partial V = S$? Is it possible to still have tubes even though $d\alpha = 0$? Sketch a picture of this situation.

As you recall, in Chap. 4 we took great pains to introduce exterior differentiation from a variety of viewpoints. We did this because exterior differentiation plays a fundamental role in this book and in differential geometry and physics in general. In Sect. 4.5 we provided what we believe is one of the nicest geometrical meanings of exterior differentiation. However, that perspective relied on understanding integration of forms, which at that point we had not yet covered. We are now ready to return to the geometric picture of exterior differentiation presented in that section to understand the geometry of exterior differentiation, at least in three dimensions, a little more fully. In that section we had looked at the two-form ω and had obtained the following formula for $d\omega(v_1, v_2, v_3)$,

$$
\begin{aligned}
d\omega(v_1, v_2, v_3) &= \lim_{h\to 0} \frac{1}{h^3} \int_{\partial(hP)} \omega \\
&= \lim_{h\to 0} \frac{1}{h^3} \left(\int_{hP_{(1,1)}} \omega - \int_{hP_{(1,0)}} \omega - \int_{hP_{(2,1)}} \omega + \int_{hP_{(2,0)}} \omega + \int_{hP_{(3,1)}} \omega - \int_{hP_{(3,0)}} \omega \right) \\
&= \underbrace{\lim_{h\to 0} \frac{1}{h^3} \left(\int_{hP_{(1,1)}} \omega - \int_{hP_{(1,0)}} \omega \right)}_{\text{see Fig. 11.34 left}} - \underbrace{\lim_{h\to 0} \frac{1}{h^3} \left(\int_{hP_{(2,1)}} \omega - \int_{hP_{(2,0)}} \omega \right)}_{\text{see Fig. 11.34 middle}} \\
&\quad + \underbrace{\lim_{h\to 0} \frac{1}{h^3} \left(\int_{hP_{(3,1)}} \omega - \int_{hP_{(3,0)}} \omega \right)}_{\text{see Fig. 11.34 right}} .
\end{aligned}
$$

But now we have a much better idea of what this represents graphically. If ω is a two-form then we think of it as tubes filling space. The integral $\int_{hP_{(1,1)}} \omega$ is the number of tubes going through the $hP_{(1,1)}$ face of the parallelepiped hP. Similarly, the integral $\int_{hP_{(1,0)}} \omega$ is the number of tubes going through the $hP_{(1,0)}$ face of hP. The difference of these integrals is simply the difference in the number of tubes between these two faces in the v_1 direction. Of course we have to pay attention to the orientation of the tubes to get the sign of the integral correct, but this comes out when we actually do the calculation.

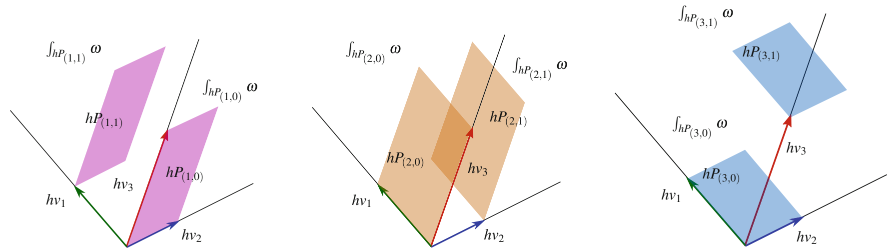


Fig. 11.34 The six faces of the cube $\partial(hP)$ which appear in the integral $\int_{\partial(hP)} \omega$. The integral $\int_{hP_{(1,1)}} \omega$ is the number of tubes going through the surface $hP_{(1,1)}$. The integral $\int_{hP_{(1,0)}} \omega$ is the number of tubes going through the surface $hP_{(1,0)}$, and so on. These six terms are added or subtracted consistent with the orientations of the face to get $\int_{\partial(hP)} \omega$

The other directions are handled similarly. When we sum the various terms we obtain the net change in the number of tubes occurring at a point in the directions v_1, v_2, and v_3 compatible with the orientations of the boundary of the parallelepiped $P = \text{span}\{v_1, v_2, v_3\}$. See Fig. 11.34.

Question 11.23 Repeat this analysis for a one-form α to explain the geometry of the exterior derivative $d\alpha$.

11.7 Summary, References, and Problems

11.7.1 Summary

Stokes' theorem is the major result regarding the integration of forms on manifolds. It states that if M is a smooth oriented n-dimensional manifold and α is an $(n-1)$-form on M then

$$\int_M d\alpha = \int_{\partial M} \alpha,$$

where ∂M is given the induced orientation.

11.7.2 References and Further Reading

There are probably no books on manifold theory or differential geometry that don't cover Stokes' theorem, but the presentation here largely followed Hubbard and Hubbard [27], Edwards [17], and Spivak [40]. While this presentation tried to be quite thorough and cover most cases, we have chosen to ignore several layers of technical details that make the theorem completely general. In particular, we have ignored the technical details related to manifolds with boundary. Also, as mentioned, it is possible to prove the whole thing on chains of k-simplices instead of k-cubes, the approach taken in Flanders [19]. In visualizing Stokes' theorem, at least in three dimensions, we have relied on the material of Chap. 5 as well as the paper by Warnick, Selfridge, and Arnold [49].

11.7.3 Problems

Question 11.24 Use Stokes' Theorem to find $\int_C \alpha$ where $\alpha = (2xy^3 + 4x)\,dx + (3x^2y^2 - 9y^2)\,dy$ is a one-form on $\mathbb{R}^2$ and C is the line segment from $(3, 2)$ to $(5, 4)$. Then find $\int_C \alpha$ for the line segment from $(5, 4)$ to $(3, 2)$.

Question 11.25 Use Stokes' Theorem to find $\int_C \alpha$ where $\alpha = (6x^2 - 3y)\,dx + (8y - 3x)\,dy$ is a one-form on $\mathbb{R}^2$ and C is the line segment from $(3, 3)$ to $(2, 2)$. Then find $\int_C \alpha$ for the line segment from $(2, 2)$ to $(3, 3)$.

Question 11.26 Let S be the region on $\mathbb{R}^2$ bounded by $y = 3x$ and $y = x^2$. Choose C_1 and C_2 such that $\partial S = C_1 + C_2$. Let $\alpha = x^2 y\, dx + y\, dy$ be a one-form on $\mathbb{R}^2$. Find $\int_S d\alpha$ directly. Then find $\int_{\partial S} \alpha$. Verify Stokes' theorem is satisfied.

Question 11.27 Let S be the quarter of the unit circle in the first quadrant of $\mathbb{R}^2$. Let C_1 be given by $\gamma_1(\theta) = \big(\cos(\theta), \sin(\theta)\big)$ for $0 \leq \theta \leq \frac{\pi}{2}$, let C_2 be the line segment from $(0, 1)$ to $(0, 0)$, and let C_3 be the line segment from $(0, 0)$ to $(1, 0)$. Hence $\partial S = C_1 + C_2 + C_3$. Let $\alpha = xy\, dx + x^2\, dy$. Use Stokes's theorem to find $\int_{C_1} \alpha$.

Question 11.28 Verify Stokes' theorem for M the unit disk $x^2 + y^2 \leq 1$ and $\alpha = xy\, dy$.

Question 11.29 Verify Stokes' theorem for $M = \big\{(u + v, u - v^2, uv)\big| 0 \leq u \leq 1,\ 0 \leq v \leq 2\big\}$ and $\alpha = 3z^2\, dx$.

Question 11.30 Verify Stokes' theorem for $M = \big\{(u^2 \cos(v), u^3 \sin(v), u^2)\big| 0 \leq u \leq 2,\ 0 \leq v \leq 2\pi\big\}$ and $\alpha = -yz\, dx + xz\, dy$.

Question 11.31 Let V be a volume in $\mathbb{R}^3$. Show that the three-volume of V is given by
$\int_{\partial V} \frac{1}{3} (z\, dx \wedge dy + y\, dz \wedge dx + x\, dy \wedge dz)$.

Question 11.32 Verify Stokes' theorem for M the unit cube $0 \leq x \leq 1, 0 \leq y \leq 1, 0 \leq z \leq 1$ and $\alpha = 2xz\, dx \wedge dy + x\, dy \wedge dz - y^2\, dz \wedge dz$.

Chapter 12
An Example: Electromagnetism

Electromagnetism is probably the first truly elegant and exciting application of differential forms in physics. Electromagnetism deals with both electrical fields and magnetic fields and Maxwell's equations are the four equations that describe how these fields act and how they are related to each other. Maxwell's complicated equations are rendered stunningly simple and beautiful when written in differential forms notation instead of the usual vector calculus notation.

We will begin the Electromagnetism chapter by first introducing the basic electromagnetic concepts and Maxwell's four equations in vector calculus notation. These four equations are generally known as

1. Gauss's law for electric fields,
2. Gauss's law for magnetic fields,
3. Faraday's law,
4. Ampère-Maxwell law (sometimes simply Ampère's law).

Since we do not assume you have already seen electromagnetism before we take our time. Section one introduces the first two of Maxwell's four equations and section two introduces the next two equations. After this introduction, in section three we will discuss Minkowski space, which is the four-dimensional space-time manifold of special relativity. The Minkowski metric, which gives the inner product on Minkowski space, is necessary for the Hodge star operator in the context of Minkowski space. We have already discussed both the inner product and the Hodge star operator in different contexts so other than the fact that we will be in four-dimensional Minkowski space there is fundamentally nothing new here. After that we derive the differential forms formulation of electromagnetism in section four.

This is certainly not meant to be an exhaustive introduction to electromagnetism. In fact, it will be a rather brief look at electromagnetism covering little more than the basics. It is meant to give you some idea of the power and utility of differential forms from the perspective of physics. Finally, please note, that in a typical mathematician's fashion we will not spend any time discussing units, which are, understandably, of vital importance to physicists and engineers. Hopefully you will not bare us too much ill-will for this omission.

12.1 Gauss's Laws for Electric and Magnetic Fields

Electric charge is an intrinsic conserved physical characteristic or property of subatomic particles. By conserved we are saying that this property remains fixed and does not change over time. On a very deep and fundamental level it may be impossible to truly understand what physical properties such as electric charge are, but on a more mundane operational and experiential level we can see, experience, test, categorize, measure, and write equations that describe these physical properties and how they interact with each other. After all, we all know an electrical shock when we get one.

There are two different kinds of electric fields. The first kind is the electrostatic field that is produced by an electric charge and the second kind of electric field is an induced electric field that is created or produced by a changing magnetic field. We will discuss electrostatic fields first.

There are two kinds of electrical charges, positive and negative, which are of course simply two completely arbitrary designations. Subatomic particles have a negative electric charge, a positive electric charge, or no electric charge. Electrons have a negative electric charge, protons have a positive electric charge, and neutrons are natural, meaning they have no electrical charge. When we discuss the charge of an object larger than a subatomic particle we are implying that there is a

J. P. Fortney, *A Visual Introduction to Differential Forms and Calculus on Manifolds*,
https://doi.org/10.1007/978-3-319-96992-3_12

net imbalance between the number of electrons and protons in the object. A negative charge means there are more negatively charged electrons than positively charged protons, and a positive charge means there are more positively charged protons than negatively charged electrons.

Particles that carry electrical charge interact with each other due to their charges and we can write equations that describe how they interact with each other. They can also interact with each other due to other reasons as well, such as gravitation or other forces, but we will not concern ourselves with these other interactions here. We will only concern ourselves with interactions that are due to electric charge. If two particles have the same electric charge, either positive or negative, they are repulsed by each other. If two particles have oppositive electric charges then they are attracted to each other.

Electrostatic fields are induced by, or created by, electric charges. We will denote electrical field by $\mathbf{E}$. Right now we will assume space is equivalent to $\mathbb{R}^3$, though of course this is just a local approximation to general relativity's four-dimensional space-time manifold. In essence, electrical fields are basically just a vector field on space, that is

$$\mathbf{E}_p = \begin{bmatrix} E_1 \\ E_2 \\ E_3 \end{bmatrix}_p$$

at each point $p \in \mathbb{R}^3$. Of course, we can't actually see this electric field, so how do we know it is there? Imagine you had a tiny positive "test particle." If you were to place this test particle at point p and then let go of it, the speed and direction that this tiny imaginary test particle moves when you let go of it determines the vector $\mathbf{E}_p$. Now imagine doing that at every point of space. At each point the movement of this tiny imaginary test particle gives the vector $\mathbf{E}_p$. Consider Fig. 12.1. If we were to place a tiny imaginary test particle somewhere close to a positive charge and let go, this test particle would move away from the positive charge. If we let go of the test particle very close to the positive charge it would move away very quickly. If we let go of the test particle further away from the positive charge it would move away more slowly. Similarly, if we were to place the test particle some distance away from the negative charge and let go it would move toward the negative charge slowly and if we were to put it very close to the negative charge and let go it would move toward the negative charge quickly.

Suppose we had a positive charge and a negative charge separated by a small distance as in Fig. 12.2. Wherever we released the test particle it would feel both a repulsion from the positive charge and an attraction to the negative charge. How strong the repulsion and attraction are relative to each other would depend on how far the test particle was from each charge. For example, a test particle released at points (a) or (c) would feel some attraction to the negative charge and some repulsion from the positive charge. A test particle released at (b) feels much more attraction to the negative charge than repulsion from the positive charge. And a test particle released at (d) feels much more repulsion from the positive charge than attraction to the negative charge.

Using the electric field we can draw electric field lines. The electric field lines are actually the integral curves of the electric field. Integral curves are the paths we get when we integrate vector fields, see for example Fig. 2.19 where the integral curves of a smooth vector field are shown. In fact, finding the integral curves of vector fields is one approach to integration on manifolds which we have not yet discussed. We will discuss it in a little more depth in Appendix B. At each point the electric field vector is given by the tangent vector to the electric field line. If a tiny positive test particle were released at a point it would travel along the electric field line with a velocity at each point given by the electric field vector at that point. The

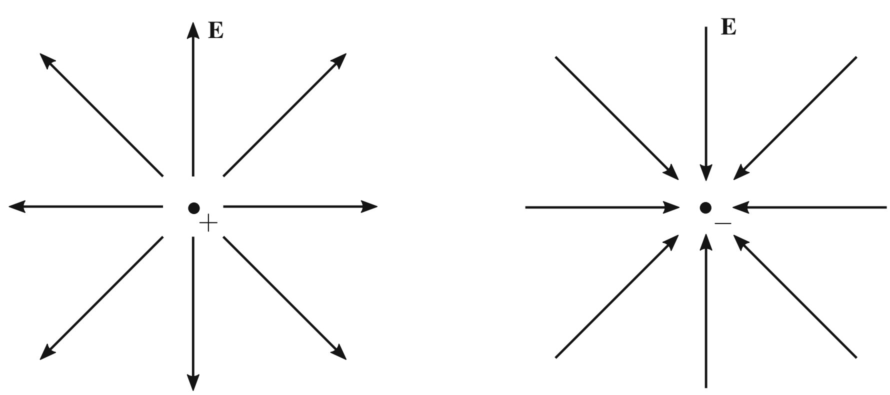


Fig. 12.1 A positively charged particle (left) and a negatively charged particle (right). Electric field lines emanating from the particles have been shown. Electric field lines are shown pointing away from positive charges and toward negative charges

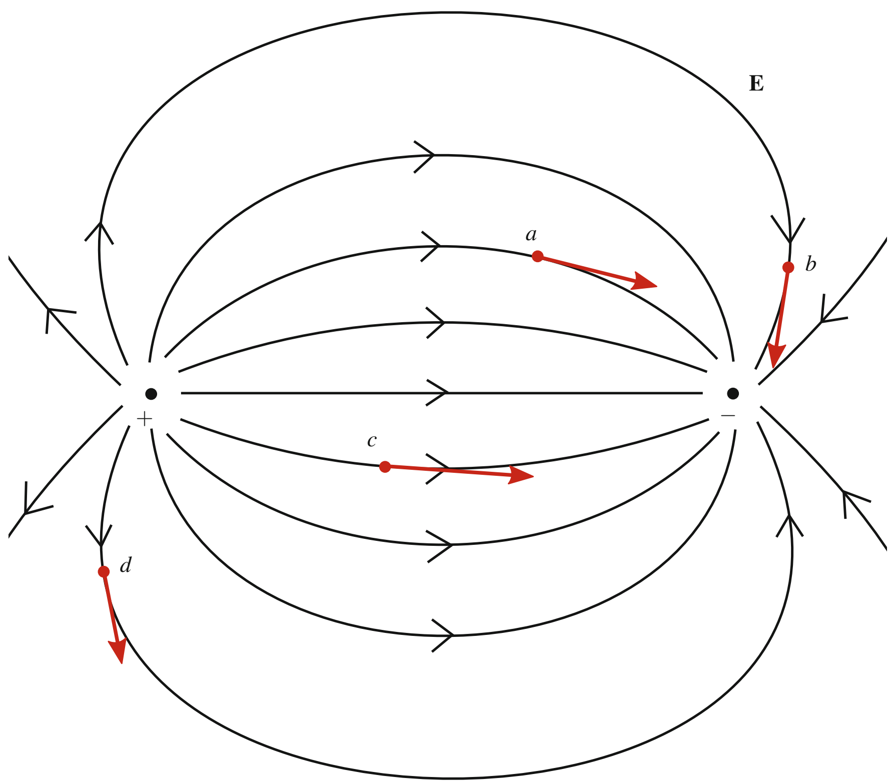


Fig. 12.2 A positively charged particle and a negatively charged particle that are close to each other. We can see some of the electric field lines that leave the positive charge go to the negative charge

length of the vector at a point tells how strong the field is at that point. Pictorially we try to show how strong the field is in a region of space by how close together we draw the electric field lines. So, in Fig. 12.2 we can see that the closer we are to the charges the denser the field lines are and so the stronger the field is. The further away we are from the charges the further apart the field lines are and the weaker the field is. Now we are ready to state **Gauss's law for electric fields**. This law concerns electrostatic fields.

An electric field is produced by an electric charge. The flux of this electric field through a closed surface is proportional to the amount of electric charge inside the closed surface.

This law can be expressed in vector calculus notation in two different ways, the integral version and the differential version. As the name implies the integral version involves an integral and the differential version involves a derivative. These two-forms are equivalent. First we give the integral form of Gauss's law for electric fields:

$$\int_S \mathbf{E} \cdot \hat{n}\, dS = \frac{q_{\text{enc.}}}{\varepsilon_0},$$

where

- $\mathbf{E}$ is the electrostatic field produced by a charge,
- S is the closed surface,
- $\hat{n}$ is the unit normal to surface S,
- dS is the area element of surface S,
- $q_{\text{enc.}}$ is the charge enclosed in surface S, and
- ε_0 is a physical constant of proportionality called the permittivity of free space.

We should recognize that $\int_S \mathbf{E} \cdot \hat{n} dS$ as the flux of the field $\mathbf{E}$ through the surface S. The flux of $\mathbf{E}$ through the closed surface S which is generated by $q_{\text{enc.}}$ is proportional to $q_{\text{enc.}}$ with a constant of proportionality given by $\frac{1}{\varepsilon_0}$. We also need to

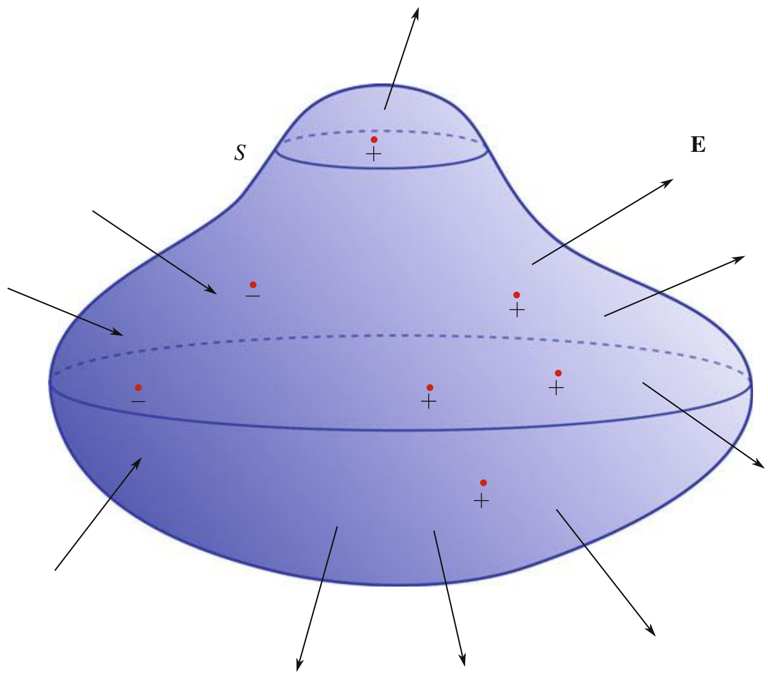


Fig. 12.3 A closed surface S enclosing some positive and negative charges. The net charge inside S is given by $q_{\text{enc.}}$. A few electric field lines penetrating S are drawn. Since there are more positive charges than negative charges there are more field lines going out of S than going into S

recognize that

$$q_{\text{enc.}} = \int_V \rho \, dV$$

where

- ρ is the charge density, a function on $\mathbb{R}^3$ that describes the amount of charge at each point.

We show what is going on in Fig. 12.3. A closed surface S encloses several point charges, some positive and some negative. There are more positive charges than negative enclosed and so when we integrate the electric vector field $\mathbf{E}$ over S we have a positive flux out of the surface.

We can move from the integral form of Gauss's law for electric fields to the differential form using the vector calculus divergence theorem,

$$\int_{S=\partial V} \mathbf{E} \cdot \hat{n} \, dS = \int_V \nabla \cdot \mathbf{E} \, dV.$$

Note that we are using the vector calculus notation of $\nabla \cdot \mathbf{E}$ for div $\mathbf{E}$. Putting this all together we get

$$\begin{aligned}
\int_V \nabla \cdot \mathbf{E} \, dV &= \int_S \mathbf{E} \cdot \hat{n} \, dS && \text{divergence theorem} \\
&= \frac{q_{\text{enc.}}}{\varepsilon_0} && \text{Gauss's law} \\
&= \frac{1}{\varepsilon_0} \int_V \rho \, dV && \text{definition of charge density} \\
&= \int_V \frac{\rho}{\varepsilon_0} \, dV
\end{aligned}$$

Equating the integrands of the first and the last terms we have the differential form of Gauss's law for electric fields,

$$\nabla \cdot \mathbf{E} = \frac{\rho}{\varepsilon_0}.$$

Question 12.1 Writing the electric field as $\mathbf{E} = E_1\hat{i} + E_2\hat{j} + E_3\hat{k}$ write Gauss's law for electric fields in terms of E_1, E_2, E_3. Do this twice, once for the integral form and once for the differential form of Gauss's law for electric fields.

Like electric charge, magnetism is an intrinsic property of subatomic particles, though in a number of ways it is a bit more complicated than electric charge and electric fields. It may be reasonable to expect that just as electric fields are produced by positive and negative electric charges, magnetic fields are produced by some sort of "magnetic charge." Unfortunately, this does not seem to be the case. Here we will consider the magnetic field produced by a spinning electron. The fields produced by spinning protons and neutrons are about a thousand times smaller than those produced by electrons and so can often be neglected.

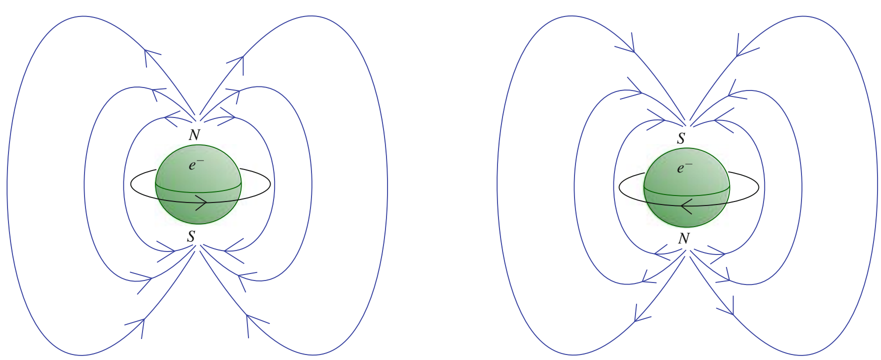
Just as with electric fields $\mathbf{E}$, magnetic fields, denoted by $\mathbf{B}$, exist at each point p in space $\mathbb{R}^3$,

$$\mathbf{B}_p = \begin{bmatrix} B_1 \\ B_2 \\ B_3 \end{bmatrix}_p .$$

The magnetic field can be used to give us magnetic field lines, which are just the integral curves of the magnetic field $\mathbf{B}$. It is these magnetic field lines, or integral curves of $\mathbf{B}$, and not $\mathbf{B}$ itself, that are drawn in Figs. 12.4, 12.5, and 12.6. Like the electric fields, the only way to know that the magnetic field is there is by observing how it interacts with tiny test particles. Like before we consider a tiny positively charged particle. However, figuring out the magnetic field vector $\mathbf{B}_p$ at a point p is somewhat more complicated than the case of finding $\mathbf{E}_p$. We won't go into the details here but it involves watching how a moving positive test particle acts.

Two spinning electrons are shown in Fig. 12.4. The magnetic field lines generated by the electron depend on what direction the electron is spinning. Of course, our picture of an electron as a tiny ball that actually spins is in itself a cartoon that helps us both visualize what an electron is like and think about how the physical properties behave. The actual quantum mechanical nature of subatomic particles like electrons is a lot more complicated than this picture would have you believe. But for our purposes this picture is sufficient.

Larger objects that produce magnetic fields are often simply called permanent magnets or simply magnets. Often when one thinks of magnets one imagines bar magnets or horseshoe shaped magnets as in Fig. 12.5, or refrigerator magnets as shown in Fig. 12.6. The materials that make up these magnets are magnetized and thus create their own persistent magnetic fields. This happens because the chemical make-up and crystalline microstructure of the material cause the magnetic fields of the subatomic particles to line up thereby producing a stronger magnetic field.



Here an electron "spins" in one direction resulting in magnetic field lines.

Here an electrons "spins" in the other direction again resulting in magnetic field lines.

Fig. 12.4 A cartoon of an electron as a little ball that can spin. The direction of the magnetic field lines depends on the spin of the electron, which we often visualize as a spinning ball. But keep in mind, the real quantum mechanical nature of electrons is very much more complicated than this picture

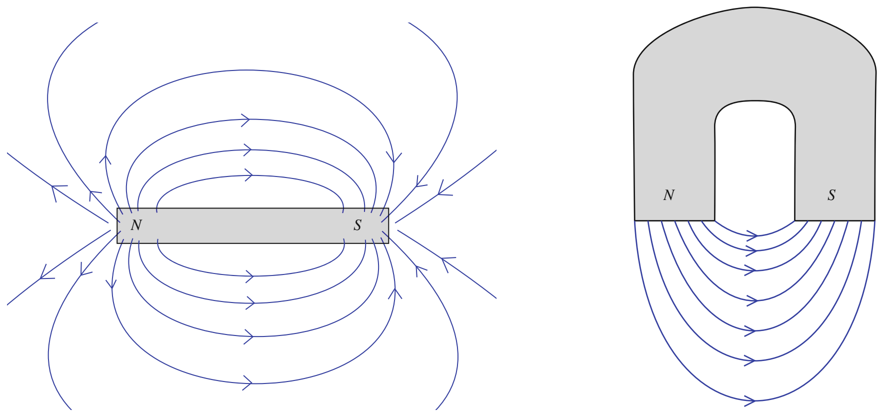


Fig. 12.5 A bar magnet (left) and a horseshoe magnet (right) with north and south poles depicted. The magnetic flux lines are shown in blue

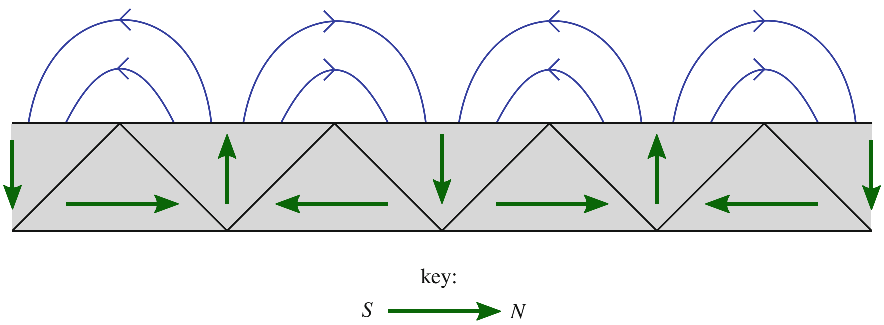


Fig. 12.6 A cross section of a refrigerator magnet. Notice the different domains where the magnetic north/south directions change. The green arrow points from the south pole to the north pole. The magnetic field lines are shown in blue. Due to the nature of the domains one side of the magnet has magnetic field lines and the other side does not

In some ways Gauss's law for magnetic fields is very similar to his law for electric fields, but its content is very different. This difference arises because it is possible to have a positive and a negative electric charge separated from each other. However, it is impossible to have a magnetic north pole separated from a magnetic south pole. That is, north poles and south poles always appear in pairs. Even a spinning electron has a north pole and a south pole, as in Fig. 12.4. No magnetic monopole, that is an isolated north pole or an isolated south pole, has ever been observed in nature or made in a laboratory. This means that the magnetic fields generated by a magnet, with both a north and a south pole, twist around back onto themselves, as we can see from Figs. 12.4, 12.5, and 12.6.

The end result of this twisting is that the total amount of magnetic field that exits a closed surface also enters that same closed surface so the net magnetic flux through the closed surface is always zero. This leads to **Gauss's law for magnetic fields**.

A magnetic field is produced by a magnet. The flux of this magnetic field through a closed surface zero.

Again, this law can all be expressed in vector calculus notation in two different ways, the integral version and the differential version. The integral version is given by

$$\int_S \mathbf{B} \cdot \hat{n} \, dS = 0,$$

where

- $\mathbf{B}$ is the magnetic field,
- S is the closed surface,

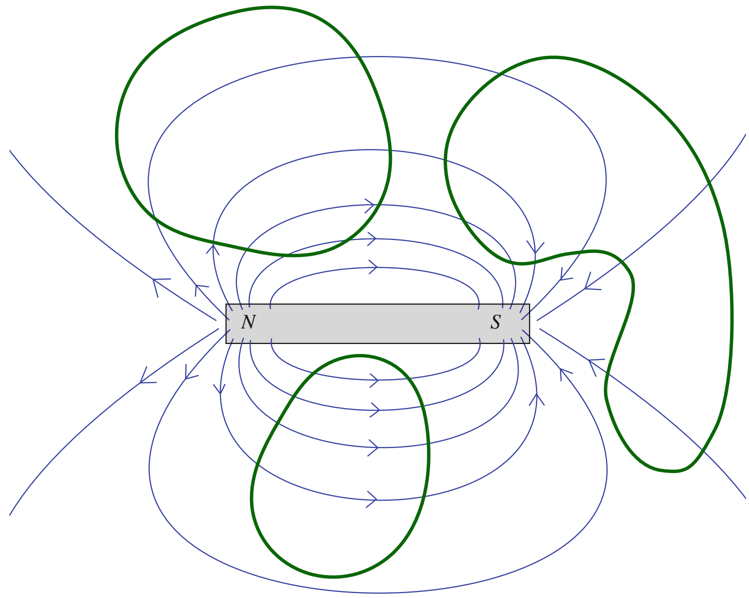


Fig. 12.7 A magnet with north and south poles is shown along with the magnetic field lines that "twist" back onto themselves. As can be seen with the green "closed surfaces," there are exactly as many magnetic field lines entering the surface as leaving the surface, and so the magnetic flux through any surface is always zero

- $\hat{n}$ is the unit normal to surface S, and
- dS is the area element of surface S.

We can see what Gauss's law for magnetic fields means by considering Fig. 12.7. The figure that is drawn is a two-dimensional cross-section of a magnet along with magnetic field lines that represent the vector field **B**. The closed blue curves represent closed surfaces. As we can see from the figure, each magnetic field line that enters a closed surface also leaves that closed surface.

Deriving the differential form of of Gauss's law for magnetic fields using the divergence theorem is then trivial

$$\begin{aligned}
\int_V \nabla \cdot \mathbf{B}\, dV &= \int_S \mathbf{B} \cdot \hat{n}\, dS && \text{divergence theorem} \\
&= 0 && \text{Gauss's law} \\
&= \int_V 0\, dV.
\end{aligned}$$

Equating the integrands of the first and last terms we have the differential form of Gauss's law for magnetic fields,

$$\nabla \cdot \mathbf{B} = 0.$$

Question 12.2 Writing the magnetic field as $\mathbf{B} = B_1\hat{i} + B_2\hat{j} + B_3\hat{k}$ write Gauss's law for magnetic fields in terms of B_1, B_2, B_3. Do this twice, once for the integral form and once for the differential form of Gauss's law for magnetic fields.

12.2 Faraday's Law and the Ampère-Maxwell Law

Both Faraday's law and the Ampère-Maxwell law explain the relationship between electric and magnetic fields. We begin by looking at Faraday's law. The induced electric field **E** in Faraday's law is similar to the electrostatic field **E** in Gauss's law for electric fields, but it is different in its structure. Both fields act as forces to accelerate a charged particle but electrostatic fields have field lines that originate on positive charges and terminate on negative charges, while induced electric fields have field lines that loop back on themselves. The statement of **Faraday's law** is quite simple.

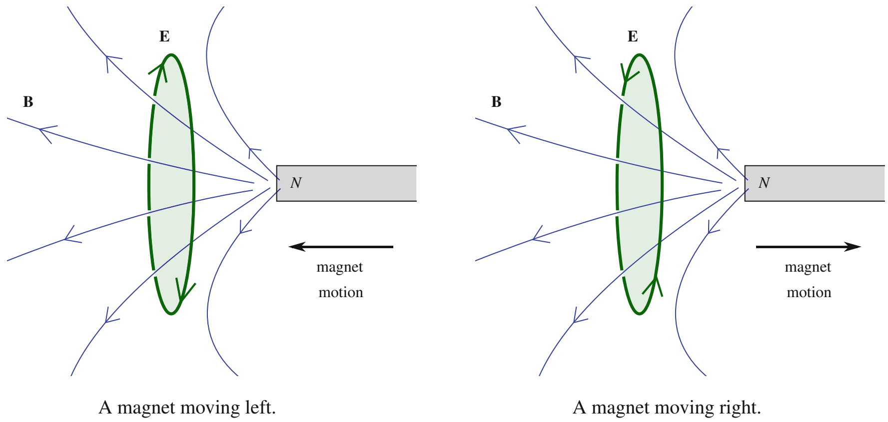


Fig. 12.8 A changing magnetic field produces a circulating electric field around the boundary of a surface. The direction the magnetic field is going in depends on if the magnetic flux is increasing or decreasing

The changing magnetic flux through a surface induces a circulating electric field around the boundary of the surface.

In other words, a changing magnetic flux through a surface S produces an electric field that circulates around the boundary of S, that is, around $C = \partial S$. Consider Fig. 12.8. When a magnet is moved then the magnetic flux going through the imaginary light green surface induces a electric field along the boundary of that surface. The direction of the magnetic field goes in depends on if the magnetic flux through the imaginary surface is increasing or decreasing. The key here is that the magnetic flux through the surface must be changing with respect to time. If the magnetic flux through the surface is not changing with respect to time then no electric field is induced.

The integral form for Faraday's law is given by

$$\int_C \mathbf{E} \cdot \hat{t} ds = -\frac{d}{dt} \int_S \mathbf{B} \cdot \hat{n} \, dS.$$

An alternative way of writing Faraday's law is given by

$$\int_C \mathbf{E} \cdot \hat{t} ds = -\int_S \frac{\partial \mathbf{B}}{\partial t} \cdot \hat{n} \, dS.$$

We will not go into the reasons why it is acceptable to switch the derivative and integral here, that we will leave for an analysis class. In both of these versions of Faraday's law we have that

- $\mathbf{E}$ is the induced electric field along the curve C,
- $C = \partial S$ is the closed curve, the boundary of surface S,
- $d\mathbf{s}$ is the line element of C,
- $\hat{t}$ is the unit tangent to C,
- $\mathbf{B}$ is the magnetic field,
- $\hat{n}$ is the unit normal to the surface S, and
- dS is the area element of S.

In order to move from the integral form of Faraday's law to the differential form of Faraday's law we will use the vector calculus version of Stokes' theorem,

$$\int_S (\nabla \times \mathbf{E}) \cdot \hat{n} \, dS = \int_{\partial S = C} \mathbf{E} \cdot \hat{t} \, ds.$$

Putting this all together we have

$$\begin{aligned}\int_S (\nabla \times \mathbf{E}) \cdot \hat{n}\, dS &= \int_{\partial S = C} \mathbf{E} \cdot \hat{t}\, ds && \text{Stoke's theorem} \\ &= -\int_S \frac{\partial \mathbf{B}}{\partial t} \cdot \hat{n}\, dS && \text{Faraday's law}.\end{aligned}$$

By equating the integrands of the first and last terms we have the differential form of Faraday's law

$$\nabla \times \mathbf{E} = -\frac{\partial \mathbf{B}}{\partial t}.$$

Question 12.3 Letting $\mathbf{E} = E_1\hat{i} + E_2\hat{j} + E_3\hat{k}$ and $\mathbf{E} = B_1\hat{i} + B_2\hat{j} + B_3\hat{k}$ write Faraday's law in terms of E_1, E_2, E_3 and B_1, B_2, B_3. You will end up with three separate equations. Do this twice, once for the integral form of Faraday's law and once for the differential form of Faraday's law.

Now we turn our attention to the Ampère-Maxwell law. This law is a little different from the other law in that the right hand side has two terms, one term that involves an electric current and another term that involves an electric flux. The **Ampère-Maxwell law** states that

An electric current or a changing electric flux through a surface produces a circulating magnetic field around the surface boundary.

Unlike the previous laws this law involves two terms on the right hand side. The first term relates to how an electric current induces a magnetic field. Suppose we have a wire with and electric current I flowing along it. An electric current is just moving electrons. If we have a surface S with a boundary $C = \partial S$ and the wire with the current goes through, or penetrates, the surface, then the current induces a magnetic field along the boundary of the surface. If the wire does not penetrate the surface, then there is no induced magnetic field along the boundary. See Fig. 12.9. Of course, this is true for any surface that the wire penetrates. Consider Fig. 12.10. The magnetic field is induced all along the length of the wire.

The second term involves a changing electric field. In Fig. 12.11 we show three charges, two negative and one positive, along with a surface S with boundary $C = \partial S$. These three charges produce an electric field. A few electric field lines are shown and as we can see, some of the electric field lines penetrate the surface. When the three charges are moving, as is represented by the velocity arrows attached to the charges, then the electric field that the moving charges induce change over time. This means that the electric flux through the surface is also changing, thereby inducing a magnetic field around the boundary of the surface.

The Ampère-Maxwell law combines these two different ways that a magnetic field can be induced. Figure 12.12 tries to show both occurring at the same time. The electric charges are moving, thus causing the electric field to change, and a current is flowing through the wire. The magnetic field along the boundary of the surface is a sum of the fields produced by

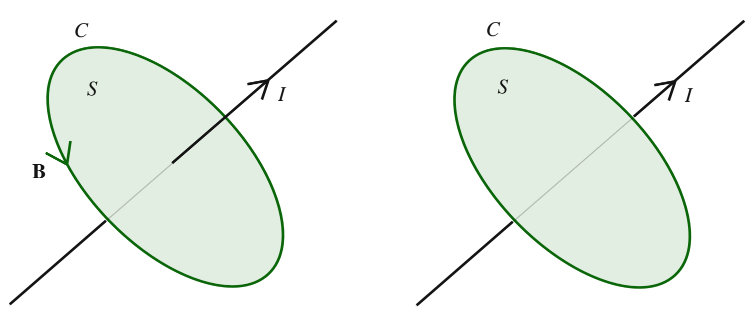


Fig. 12.9 An electric current I traveling along a wire. On the left the wire with the current "penetrates" the green surface S thereby producing a circulating magnetic field around the boundary of the surface $C = \partial S$. Since the current penetrates the surface we would write $I_{\text{enc.}}$. On the right the wire with the electric current does not penetrate the surface resulting in no magnetic field around the boundary

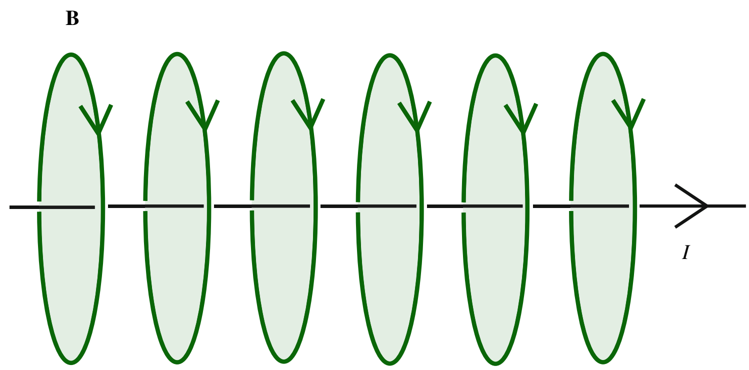


Fig. 12.10 Similar to Fig. 12.9, an electric current traveling down a wire induces magnetic fields around the wire. Of course the magnetic field is not restricted to the boundary of one surface, it occurs on the boundary of every surface through which the wire penetrates

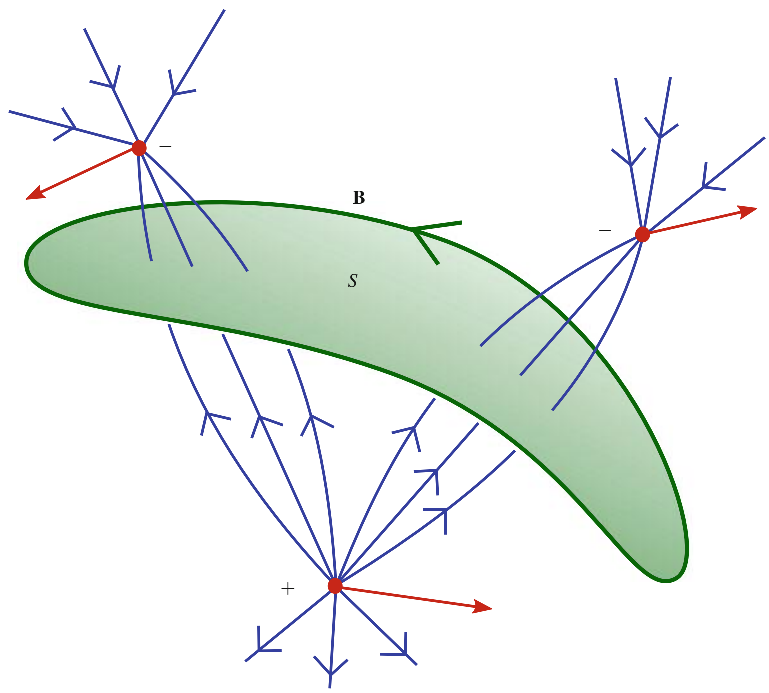


Fig. 12.11 Here three charges (red) induce and electric field. A few electric field lines are shown (blue), which penetrate the surface S (green). If the charges are moving, as represented by the velocity vectors emanating from the charges (red), then the electric field is changing with respect to time, thereby inducing a magnetic field along the surface boundary, ∂S

each of these. We are now ready for the mathematical statement of the Ampère-Maxwell law. The integral form is given by

$$\int_C \mathbf{B} \cdot \hat{t}\, ds = \mu_0 I_{\text{enc.}} + \mu_0 \varepsilon_0 \frac{d}{dt} \int_S \mathbf{E} \cdot \hat{n}\, dS.$$

An alternative way of writing the Ampère-Maxwell law is given by

$$\int_C \mathbf{B} \cdot \hat{t}\, ds = \mu_0 I_{\text{enc.}} + \mu_0 \varepsilon_0 \int_S \frac{\partial \mathbf{E}}{\partial t} \cdot \hat{n}\, dS.$$

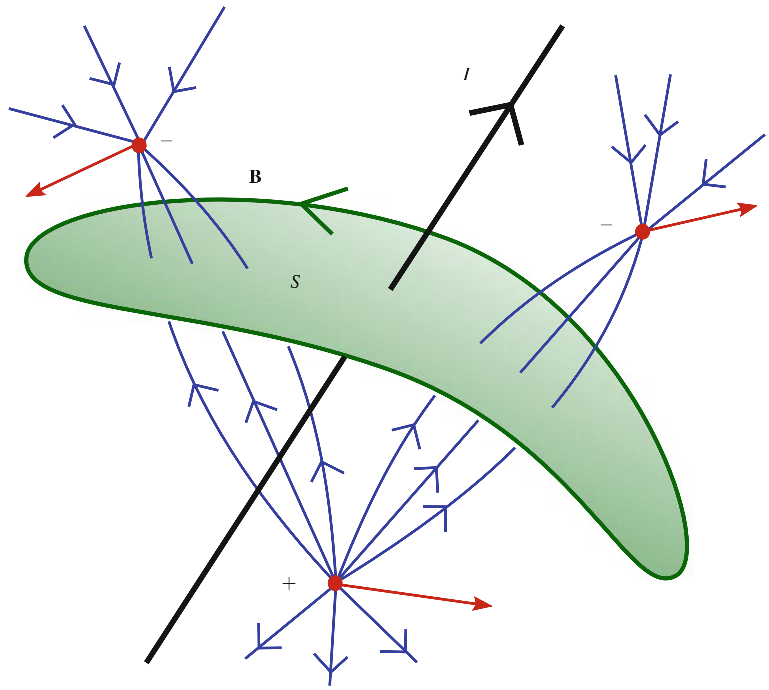


Fig. 12.12 Here the effects of both a current along a wire, as in Fig. 12.9 left, and a changing electric field, as in Fig. 12.11, are combined to product the magnetic field along ∂S

Again, we will not go into the reasons why it is acceptable to switch the derivative and integral sign here. In both of these versions we have

- μ_0 is the magnetic permittivity of space, and
- $I_{\text{enc.}}$ is the enclosed electrical current.

Since electrical current is nothing more than flowing electrons, we have that the current I is the integral of the flowing charge density $\mathbf{J} = J_1\hat{i} + J_2\hat{j} + J_3\hat{k}$,

$$I = \int_S \mathbf{J} \cdot \hat{n}\, dS.$$

Again, using the vector calculus version of Stokes' theorem we have

$$\begin{aligned}
\int_S (\nabla \times \mathbf{B}) \cdot \hat{n}\, dS &= \int_C \mathbf{B} \cdot \hat{t}\, ds && \text{Stoke's theorem} \\
&= \mu_0 I_{\text{enc.}} + \mu_0\varepsilon_0 \int_S \frac{\partial \mathbf{E}}{\partial t} \cdot \hat{n}\, dS && \text{Ampère-Maxwell law} \\
&= \mu_0 \int_S \mathbf{J} \cdot \hat{n}\, dS + \mu_0\varepsilon_0 \int_S \frac{\partial \mathbf{E}}{\partial t} \cdot \hat{n}\, dS && \text{Def. charge density} \\
&= \int_S \mu_0 \mathbf{J} \cdot \hat{n}\, dS + \int_S \mu_0\varepsilon_0 \frac{\partial \mathbf{E}}{\partial t} \cdot \hat{n}\, dS \\
&= \int_S \left(\mu_0 \mathbf{J} + \mu_0\varepsilon_0 \frac{\partial \mathbf{E}}{\partial t} \right) \cdot \hat{n}\, dS.
\end{aligned}$$

The second last equality is simply a change in notation. Equating the integrands of the first and last terms we get the differential form of Ampère-Maxwell law

$$\nabla \times \mathbf{B} = \mu_0 \mathbf{J} + \mu_0 \varepsilon_0 \frac{\partial \mathbf{E}}{\partial t}.$$

Question 12.4 Letting $\mathbf{E} = E_1\hat{i} + E_2\hat{j} + E_3\hat{k}$, $\mathbf{E} = B_1\hat{i} + B_2\hat{j} + B_3\hat{k}$, and $\mathbf{J} = J_1\hat{i} + J_2\hat{j} + J_3\hat{k}$, write Ampère-Maxwell law in terms of $E_1, E_2, E_3, B_1, B_2, B_3$, and J_1, J_2, J_3. You will end up with three separate equations. Do this twice, once for the integral form of Ampère-Maxwell law and once for the differential form of Ampère-Maxwell law.

We will henceforth blithely assume that units are chosen so that $\mu_0 = \varepsilon_0 = 1$. This saves us from having to carry a bunch of constants around with us everywhere and drives physicists and engineers crazy. This gives us the following summary of Maxwell's equations,

Maxwell's Equations:
version one

Integral Form		Differential Form
$\int_S \mathbf{E} \cdot \hat{n}\, dS = q_{\text{enc.}}$	$\Longleftrightarrow$	$\nabla \cdot \mathbf{E} = \rho$
$\int_S \mathbf{B} \cdot \hat{n}\, dS = 0$	$\Longleftrightarrow$	$\nabla \cdot \mathbf{B} = 0$
$\int_C \mathbf{E} \cdot \hat{t}\, ds = -\int_S \frac{\partial \mathbf{B}}{\partial t} \cdot \hat{n}\, dS$	$\Longleftrightarrow$	$\nabla \times \mathbf{E} = -\frac{\partial \mathbf{B}}{\partial t}$
$\int_C \mathbf{B} \cdot \hat{t}\, ds = I_{\text{enc.}} + \int_S \frac{\partial \mathbf{E}}{\partial t} \cdot \hat{n}\, dS$	$\Longleftrightarrow$	$\nabla \times \mathbf{B} = \mathbf{J} + \frac{\partial \mathbf{E}}{\partial t}$

12.3 Special Relativity and Hodge Duals

Now we want to recast electromagnetism in terms of differential forms. There are actually a couple ways to do this. First, it could be done using forms on a three-dimensional Euclidian space and keeping time as a separate variable. However, this approach ends up being less than elegant. In fact, this approach fails to harness the full power of differential forms and fails to illuminate one of electromagnetism's most interesting features, that it is compatible with special relativity. It is interesting to note that of the two great areas of nineteenth century physics, classical mechanics and classical electromagnetism, it is classical electromagnetism that is compatible with special relativity and not classical mechanics.

In other words, Maxwell's equations require no modifications at high speeds, not surprising since they are related to electromagnetic fields which propagate at the speed of light. However, classical mechanics is drastically inaccurate at velocities that approach the speed of light. Maxwell's theory of electromagnetism is already fully relativistic, a fact that is clearly seen when we move from the vector calculus formulation to the differential forms formulation of the theory.

However, before moving to the differential forms formulation of Maxwell's equations we need to make a few modifications. First, we will need to move to four dimensions. But instead of our standard Euclidian space $\mathbb{R}^4$ we will consider Minkowski space, the four dimensional manifold on which special relativity is built. The Minkowski manifold is sometimes denoted M^4 or $\mathbb{R}^{1,3}$, where the time variable is separated from the three space variables. We will use the $\mathbb{R}^{1,3}$ notation.

Minkowski space $\mathbb{R}^{1,3}$ differs from Euclidian space $\mathbb{R}^4$ because it has a different metric tensor. Metrics were discussed informally in Sect. 10.1 and will be defined and discussed formally in Sect. A.6. This different metric tensor of Minkowski space means that $\mathbb{R}^{1,3}$ has a different inner product, or dot product, than $\mathbb{R}^4$ does. We first consider that the Euclidian metric with respect to the standard basis on $\mathbb{R}^4$. This Euclidian metric can be denoted by the matrix

$$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

The inner product, or dot product, can be computed using this matrix. This is something we saw in Sect. 5.6. Let v and w be two vectors on $\mathbb{R}^4$ at a point $p \in \mathbb{R}^4$. That is, $v, w \in T_p\mathbb{R}^4$. If the standard basis for $\mathbb{R}^4$ is denoted by e_1, e_2, e_3, e_4 then we have $v = v^i e_i$ and $w = w^i e_i$, $1 \leq i \leq 4$, in Einstein summation notation. Then the inner product can be determined by

$$\begin{aligned} v \cdot w \equiv \langle v, w \rangle \equiv v^T \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} w \\ = [v^1, v^2, v^3, v^4] \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} w^1 \\ w^2 \\ w^3 \\ w^4 \end{bmatrix} \\ = v^1 w^1 + v^2 w^2 + v^3 w^3 + v^4 w^4. \end{aligned}$$

The metric on Minkowski space $\mathbb{R}^{1,3}$, which is called the Minkowski metric, is defined a little differently. We will first denote the standard basis of $\mathbb{R}^{1,3}$ by e_0, e_1, e_2, e_3, with e_0 being the unit vector in the time direction, e_1 the unit vector in the x direction, e_2 the unit vector in the y direction, and e_3 the unit vector in the z direction. The metric on $\mathbb{R}^{1,3}$ with respect to this basis is denoted by the matrix

$$\begin{bmatrix} -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

We should note here that the matrix

$$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{bmatrix}$$

is used in some books instead. This simply results in all signs being switched. Which version of this matrix is used is really a matter of taste and convention, though the first one seems to be used somewhat more often. But you should double check to see which convention is being used in whatever book you are reading.

Let v and w be two vectors on $\mathbb{R}^{1,3}$ at a point $p \in \mathbb{R}^{1,3}$, that is, $v, w \in T_p\mathbb{R}^{1,3}$. Then we have $v = v^i e_i$ and $w = w^i e_i$, $0 \leq i \leq 3$, in Einstein summation notation. The inner product can be determined by

$$
\begin{aligned}
\langle v, w \rangle &\equiv v^T \begin{bmatrix} -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} w \\
&= [v^0, v^1, v^2, v^3] \begin{bmatrix} -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} w^0 \\ w^1 \\ w^2 \\ w^3 \end{bmatrix} \\
&= -v^0 w^0 + v^1 w^1 + v^2 w^2 + v^3 w^3.
\end{aligned}
$$

Notice that we did not use the dot product notation here. The dot product $v \cdot w$ is generally reserved for when we are using the Euclidian metric and the bracket notation $\langle v, w \rangle$ is often used with other metrics. Also, if you recall, we have used the bracket notation for the pairing between one-forms and vectors, $\langle \alpha, v \rangle \equiv \alpha(v)$, and once in a while you will even see it used for k-forms, $\langle \alpha, (v_1, \ldots, v_k) \rangle \equiv \alpha(v_1, \ldots, v_k)$. It is easy enough to tell in what sense $\langle \cdot, \cdot \rangle$ is being used, just look to see what is inside, two vectors or a form and a vector or set of vectors.

We now need to make some definitions and get our notation in order. First, we will use x^1 to denote x, x^2 to denote y, x^3 to denote z. Either t is left unchanged or else it is denoted by x^0. For the moment we will continue to use t instead of x^0 to help you remember that it is the time variable. We have the following vector spaces defined at each point of Minkowski space $\mathbb{R}^{1,3}$:

$$
\begin{aligned}
f, g &\in \bigwedge\nolimits_p^0 \left(\mathbb{R}^{1,3}\right) = \text{ continuous functions on } \mathbb{R}^{1,3}, \\
\alpha, \beta &\in \bigwedge\nolimits_p^1 \left(\mathbb{R}^{1,3}\right) = T_p^*\mathbb{R}^{1,3} = \text{span}\left\{dt, dx^1, dx^2, dx^3\right\}, \\
\gamma, \delta &\in \bigwedge\nolimits_p^2 \left(\mathbb{R}^{1,3}\right) = \text{span}\left\{dt \wedge dx^1, dt \wedge dx^2, dt \wedge dx^3, dx^1 \wedge dx^2, dx^2 \wedge dx^3, dx^3 \wedge dx^1\right\}, \\
\omega, \nu &\in \bigwedge\nolimits_p^3 \left(\mathbb{R}^{1,3}\right) = \text{span}\left\{dt \wedge dx^1 \wedge dx^2, dt \wedge dx^3 \wedge dx^1, dt \wedge dx^2 \wedge dx^3, dx^1 \wedge dx^2 \wedge dx^3\right\}, \\
\varepsilon, \theta &\in \bigwedge\nolimits_p^4 \left(\mathbb{R}^{1,3}\right) = \text{span}\left\{dt \wedge dx^1 \wedge dx^2 \wedge dx^3\right\}.
\end{aligned}
$$

As a matter of convention when writing the basis elements of the various spaces we write dt first and the other terms in cyclic order. We will write

$$
\begin{aligned}
\alpha &= \alpha_0 dt + \alpha_1 dx^1 + \alpha_2 dx^2 + \alpha_3 dx^3, \\
\gamma &= \gamma_{01} dt \wedge dx^1 + \gamma_{02} dt \wedge dx^2 + \gamma_{03} dt \wedge dx^3 + \gamma_{12} dx^1 \wedge dx^2 + \gamma_{23} dx^2 \wedge dx^3 + \gamma_{31} dx^3 \wedge dx^1, \\
\omega &= \omega_{012} dt \wedge dx^1 \wedge dx^2 + \omega_{031} dt \wedge dx^3 \wedge dx^1 + \omega_{023} dt \wedge dx^2 \wedge dx^3 + \omega_{123} dx^1 \wedge dx^2 \wedge dx^3, \\
\varepsilon &= \varepsilon_{0123} dt \wedge dx^1 \wedge dx^2 \wedge dx^3,
\end{aligned}
$$

with $\beta, \delta, \nu, \theta$ being written similarly. The subscripts that keep track of our coefficients correspond with the basis elements.

We get the space $\bigwedge^0 \left(\mathbb{R}^{1,3}\right)$ by taking the disjoint union of all the spaces $\bigwedge_p^0 \left(\mathbb{R}^{1,3}\right)$ for each $p \in \mathbb{R}^{1,3}$, thereby making $\bigwedge^0 \left(\mathbb{R}^{1,3}\right)$ a bundle over the manifold $\mathbb{R}^{1,3}$. The spaces $\bigwedge^1 \left(\mathbb{R}^{1,3}\right)$, $\bigwedge^2 \left(\mathbb{R}^{1,3}\right)$, $\bigwedge^3 \left(\mathbb{R}^{1,3}\right)$, and $\bigwedge^4 \left(\mathbb{R}^{1,3}\right)$ are obtained similarly. We can then define inner products on these spaces.

Due to the number of basis elements for some of these spaces we leave it as an exercise to write the metric matrices of these vector spaces. However, all inner products are essentially defined along the same lines as the inner product was defined when we used the Minkowski metric of the underlying manifold $\mathbb{R}^{1,3}$; for any term that had a dt in it we have a negative sign. Thus we have

$$
\begin{aligned}
f, g &\in \bigwedge^0\left(\mathbb{R}^{1,3}\right); \quad \langle f, g\rangle = f \cdot g,\\
\alpha, \beta &\in \bigwedge^1\left(\mathbb{R}^{1,3}\right); \quad \langle \alpha, \beta\rangle = -\alpha_0\beta_0 + \alpha_1\beta_1 + \alpha_2\beta_2 + \alpha_3\beta_3,\\
\gamma, \delta &\in \bigwedge^2\left(\mathbb{R}^{1,3}\right); \quad \langle \gamma, \delta\rangle = -\gamma_{01}\delta_{01} - \gamma_{02}\delta_{02} - \gamma_{03}\delta_{03} + \gamma_{12}\delta_{12} + \gamma_{23}\delta_{23} + \gamma_{31}\delta_{31},\\
\omega, \nu &\in \bigwedge^3\left(\mathbb{R}^{1,3}\right); \quad \langle \omega, \nu\rangle = -\omega_{012}\nu_{012} - \omega_{023}\nu_{023} - \omega_{031}\nu_{031} + \omega_{123}\nu_{123},\\
\varepsilon, \theta &\in \bigwedge^4\left(\mathbb{R}^{1,3}\right); \quad \langle \varepsilon, \theta\rangle = -\varepsilon_{0123}\theta_{0123}.
\end{aligned}
$$

The inner products are defined pointwise, so for example at the point p we have $\langle f, g\rangle_p = f(p)g(p)$, $\langle \alpha, \beta\rangle_p = -\alpha_0(p)\beta_0(p) + \alpha_1(p)\beta_1(p) + \alpha_2(p)\beta_2(p) + \alpha_3(p)\beta_3(p)$, and so on. Thus if all functions used to define the forms are smooth functions on the manifold $\mathbb{R}^{1,3}$ then the inner product of two k-forms, $k = 0, 1, 2, 3, 4$, is a smooth function on the $\mathbb{R}^{1,3}$. For example, we have $\langle f, g\rangle : \mathbb{R}^{1,3} \to \mathbb{R}$, and so on.

Question 12.5 Write the metric matrices for the spaces $\bigwedge^0\left(\mathbb{R}^{1,3}\right)$, $\bigwedge^1\left(\mathbb{R}^{1,3}\right)$, $\bigwedge^2\left(\mathbb{R}^{1,3}\right)$, $\bigwedge^3\left(\mathbb{R}^{1,3}\right)$, and $\bigwedge^4\left(\mathbb{R}^{1,3}\right)$.

We need to figure out the Hodge star operator using these new inner products. On $\mathbb{R}^{1,3}$ the Hodge star dual of a differential form β will be given by the formula

$$\alpha \wedge *\beta = \langle \alpha, \beta\rangle \, dt \wedge dx^1 \wedge dx^2 \wedge dx^3.$$

If α is a basis element then to find $*\alpha$ the formula reduces to

$$\alpha \wedge *\alpha = \langle \alpha, \alpha\rangle \, dt \wedge dx^1 \wedge dx^2 \wedge dx^3 = \pm dt \wedge dx^1 \wedge dx^2 \wedge dx^3$$

where the minus or plus depends on whether the element α had a dt in it or not.

Now we find the Hodge star dual of the one-form basis elements. Notice that we always put dt first and all other elements in cyclic order,

$$
\begin{aligned}
dt \wedge *dt &= \langle dt, dt\rangle \, dt \wedge dx^1 \wedge dx^2 \wedge dx^3\\
&= -dt \wedge dx^1 \wedge dx^2 \wedge dx^3\\
\Rightarrow *dt &= -dx^1 \wedge dx^2 \wedge dx^3,
\end{aligned}
$$

$$
\begin{aligned}
dx^1 \wedge *dx^1 &= \langle dx^1, dx^1\rangle \, dt \wedge dx^1 \wedge dx^2 \wedge dx^3\\
&= dt \wedge dx^1 \wedge dx^2 \wedge dx^3\\
&= -dx^1 \wedge dt \wedge dx^2 \wedge dx^3\\
\Rightarrow *dx^1 &= -dt \wedge dx^2 \wedge dx^3,
\end{aligned}
$$

The calculations for the other two basis elements are similar. Thus the Hodge duals of the one-form basis elements are

$$
\begin{aligned}
*dt &= -dx^1 \wedge dx^2 \wedge dx^3,\\
*dx^1 &= -dt \wedge dx^2 \wedge dx^3,\\
*dx^2 &= -dt \wedge dx^3 \wedge dx^1,\\
*dx^3 &= -dt \wedge dx^1 \wedge dx^2.
\end{aligned}
$$

Question 12.6 Do the calculation to find the Hodge star duals of dx_2 and dx_3.

The Hodge dual of the two-form basis elements are found similarly. Again we always put dt first and put the other elements in cyclic order as necessary,

$$\begin{aligned}(dt \wedge dx^1) \wedge *(dt \wedge dx^1) &= \langle dt \wedge dx^1, dt \wedge dx^1 \rangle \, dt \wedge dx^1 \wedge dx^2 \wedge dx^3 \\ &= -(dt \wedge dx^1) \wedge (dx^2 \wedge dx^3) \\ \Rightarrow *(dt \wedge dx^1) &= -dx^2 \wedge dx^3,\end{aligned}$$

$$\begin{aligned}(dt \wedge dx^2) \wedge *(dt \wedge dx^2) &= \langle dt \wedge dx^2, dt \wedge dx^2 \rangle \, dt \wedge dx^1 \wedge dx^2 \wedge dx^3 \\ &= -dt \wedge dx^2 \wedge dx^3 \wedge dx^1 \\ \Rightarrow *(dt \wedge dx^2) &= -dx^3 \wedge dx^1.\end{aligned}$$

The calculations for the other basis elements are similar. In summary the Hodge duals of the two-form basis elements are

$$\begin{aligned}*(dt \wedge dx^1) &= -dx^2 \wedge dx^3, \\ *(dt \wedge dx^2) &= -dx^3 \wedge dx^1, \\ *(dt \wedge dx^3) &= -dx^1 \wedge dx^2, \\ *(dx^1 \wedge dx^2) &= dt \wedge dx^2, \\ *(dx^2 \wedge dx^3) &= dt \wedge dx^1, \\ *(dx^3 \wedge dx^1) &= dt \wedge dx^2.\end{aligned}$$

Question 12.7 Do the calculation to find the Hodge star dual of the remaining two-form basis elements.

Question 12.8 Find $*\alpha$ for $\alpha = 7dt \wedge dx^2 - 20dt \wedge d^3 + 6dx^2 \wedge dx^3 - 2dx^3 \wedge dx^1$.

And finally the Hodge star duals of the three-form basis elements are given by

$$\begin{aligned}*(dt \wedge dx^1 \wedge dx^2) &= -dx^3, \\ *(dt \wedge dx^2 \wedge dx^3) &= -dx^1, \\ *(dt \wedge dx^3 \wedge dx^1) &= -dx^2, \\ *(dx^1 \wedge dx^2 \wedge dx^3) &= -dt.\end{aligned}$$

Question 12.9 Do the calculations to find the Hodge star duals of the three-form basis elements.

12.4 Differential Forms Formulation

Now that we have the necessary mathematical foundations in place we are ready to embark on the main purpose of this chapter, to write Maxwell's equations using differential forms. Recall that in the vector calculus notation electric and magnetic fields were written as

$$\begin{aligned}\mathbf{E} &= E_1\hat{i} + E_2\hat{j} + E_3\hat{k}, \\ \mathbf{B} &= B_1\hat{i} + B_2\hat{j} + B_3\hat{k}.\end{aligned}$$

We will combine these two fields to make a particular kind of two-form on $\mathbb{R}^{1,3}$ called the Faraday two-form. Even though forms are generally denoted by Greek letters, for historical reasons this particular form is denoted by $\mathbf{F}$. Also, the subscripts on the coefficients of this two-form do not follow the subscript convention for two-forms, instead the same subscripts from $\mathbf{E}$ and $\mathbf{B}$ are retained. The Faraday two-form is

Faraday two-form
$$\boxed{\begin{aligned}\mathbf{F} = &-E_1\, dt \wedge dx^1 - E_2\, dt \wedge dx^2 - E_3\, dt \wedge dx^3 \\ &+B_1\, dx^2 \wedge dx^3 + B_2\, dx^3 \wedge dx^1 + B_3\, dx^1 \wedge dx^2.\end{aligned}}$$

We now slog through the computation of $d\mathbf{F}$,

$$\begin{aligned}
d\mathbf{F} = &-dE_1 \wedge dt \wedge dx^1 - dE_2 \wedge dt \wedge dx^2 - dE_3 \wedge dt \wedge dx^3 \\
&+ dB_1 \wedge dx^2 \wedge dx^3 + dB_2 \wedge dx^3 \wedge dx^1 + dB_3 \wedge dx^1 \wedge dx^2 \\
\\
= &-\left(\frac{\partial E_1}{\partial t}dt + \frac{\partial E_1}{\partial x^1}dx^1 + \frac{\partial E_1}{\partial x^2}dx^2 + \frac{\partial E_1}{\partial x^3}dx^3\right) \wedge dt \wedge dx^1 \\
&-\left(\frac{\partial E_2}{\partial t}dt + \frac{\partial E_2}{\partial x^2}dx^1 + \frac{\partial E_2}{\partial x^2}dx^2 + \frac{\partial E_2}{\partial x^3}dx^3\right) \wedge dt \wedge dx^2 \\
&-\left(\frac{\partial E_3}{\partial t}dt + \frac{\partial E_3}{\partial x^2}dx^1 + \frac{\partial E_3}{\partial x^2}dx^2 + \frac{\partial E_3}{\partial x^3}dx^3\right) \wedge dt \wedge dx^3 \\
&+\left(\frac{\partial B_1}{\partial t}dt + \frac{\partial B_1}{\partial x^1}dx^1 + \frac{\partial B_1}{\partial x^2}dx^2 + \frac{\partial B_1}{\partial x^3}dx^3\right) \wedge dx^2 \wedge dx^3 \\
&+\left(\frac{\partial B_2}{\partial t}dt + \frac{\partial B_2}{\partial x^2}dx^1 + \frac{\partial B_2}{\partial x^2}dx^2 + \frac{\partial B_2}{\partial x^3}dx^3\right) \wedge dx^3 \wedge dx^1 \\
&+\left(\frac{\partial B_3}{\partial t}dt + \frac{\partial B_3}{\partial x^2}dx^1 + \frac{\partial B_3}{\partial x^2}dx^2 + \frac{\partial B_3}{\partial x^3}dx^3\right) \wedge dx^1 \wedge dx^2 \\
\\
= &-\frac{\partial E_1}{\partial x^2}dx^2 \wedge dt \wedge dx^1 - \frac{\partial E_1}{\partial x^3}dx^3 \wedge dt \wedge dx^1 \\
&-\frac{\partial E_2}{\partial x^1}dx^1 \wedge dt \wedge dx^2 - \frac{\partial E_2}{\partial x^3}dx^2 \wedge dt \wedge dx^2 \\
&-\frac{\partial E_3}{\partial x^1}dx^1 \wedge dt \wedge dx^3 - \frac{\partial E_3}{\partial x^2}dx^2 \wedge dt \wedge dx^3 \\
&+\frac{\partial B_1}{\partial t}dt \wedge dx^2 \wedge dx^3 + \frac{\partial B_1}{\partial x^1}dx^1 \wedge dx^2 \wedge dx^3 \\
&+\frac{\partial B_2}{\partial t}dt \wedge dx^3 \wedge dx^1 + \frac{\partial B_2}{\partial x^2}dx^2 \wedge dx^3 \wedge dx^1 \\
&+\frac{\partial B_3}{\partial t}dt \wedge dx^1 \wedge dx^2 + \frac{\partial B_3}{\partial x^3}dx^3 \wedge dx^1 \wedge dx^2 \\
\\
= &\left(\frac{\partial E_2}{\partial x^1} - \frac{\partial E_1}{\partial x^2} + \frac{\partial B_3}{\partial t}\right) dt \wedge dx^1 \wedge dx^2 \\
&+\left(\frac{\partial E_3}{\partial x^2} - \frac{\partial E_2}{\partial x^3} + \frac{\partial B_1}{\partial t}\right) dt \wedge dx^2 \wedge dx^3
\end{aligned}$$

$$+\left(\frac{\partial E_1}{\partial x^3} - \frac{\partial E_3}{\partial x^1} + \frac{\partial B_2}{\partial t}\right) dt \wedge dx^3 \wedge dx^1$$
$$+\left(\frac{\partial B_1}{\partial x^1} + \frac{\partial B_2}{\partial x^2} + \frac{\partial B_3}{\partial x^3}\right) dx^1 \wedge dx^2 \wedge dx^3.$$

By letting $d\mathbf{F} = 0$ we have that each of the coefficients is equal to zero, thereby giving us the following four equations

$$\frac{\partial E_2}{\partial x^1} - \frac{\partial E_1}{\partial x^2} + \frac{\partial B_3}{\partial t} = 0,$$
$$\frac{\partial E_3}{\partial x^2} - \frac{\partial E_2}{\partial x^3} + \frac{\partial B_1}{\partial t} = 0,$$
$$\frac{\partial E_1}{\partial x^3} - \frac{\partial E_3}{\partial x^1} + \frac{\partial B_2}{\partial t} = 0,$$
$$\frac{\partial B_1}{\partial x^1} + \frac{\partial B_2}{\partial x^2} + \frac{\partial B_3}{\partial x^3} = 0.$$

Consider the differential form of Gauss's law for magnetic fields, $\nabla \cdot \mathbf{B} = 0$. Recalling that $\mathbf{B} = B_1\hat{i} + B_2\hat{j} + B_3\hat{k}$ we have that

$$\nabla \cdot \mathbf{B} = 0 \quad \Longleftrightarrow \quad \frac{\partial B_1}{\partial x^1} + \frac{\partial B_2}{\partial x^2} + \frac{\partial B_3}{\partial x^3} = 0.$$

Thus the fourth term is equivalent to Gauss's law for magnetic fields. Now consider the differential form of Faraday's law, $\nabla \times \mathbf{E} = -\frac{\partial \mathbf{B}}{\partial t}$. Recalling the definition of $\mathbf{B}$ as well as the definition of $\mathbf{E} = E_1\hat{i} + E_2\hat{j} + E_3\hat{k}$ it is straightforward to show that

$$\nabla \times \mathbf{E} = -\frac{\partial \mathbf{B}}{\partial t} \quad \Longleftrightarrow \quad \begin{cases} \frac{\partial E_2}{\partial x^1} - \frac{\partial E_1}{\partial x^2} + \frac{\partial B_3}{\partial t} = 0, \\ \frac{\partial E_3}{\partial x^2} - \frac{\partial E_2}{\partial x^3} + \frac{\partial B_1}{\partial t} = 0, \\ \frac{\partial E_1}{\partial x^3} - \frac{\partial E_3}{\partial x^1} + \frac{\partial B_2}{\partial t} = 0. \end{cases}$$

Thus we have that $d\mathbf{F} = 0$ is equivalent to both Gauss's law for magnetic fields and Faraday's law,

$$\boxed{d\mathbf{F} = 0 \quad \Longleftrightarrow \quad \nabla \cdot \mathbf{B} = 0 \ \text{ and } \ \nabla \times \mathbf{E} = -\frac{\partial \mathbf{B}}{\partial t}.}$$

Recall now the differential form of Gauss's law of electric fields $\nabla \cdot \mathbf{E} = \rho$ and the differential form of the Ampère-Maxwell law $\nabla \times \mathbf{B} = \mathbf{J} + \frac{\partial \mathbf{E}}{\partial t}$, where ρ is the charge density function and $\mathbf{J}$ is the current flux vector field, given by

$$\mathbf{J} = J_1\hat{i} + J_2\hat{j} + J_3\hat{k}.$$

We will now define an electric current density one-form $\mathbf{J}$ as

Electric current density one-form	$\mathbf{J} = -\rho dt + J_1 dx^1 + J_2 dx^2 + J_3 dx^3.$

We will show that the differential form of Gauss's law for electric currents and the differential form of the Ampère-Maxwell law are equivalent to $*d*\mathbf{F} = \mathbf{J}$. We will compute the left hand side of this equation in steps. First we have

$$\mathbf{F} = -E_1\, dt \wedge dx^1 - E_2\, dt \wedge dx^2 - E_3\, dt \wedge dx^3$$
$$+ B_1\, dx^2 \wedge dx^3 + B_2\, dx^3 \wedge dx^1 + B_3\, dx^1 \wedge dx^2.$$

Taking the Hodge star dual we get

$$
\begin{aligned}
*\mathbf{F} = & -E_1 \ *(dt \wedge dx^1) - E_2 \ *(dt \wedge dx^2) - E_3 \ *(dt \wedge dx^3) \\
& + B_1 \ *(dx^2 \wedge dx^3) + B_2 \ *(dx^3 \wedge dx^1) + B_3 \ *(dx^1 \wedge dx^2)
\end{aligned}
$$

$$
\begin{aligned}
= & \ E_1 dx^2 \wedge dx^3 + E_2 dx^3 \wedge dx^1 + E_3 dx^1 \wedge dx^2 \\
& + B_1 dt \wedge dx^1 + B_2 dt \wedge dx^2 + B_3 dt \wedge dx^3.
\end{aligned}
$$

Now we take the exterior derivative of $*\mathbf{F}$ to get

$$
\begin{aligned}
d * \mathbf{F} = & \ dE_1 \wedge dx^2 \wedge dx^3 + dE_2 \wedge dx^3 \wedge dx^1 + dE_3 \wedge dx^1 \wedge dx^2 \\
& + dB_1 \wedge dt \wedge dx^1 + dB_2 \wedge dt \wedge dx^2 + dB_3 \wedge dt \wedge dx^3
\end{aligned}
$$

$$
\begin{aligned}
= & \ \frac{\partial E_1}{\partial t} \, dt \wedge dx^2 \wedge dx^3 + \frac{\partial E_1}{\partial x^1} \, dx^1 \wedge dx^2 \wedge dx^3 \\
& + \frac{\partial E_2}{\partial t} \, dt \wedge dx^2 \wedge dx^1 + \frac{\partial E_2}{\partial x^2} \, dx^2 \wedge dx^3 \wedge dx^1 \\
& + \frac{\partial E_3}{\partial t} \, dt \wedge dx^1 \wedge dx^2 + \frac{\partial E_3}{\partial x^3} \, dx^3 \wedge dx^1 \wedge dx^2 \\
& + \frac{\partial B_1}{\partial x^2} \, dx^2 \wedge dt \wedge dx^1 + \frac{\partial B_1}{\partial x^3} \, dx^3 \wedge dt \wedge dx^1 \\
& + \frac{\partial B_2}{\partial x^1} \, dx^1 \wedge dt \wedge dx^2 + \frac{\partial B_2}{\partial x^3} \, dx^3 \wedge dt \wedge dx^2 \\
& + \frac{\partial B_3}{\partial x^1} \, dx^1 \wedge dt \wedge dx^3 + \frac{\partial B_3}{\partial x^2} \, dx^2 \wedge dt \wedge dx^3
\end{aligned}
$$

$$
\begin{aligned}
= & \left(\frac{\partial E_1}{\partial x^1} + \frac{\partial E_2}{\partial x^2} + \frac{\partial E_3}{\partial x^3} \right) dx^1 \wedge dx^2 \wedge dx^3 \\
& + \left(\frac{\partial E_3}{\partial t} + \frac{\partial B_1}{\partial x^2} - \frac{\partial B_2}{\partial x^1} \right) dt \wedge dx^1 \wedge dx^2 \\
& + \left(\frac{\partial E_1}{\partial t} + \frac{\partial B_2}{\partial x^3} - \frac{\partial B_3}{\partial x^2} \right) dt \wedge dx^2 \wedge dx^3 \\
& + \left(\frac{\partial E_2}{\partial t} + \frac{\partial B_3}{\partial x^1} - \frac{\partial B_1}{\partial x^3} \right) dt \wedge dx^3 \wedge dx^1.
\end{aligned}
$$

Using this and the Hodge star dual we find

$$
\begin{aligned}
*d * \mathbf{F} = & -\left(\frac{\partial E_1}{\partial x^1} + \frac{\partial E_2}{\partial x^2} + \frac{\partial E_3}{\partial x^3} \right) dt \\
& -\left(\frac{\partial E_3}{\partial t} + \frac{\partial B_1}{\partial x^2} - \frac{\partial B_2}{\partial x^1} \right) dx^3 \\
& -\left(\frac{\partial E_1}{\partial t} + \frac{\partial B_2}{\partial x^3} - \frac{\partial B_3}{\partial x^2} \right) dx^1 \\
& -\left(\frac{\partial E_2}{\partial t} + \frac{\partial B_3}{\partial x^1} - \frac{\partial B_1}{\partial x^3} \right) dx^2.
\end{aligned}
$$

By letting $*d*\mathbf{F}=\mathbf{J}$ we have the following system of equations

$$
\begin{aligned}
-\frac{\partial E_1}{\partial x^1}-\frac{\partial E_2}{\partial x^2}-\frac{\partial E_3}{\partial x^3} &= -\rho,\\
-\frac{\partial E_1}{\partial t}-\frac{\partial B_2}{\partial x^3}+\frac{\partial B_3}{\partial x^2} &= J_1,\\
-\frac{\partial E_2}{\partial t}-\frac{\partial B_3}{\partial x^1}+\frac{\partial B_1}{\partial x^3} &= J_2,\\
-\frac{\partial E_3}{\partial t}-\frac{\partial B_1}{\partial x^2}+\frac{\partial B_2}{\partial x^1} &= J_3.
\end{aligned}
$$

It is now straight-forward to show that for Gauss's law of electric fields we have

$$\nabla\cdot\mathbf{E}=\rho \quad\Longleftrightarrow\quad \frac{\partial E_1}{\partial x^1}+\frac{\partial E_2}{\partial x^2}+\frac{\partial E_3}{\partial x^3}=\rho,$$

which is the first equation in the system directly above. Similarly, it is straight-forward to show that for the Ampère-Maxwell law we have

$$-\frac{\partial\mathbf{E}}{\partial t}+\nabla\times\mathbf{B}=\mathbf{J} \quad\Longleftrightarrow\quad \begin{cases} -\frac{\partial E_1}{\partial t}-\frac{\partial B_2}{\partial x^3}+\frac{\partial B_3}{\partial x^2}=J_1,\\ -\frac{\partial E_2}{\partial t}-\frac{\partial B_3}{\partial x^1}+\frac{\partial B_1}{\partial x^3}=J_2,\\ -\frac{\partial E_3}{\partial t}-\frac{\partial B_1}{\partial x^2}+\frac{\partial B_2}{\partial x^1}=J_3, \end{cases}$$

which are the last there equations in the system of equations above. Thus we have found that

$$\boxed{*d*\mathbf{F}=\mathbf{J} \quad\Longleftrightarrow\quad \nabla\cdot\mathbf{E}=\rho \ \text{ and } \ \nabla\times\mathbf{B}=\mathbf{J}+\frac{\partial\mathbf{E}}{\partial t}.}$$

In summary, by using the Faraday two-form and the electric current density one-form we have managed to write Maxwell's system of four equations into a system of the two equations,

$$\boxed{\begin{array}{c}\text{Maxwell's Equations:}\\ \text{version two}\\ \\ d\mathbf{F}=0\\ *d*\mathbf{F}=\mathbf{J}\end{array}}$$

Question 12.10 Show that $*d*\mathbf{F}=\mathbf{J} \Leftrightarrow d*\mathbf{F}=*\mathbf{J}$.

The fact that these calculations turned out so nicely using the Hodge star operator, which was defined in terms of the Minkowski inner product defined Minkowski space of special relativity, $\mathbb{R}^{1,3}$, indicates that Maxwell's equations were already compatible with special relativity. We will now go one step further. Since $d\mathbf{F}=0$ we have that $\mathbf{F}$ is closed. We now ask if there is some one-form $\mathbf{A}$ such that $\mathbf{F}=d\mathbf{A}$. Based on the Poincaré lemma of Chap. 8 we know that there is. This one-form $\mathbf{A}$ is called the electromagnetic potential one-form,

$$\mathbf{A}=A_0dt+A_1dx^1+A_2dx^2+A_3dx^3.$$

By looking at this relationship between the Faraday two-form and the electromagnetic potential one-form we are able to see the relationship between the scalar potential, the electric potential, and the electric and magnetic fields. We compute $d\mathbf{A}$, which is then a two-form, and set that equal to the two-form $\mathbf{F}$.

$$d\mathbf{A} = dA_0 \wedge dt + dA_1 \wedge dx^1 + dA_2 \wedge dx^2 + dA_3 \wedge dx^3$$

$$\begin{aligned}
&= \frac{\partial A_0}{\partial x^1} dx^1 \wedge dt + \frac{\partial A_0}{\partial x^2} dx^2 \wedge dt + \frac{\partial A_0}{\partial x^3} dx^3 \wedge dt \\
&\quad + \frac{\partial A_1}{\partial t} dt \wedge dx^1 + \frac{\partial A_1}{\partial x^2} dx^2 \wedge dx^1 + \frac{\partial A_1}{\partial x^3} dx^3 \wedge dx^1 \\
&\quad + \frac{\partial A_2}{\partial t} dt \wedge dx^2 + \frac{\partial A_2}{\partial x^1} dx^1 \wedge dx^2 + \frac{\partial A_2}{\partial x^3} dx^3 \wedge dx^2 \\
&\quad + \frac{\partial A_3}{\partial t} dt \wedge dx^3 + \frac{\partial A_3}{\partial x^1} dx^1 \wedge dx^3 + \frac{\partial A_3}{\partial x^2} dx^2 \wedge dx^3
\end{aligned}$$

$$\begin{aligned}
&= \left(\frac{\partial A_1}{\partial t} - \frac{\partial A_0}{\partial x^1}\right) dt \wedge dx^1 + \left(\frac{\partial A_2}{\partial t} - \frac{\partial A_0}{\partial x^2}\right) dt \wedge dx^2 + \left(\frac{\partial A_3}{\partial t} - \frac{\partial A_0}{\partial x^3}\right) dt \wedge dx^3 \\
&\quad + \left(\frac{\partial A_2}{\partial x^1} - \frac{\partial A_1}{\partial x^2}\right) dx^1 \wedge dx^2 + \left(\frac{\partial A_3}{\partial x^2} - \frac{\partial A_2}{\partial x^3}\right) dx^2 \wedge dx^3 + \left(\frac{\partial A_1}{\partial x^3} - \frac{\partial A_3}{\partial x^1}\right) dx^3 \wedge dx^1
\end{aligned}$$

$$\begin{aligned}
&= -E_1\, dt \wedge dx^1 - E_2\, dt \wedge dx^2 - E_3\, dt \wedge dx^3 \\
&\quad + B_1\, dx^2 \wedge dx^3 + B_2\, dx^3 \wedge dx^1 + B_3\, dx^1 \wedge dx^2.
\end{aligned}$$

This identity gives us the following system of six equations

$$\begin{aligned}
E_1 &= -\frac{\partial A_1}{\partial t} + \frac{\partial A_0}{\partial x^1}, & B_1 &= \frac{\partial A_3}{\partial x^2} - \frac{\partial A_2}{\partial x^3}, \\
E_2 &= -\frac{\partial A_2}{\partial t} + \frac{\partial A_0}{\partial x^2}, & B_2 &= \frac{\partial A_1}{\partial x^3} - \frac{\partial A_3}{\partial x^1}, \\
E_3 &= -\frac{\partial A_3}{\partial t} + \frac{\partial A_0}{\partial x^3}, & B_3 &= \frac{\partial A_2}{\partial x^1} - \frac{\partial A_1}{\partial x^2}.
\end{aligned}$$

By letting $\mathbf{A} = A_1\hat{i} + A_2\hat{j} + A_3\hat{k}$ it is straight forward to to see that

$$\left.\begin{aligned}
E_1 &= -\tfrac{\partial A_1}{\partial t} + \tfrac{\partial A_0}{\partial x^1} \\
E_2 &= -\tfrac{\partial A_2}{\partial t} + \tfrac{\partial A_0}{\partial x^2} \\
E_3 &= -\tfrac{\partial A_3}{\partial t} + \tfrac{\partial A_0}{\partial x^3}
\end{aligned}\right\} \quad \Longleftrightarrow \quad \mathbf{E} = -\frac{\partial \mathbf{A}}{\partial t} + \nabla A_0$$

and that

$$\left.\begin{aligned}
B_1 &= \tfrac{\partial A_3}{\partial x^2} - \tfrac{\partial A_2}{\partial x^3} \\
B_2 &= \tfrac{\partial A_1}{\partial x^3} - \tfrac{\partial A_3}{\partial x^1} \\
B_3 &= \tfrac{\partial A_2}{\partial x^1} - \tfrac{\partial A_1}{\partial x^2}
\end{aligned}\right\} \quad \Longleftrightarrow \quad \mathbf{B} = \nabla \times \mathbf{A}.$$

Question 12.11 Using the identities obtained from setting $\mathbf{F} = d\mathbf{A}$ show $\mathbf{E} = -\frac{\partial \mathbf{A}}{\partial t} + \nabla A_0$ and $\mathbf{B} = \nabla \times \mathbf{A}$.

In physics $-A_0$ is called the scalar potential or the electric potential and is usually denoted by V. The vector field $\mathbf{A} = A_1\hat{i} + A_2\hat{j} + A_3\hat{k}$ is called the vector potential or the magnetic potential. The electric/scalar potential and the magnetic/vector potential are related to the electric field $\mathbf{E}$ and the magnetic field $\mathbf{B}$ according to

$$\mathbf{E} = -\frac{\partial \mathbf{A}}{\partial t} - \nabla V \quad \text{and} \quad \mathbf{B} = \nabla \times \mathbf{A}.$$

The one-form

Electromagnetic potential one-form	$\mathbf{A} = -V\,dt + A_1\,dx^1 + A_2\,dx^2 + A_3\,dx^3$

is called the electromagnetic potential one-form or sometimes the electromagnetic 4-potential, where the four represents the four dimensions of the Minkowski space-time manifold.

By having $\mathbf{F} = d\mathbf{A}$ we automatically have $d\mathbf{F} = dd\mathbf{A} = 0$ so Maxwell's equations

$$d\mathbf{F} = 0,$$
$$*d*\mathbf{F} = \mathbf{J}$$

can be combined into one equation

Maxwell's Equations:
version three

$$*d*d\mathbf{A} = \mathbf{J}.$$

This one equation contains all of electromagnetism.

As a final comment it is interesting to look at the ways that the Poincaré lemma shows up and can be used. Notice that $\mathbf{A}$ is a one-form such that $\mathbf{F} = d\mathbf{A}$. However, $\mathbf{A}$ is not unique. If $\mathbf{A}' = \mathbf{A} + d\lambda$ for some zero-form (function) λ then we have

$$\mathbf{F} = d\mathbf{A}' = d(\mathbf{A} + d\lambda) = d\mathbf{A} + dd\lambda = d\mathbf{A}.$$

This follows directly from the Poincaré lemma. The mapping $\mathbf{A} \mapsto \mathbf{A}' = \mathbf{A} + d\lambda$ is a version of what is called a gauge transformation in physics. Also, consider the relationship between the magnetic potential $\mathbf{A}$ and the magnetic field $\mathbf{B}$ given by $\mathbf{B} = \nabla \times \mathbf{A}$. Then Gauss's law for magnetic fields, $\nabla \cdot \mathbf{B} = 0$ is simply an example of the Poincaré lemma, $\nabla \cdot (\nabla \times \mathbf{A}) = 0$. Also consider the relationship between the electric potential V and the electric field $\mathbf{E}$ given by $\mathbf{E} = -\dot{\mathbf{A}} - \nabla V$. Taking the curl of this relationship we get $\nabla \times \mathbf{E} = -\nabla \times \dot{\mathbf{A}} - \nabla \times (\nabla V)$. By the Poincaré lemma we have $\nabla \times (\nabla V) = 0$ and so $\nabla \times \mathbf{E} = -\nabla \times \dot{\mathbf{A}} = -\dot{\mathbf{B}}$, which is exactly Faraday's law.

12.5 Summary, References, and Problems

12.5.1 Summary

Electromagnetism is described by four equations known as Maxwell's equations. These equations describe the relationship between electrical fields $\mathbf{E}$, magnetic fields $\mathbf{B}$, electrical current flux vector field $\mathbf{J}$, and the charge density function ρ. They are summarized in vector calculus notation as

Maxwell's Equations:
version one

Integral Form		Differential Form
$\int_S \mathbf{E} \cdot \hat{n}\, dS = q_{\text{enc.}}$	$\Longleftrightarrow$	$\nabla \cdot \mathbf{E} = \rho$
$\int_S \mathbf{B} \cdot \hat{n}\, dS = 0$	$\Longleftrightarrow$	$\nabla \cdot \mathbf{B} = 0$
$\int_C \mathbf{E} \cdot d\mathbf{r} = -\frac{d}{dt} \int_S \mathbf{B} \cdot \hat{n}\, dS$	$\Longleftrightarrow$	$\nabla \times \mathbf{E} = -\frac{\partial \mathbf{B}}{\partial t}$
$\int_C \mathbf{B} \cdot d\mathbf{r} = I_{\text{enc.}} + \frac{d}{dt} \int_S \mathbf{E} \cdot \hat{n}\, dS$	$\Longleftrightarrow$	$\nabla \times \mathbf{B} = \mathbf{J} + \frac{\partial \mathbf{E}}{\partial t}.$

The Faraday two-form can be defined using the components of the electric and magnetic fields as

Faraday two-form
$$\begin{aligned}\mathbf{F} = &-E_1\, dt \wedge dx^1 - E_2\, dt \wedge dx^2 - E_3\, dt \wedge dx^3 \\ &+B_1\, dx^2 \wedge dx^3 + B_2\, dx^3 \wedge dx^1 + B_3\, dx^1 \wedge dx^2,\end{aligned}$$

while the electric current density one-form can be defined using the components of the electrical current flux vector field and the charge density function as

Electric current density one-form
$$\mathbf{J} = -\rho dt + J_1 dx^1 + J_2 dx^2 + J_3 dx^3.$$

Using these two differential forms Maxwell's equations can be written as

Maxwell's Equations:
version two

$$d\mathbf{F} = 0$$
$$*d * \mathbf{F} = \mathbf{J}.$$

The electromagnetic potential one-form is defined as

Electromagnetic potential one-form
$$\mathbf{A} = -V\, dt + A_1\, dx^1 + A_2\, dx^2 + A_3\, dx^3,$$

were $\mathbf{F} = d\mathbf{A}$. This allows us to write Maxwell's equations as

Maxwell's Equations:
version three

$$*d * d\mathbf{A} = \mathbf{J}.$$

Hence using differential forms we have been able to write all of electromagnetism in one simple elegant equation.

12.5.2 References and Further Reading

Electromagnetism is very often the first, and one of the most natural, topics in which one can illustrate the utility and power of differential forms. All of Maxwell's rather complicated equations can be written very succinctly using differential forms notation. Here we have assumed the reader may not be familiar with Maxwell's equations and thus have taken some effort to introduce them and explain their meaning more or less following Fleisch [20] and, to a lessor extent, Stevens [42]. The Hodge dual in the context of special relativity is discussed in both Dray [16] and Abraham, Marsden and Ratiu [1]. Presentations of the differential forms formulation of Electromagnatism can be found in Abraham, Marsden and Ratiu [1], Martin [33], and Frankel [21]. Additionally, the papers by Warnick and Russer [48], Warnick, Selfridge, and Arnold [49], and DesChampes [13] were useful.

After writing this chapter we became aware of a recent book by Garrity [23] that seems to be a wonderful introduction to Electromagnetism from a mathematical point of view. As such it appears covers much of the material found in this chapter in greater depth but also goes well beyond it, while still being very readable. In particular Garrity's exposition of the Hodge star is very similar, though he uses the opposite sign convention.

12.5.3 Problems

Question 12.12 Does the electric flux through the surface of a sphere containing 10 electrons and seven protons depend on the size of the sphere. Find the electric flux through the sphere.

Question 12.13 Does the magnetic flux through the surface of a sphere containing a small magnet depend on the size of the sphere? Find the magnetic flux through the sphere. Suppose the magnet is just outside of the sphere. Find the magnetic flux.

Question 12.14 Use Faraday's law to find the induced current in a square loop with sides of length a in the xy-plane in a region where the magnetic field changes over time as $\mathbf{B}(t) = B_0 e^{-5t/t_0}\hat{i}$.

Question 12.15 Two parallel wires carry currents I and $2I$ in opposite directions. Use the Ampère-Maxwell law to find the magnetic field at a point midway between the two wires.

Question 12.16 On the manifold $\mathbb{R}^6$ equipped with the Euclidian metric, find the Hodge star of the following forms:

$a)\ dx_1$	$e)\ dx_1 \wedge dx_2 \wedge dx_3$	$i)\ dx_1 \wedge dx_2 \wedge dx_4 \wedge dx_6$
$b)\ dx_2$	$f)\ dx_2 \wedge dx_4 \wedge dx_5$	$j)\ dx_1 \wedge dx_3 \wedge dx_5 \wedge dx_6$
$c)\ dx_5$	$g)\ dx_2 \wedge dx_5 \wedge dx_6$	$k)\ dx_2 \wedge dx_3 \wedge dx_4 \wedge dx_5$
$d)\ dx_6$	$h)\ dx_3 \wedge dx_4 \wedge dx_5$	$l)\ dx_1 \wedge dx_2 \wedge dx3 \wedge dx_4 \wedge dx_6$

Question 12.17 Suppose we know an inner product on $\bigwedge^1(\mathbb{R}^n)$ with ordered basis $dx_1, \ldots, dx_n$. A basis on $\bigwedge^k(\mathbb{R}^n)$ is formed from the $dx_I = dx_{i_1} \wedge \cdots \wedge dx_{i_k}$. We can define an inner product on $\bigwedge^k(\mathbb{R}^n)$ by setting, for $I = (i_1, \ldots, i_k)$ and $J = (j_1, \ldots, j_k)$,

$$\langle dx_I, dx_J \rangle = \sum_{\sigma \in S_k} sgn(\sigma) \langle dx_{i_1}, dx_{j_{\sigma(1)}} \rangle \langle dx_{i_2}, dx_{j_{\sigma(2)}} \rangle \cdots \langle dx_{i_k}, dx_{j_{\sigma(k)}} \rangle,$$

where S_k is the group of permutation on k elements. Given the inner product on $\bigwedge^1\left(\mathbb{R}^{1,3}\right)$ show that this was how the inner products on $\bigwedge^3\left(\mathbb{R}^{1,3}\right)$, $\bigwedge^3\left(\mathbb{R}^{1,3}\right)$, and $\bigwedge^4\left(\mathbb{R}^{1,3}\right)$ were defined.

Question 12.18 Define an inner product on $\bigwedge^1\left(\mathbb{R}^3\right)$ by setting $\langle dx, dx \rangle = 3$, $\langle dy, dy \rangle = 5$, $\langle dz, dz \rangle = 2$, $\langle dx, dy \rangle = 1$, $\langle dx, dz \rangle = 0$, and $\langle dy, dz \rangle = 2$. Write the inner product as a 3×3 symmetric matrix using the ordering dx, dy, dz for the vector space basis.

Question 12.19 With the inner product on $\bigwedge^1\left(\mathbb{R}^3\right)$ in Question 12.18 and the definition given in problem 12.17 find the corresponding inner products for the spaces $\bigwedge^2\left(\mathbb{R}^3\right)$ and $\bigwedge^3\left(\mathbb{R}^3\right)$. Using the ordering $dx \wedge dy, dy \wedge dz, dz \wedge dx$, write the inner product on $\bigwedge^2\left(\mathbb{R}^3\right)$ as a 3×3 matrix.

Question 12.20 Using the inner products for the preceding two problems find the Hodge star duals of all the basis elements of $\bigwedge^2\left(\mathbb{R}^3\right)$, $\bigwedge^2\left(\mathbb{R}^3\right)$, and $\bigwedge^3\left(\mathbb{R}^3\right)$.

Question 12.21 Define an inner product on $\bigwedge^1(\mathbb{R}^3)$ by setting $\langle dx, dx\rangle = 2$, $\langle dy, dy\rangle = 4$, $\langle dz, dz\rangle = 6$, $\langle dx, dy\rangle = 1$, $\langle dx, dz\rangle = 2$, and $\langle dy, dz\rangle = 3$. Write the inner product as a 3×3 symmetric matrix using the ordering dx, dy, dz for the vector space basis.

Question 12.22 With the inner product on $\bigwedge^1(\mathbb{R}^3)$ in Question 12.21 and the definition given in problem 12.17 find the corresponding inner products for the spaces $\bigwedge^2(\mathbb{R}^3)$ and $\bigwedge^3(\mathbb{R}^3)$. Using the ordering $dx \wedge dy$, $dy \wedge dz$, $dz \wedge dx$, write the inner product on $\bigwedge^2(\mathbb{R}^3)$ as a 3×3 matrix.

Question 12.23 Using the inner products for the preceding two problems find the Hodge star duals of all the basis elements of $\bigwedge^2(\mathbb{R}^3)$, $\bigwedge^2(\mathbb{R}^3)$, and $\bigwedge^3(\mathbb{R}^3)$.

Appendix A
Introduction to Tensors

This appendix focuses on tensors. There is a deep enough relationship between differential forms and tensors that covering this material is important. However, this material is also distinct enough from differential forms that relegating it to an appendix is appropriate. Indeed, If we were to cover the material on tensors in the same spirit and with the same diligence as we covered differential forms this book would be twice the current size. Therefore this appendix is more of a quick overview of the essentials, though hopefully with enough detail that the relationship between differential forms and tensors is clear.

But there is also another reason for including this appendix, our strong desire to include a full proof of the global formula for exterior differentiation given at the end of Sect. A.7. Strangely enough, the full global exterior derivative formula for differential forms uses, in its proof, the global Lie derivative formula for differential forms. The material presented in this appendix is necessary for understanding that proof.

In section one we give a brief big-picture overview of tensors and why they play such an important role in mathematics and physics. After that we define tensors using the strategy of multilinear maps and then use this, along with what we already know, to introduce the coordinate transformation formulas. Section two looks at rank-one tensors, section three at rank-two tensors, and section four looks at general tensors. Section five introduces differential forms as a special kind of tensor and then section six gives the real definition of a metric tensor. We have already discussed metrics in an imprecise manor several times in this book. Finally, section seven introduces the Lie derivative, one of the fundamental ways of taking derivatives on manifolds, and then goes on to prove many important identities and formulas that are related to Lie derivatives and their relationship to exterior derivatives. With this material we are finally able to give a rigorous proof of the global formula for exterior differentiation.

A.1 An Overview of Tensors

This book, as the title clearly indicates, is a book about differential forms and calculus on manifolds. In particular, it is a book meant to help you develop a good geometrical understanding of, and intuition for, differential forms. Thus, we have chosen one particular pedagogical path among several possible pedagogical paths, what we believe is the most geometrically intuitive pedagogical path, to introduce the concepts of differential forms and basic ideas regarding calculus on manifolds. Now that we have developed some understanding of differential forms we will use the concept of differential forms to introduce tensors. It is, however, perfectly possible, and in fact very often done, to introduce tensors first and then introduce differential forms as a special case of tensors. In particular, differential forms are skew-symmetric covariant tensors. The word anti-symmetric is often used instead of skew-symmetric. Using this definition, formulas for the wedgeproduct, exterior derivatives, and all the rest can be developed.

The problem with this approach is that students are often overwhelmed by the flurry of indices and absolutely mystified by what it all actually means geometrically. The topic quickly becomes simply a confusing mess of difficult to understand formulas and equations. Much of this book has been an attempt to provide an easier to understand, and more geometrically based, introduction to differential forms. In this appendix we will continue in this spirit as much as is reasonably possible. In this chapter we will define tensors and introduce the tensorial "index" notation that you will likely see in future physics or differential geometry classes. Our hope is that when you see tensors again you will be better positioned to understand them.

Broadly speaking, there are three generally different strategies that are used to introduce tensors. They are as follows:

J. P. Fortney, *A Visual Introduction to Differential Forms and Calculus on Manifolds*,
https://doi.org/10.1007/978-3-319-96992-3

1. Transformation rules are given. Here, the definition of tensors essentially boils down to the rather trite phrase that "tensors are objects that transform like tensors." This strategy is certainly the one that is most often used in undergraduate physics classes. Its advantage is that it allows students to hit the ground running with computations. Its disadvantage is that there is generally little or no deeper understanding of what tensors are and it often leaves students feeling a bit mystified or confused.
2. As multi-linear maps. Here, tensors are introduced as a certain kind of multilinear map. This strategy is often used in more theoretical physics classes or in undergraduate math classes. This is also the strategy that we will employ in this chapter. In particular, we will make a concerted effort to link this approach to the transformation rules approach mentioned above. This way we can gain both a more abstract understanding of what tensors are as well as some facility with computations.
3. An abstract mathematical approach. This approach is generally used in theoretical mathematics and is quite beyond the scope of this book, though it is hinted at. A great deal of heavy mathematical machinery has too be well understood for this approach to introducing tensors.

It is natural to ask why we should care about tensors and what they are good for. What is so important about them? The fundamental idea is that the laws of physics should not (and in fact do not) depend on the coordinate system that is used to write down the mathematical equations that describe those laws. For example, the gravitational interaction of two particles simply is what it is and does not depend on if we use cartesian coordinates or spherical coordinates or cylindrical coordinates or some other system of coordinates to describe the gravitational interaction of these two particles. What physically happens to these two particles out there in the real world is totally independent of the coordinate system we use to describe what happens. It is also independent of where we decide to place the origin of our coordinate system or the orientation of that coordinate system. This is absolutely fundamental to our idea of how the universe behaves.

The equations we write down for a particular coordinate system maybe be complicated or simple, but once we solve them the actual behavior those equations describe is always the same. So, we are interested in finding ways to write down the equations for physical laws that

1. allow us to do computations,
2. enable us to switch between coordinate systems easily,
3. and are guaranteed to be independent of the coordinate system used. This means the actual results we get do not depend on, that is, are independent of, the coordinate system we used in doing the calculations.

Actually, this is quite a tall order. Mathematicians like to write things down in coordinate-invariant ways, that is, using notations and ideas that are totally independent of any coordinate system that may be used. In general this makes proving things much easier but makes doing specific computations, like computing the time evolution of two particular gravitationally interacting particles, impossible. In order to do any sort of specific computations related to a particular situation requires one to define a coordinate system first. It turns out that tensors are exactly what is needed. While it is true that they can be quite complicated, there is a lot of power packed into them. They ensure that all three of the above requirements are met.

It is, in fact, this ability to switch, or transform, between coordinate systems easily that gives the physics "definition" of tensors as "objects that transform like tensors." A tensor, when transformed to another coordinate system, is still a tensor. Relationships that hold between tensors in one coordinate system automatically hold in all other coordinate systems because the transformations work the same way on all tensors. This is not meant to be an exhaustive or rigorous introduction to tensors. We are interested in understanding the big picture and the relationship between tensors and differential forms. We will employ the Einstein summation notation throughout, which means we must be very careful about upper and lower indices. There is a reason Elie Cartan, who did a great deal of work in mathematical physics and differential geometry, said that the tensor calculus is the "debauch of indices."

A.2 Rank One Tensors

A tensor on a manifold is a multilinear map

$$
\begin{aligned}
T : \underbrace{T^*M \times \cdots \times T^*M}_{\substack{r \\ \text{contravariant} \\ \text{degree}}} \times \underbrace{TM \times \cdots \times TM}_{\substack{s \\ \text{covariant} \\ \text{degree}}} &\longrightarrow \mathbb{R} \\
T\big(\alpha_1, \ldots, \alpha_r, v_1, \ldots, v_s\big) \qquad\qquad &\longrightarrow \mathbb{R}.
\end{aligned}
$$

That is, the map T eats r one-forms and s vectors and produces a real number. We will discuss the multilinearity part later so will leave it for now. The number r, which is the number of one-forms that the tensor T eats, is called the **contravariant** degree and the number s, which is the number of vectors the tensor T eats, is called the **covariant** degree. We will also discuss this terminology later. The tensor T would be called a rank-r contravariant, rank-s covariant tensor, or a rank (r, s)-tensor for short. Before delving deeper into general tensors let us take a more detailed look at two specific tensors to see how they work and to start get a handle on the notation.

Rank (0, 1)-Tensors (Rank-One Covariant Tensors)
A (0, 1)-tensor, or a rank-one covariant tensor, is a linear mapping

$$T : TM \longrightarrow \mathbb{R}.$$

But of course we recognize exactly what these are, they are exactly one-forms. Thus a rank one covariant tensor is another way of saying a differential one-form, which we are very familiar with. Suppose that M is an n dimensional manifold with coordinate functions $(x^1, x^2, \ldots, x^n)$. When we wish to make it clear which coordinates we are using on M we will write $M_{(x^1,x^2,\ldots,x^n)}$. We will now use superscripts for our manifold's coordinate functions instead of the subscripts more commonly encountered in math classes so all of our Einstein summation notation works out. Thus we have

$$T \in T^*M_{(x^1,x^2,\ldots,x^n)} = \operatorname{span}\Big\{dx^1, dx^2, \ldots, dx^n\Big\}$$

which can therefore be written as

$$\begin{aligned} T &= T_1 dx^1 + T_2 dx^2 + \cdots + T_n dx^n \\ &= \sum_{i=1}^{n} T_i dx^i \\ &\equiv T_i dx^i, \end{aligned}$$

where the T_i are either numbers or real-valued functions on M. Also take a moment to recall how Einstein summation notation works, we sum over matching lower and upper indices. Since T_i has a lower index i, which matches the upper index i in dx^i we sum over i. The numbers, or real-valued functions, $T_1, T_2, \ldots, T_n$ are called the components of the tensor T. Notice that we are not keeping track of the base point in our notation. The notation will end up complicated enough and writing the base point can seem like overkill.

Now suppose we change coordinate systems on the manifold M. The manifold itself does not change, but the coordinates that we use to name particular points p on the manifold changes. We will consider changes of coordinates to be a mapping of the manifold M to itself. Suppose we change the coordinates from $(x^1, x^2, \ldots, x^n)$ to $(u^1, u^2, \ldots, u^n)$ using the n functions

$$\begin{aligned} u^1(x^1, x^2, \ldots, x^n) &= u_1, \\ u^2(x^1, x^2, \ldots, x^n) &= u_2, \\ &\vdots \\ u^n(x^1, x^2, \ldots, x^n) &= u_n. \end{aligned}$$

We depict this coordinate change as a mapping

$$M_{(x^1,x^2,\ldots,x^n)} \longrightarrow M_{(u^1,u^2,\ldots,u^n)}.$$

What we want to do is write the tensor T in terms of these new coordinates $u^1, u^2, \ldots, u^n$. In other words, we want to find new components $\widetilde{T}_1, \widetilde{T}_2, \ldots, \widetilde{T}_n$ such that

$$\begin{aligned}
T &= \widetilde{T}_1 du^1 + \widetilde{T}_2 du^2 + \cdots + \widetilde{T}_n du^n \\
&= \sum_{i=1}^{n} \widetilde{T}_i du^i \\
&\equiv \widetilde{T}_i du^i .
\end{aligned}$$

Notice from our notation, only the components of tensor T change, the tensor T itself does not change at all. We will assume that our change of coordinates is invertible, which means we may have to restrict ourselves to some subset, or coordinate patch, on the manifold M. With an invertible change of coordinates we can also write

$$\begin{aligned}
x^1(u^1, u^2, \ldots, u^n) &= x_1, \\
x^2(u^1, u^2, \ldots, u^n) &= x_2, \\
&\vdots \\
x^n(u^1, u^2, \ldots, u^n) &= x_n.
\end{aligned}$$

Now we will figure out how to write $\widetilde{T}_i$ in terms of T_i. For the coordinate functions x^i, taking the exterior derivatives we get the following system of identities relating the one-forms dx^i with the one-forms du^j,

$$\begin{aligned}
dx^1 &= \frac{\partial x^1}{\partial u^1} du^1 + \frac{\partial x^1}{\partial u^2} du^2 + \cdots + \frac{\partial x^1}{\partial u^n} du^n, \\
dx^2 &= \frac{\partial x^2}{\partial u^1} du^1 + \frac{\partial x^2}{\partial u^2} du^2 + \cdots + \frac{\partial x^2}{\partial u^n} du^n, \\
&\vdots \\
dx^n &= \frac{\partial x^n}{\partial u^1} du^1 + \frac{\partial x^n}{\partial u^2} du^2 + \cdots + \frac{\partial x^n}{\partial u^n} du^n.
\end{aligned}$$

Using these identities we can rewrite the tensor T,

$$\begin{aligned}
T =&\ T_1 dx^1 + T_2 dx^2 + \cdots + T_n dx^n \\
=&\ T_1 \Big(\frac{\partial x^1}{\partial u^1} du^1 + \frac{\partial x^1}{\partial u^2} du^2 + \cdots + \frac{\partial x^1}{\partial u^n} du^n \Big) \\
&+ T_2 \Big(\frac{\partial x^2}{\partial u^1} du^1 + \frac{\partial x^2}{\partial u^2} du^2 + \cdots + \frac{\partial x^2}{\partial u^n} du^n \Big) \\
&+ \cdots \\
&+ T_n \Big(\frac{\partial x^n}{\partial u^1} du^1 + \frac{\partial x^n}{\partial u^2} du^2 + \cdots + \frac{\partial x^n}{\partial u^n} du^n \Big) \\
=&\ \Big(T_1 \frac{\partial x^1}{\partial u^1} + T_2 \frac{\partial x^2}{\partial u^1} + \cdots + T_n \frac{\partial x^n}{\partial u^1} \Big) du^1 \\
&+ \Big(T_1 \frac{\partial x^1}{\partial u^2} + T_2 \frac{\partial x^2}{\partial u^2} + \cdots + T_n \frac{\partial x^n}{\partial u^2} \Big) du^2
\end{aligned}$$

$$
\begin{aligned}
&+\cdots \\
&+\left(T_1\frac{\partial x^1}{\partial u^n}+T_2\frac{\partial x^2}{\partial u^n}+\cdots+T_n\frac{\partial x^n}{\partial u^n}\right)du^n \\
=\ & \widetilde{T}_1 du^1+\widetilde{T}_2 du^2+\cdots+\widetilde{T}_n du^n.
\end{aligned}
$$

Thus we have the system of equations

$$
\begin{aligned}
\widetilde{T}_1 &= T_1\frac{\partial x^1}{\partial u^1}+T_2\frac{\partial x^2}{\partial u^1}+\cdots+T_n\frac{\partial x^n}{\partial u^1}, \\
\widetilde{T}_2 &= T_1\frac{\partial x^1}{\partial u^2}+T_2\frac{\partial x^2}{\partial u^2}+\cdots+T_n\frac{\partial x^n}{\partial u^2}, \\
&\vdots \\
\widetilde{T}_n &= T_1\frac{\partial x^1}{\partial u^n}+T_2\frac{\partial x^2}{\partial u^n}+\cdots+T_n\frac{\partial x^n}{\partial u^n},
\end{aligned}
$$

which can be rewritten as the matrix equation

$$
\begin{bmatrix} \widetilde{T}_1 \\ \widetilde{T}_2 \\ \vdots \\ \widetilde{T}_n \end{bmatrix}
=
\begin{bmatrix}
\frac{\partial x^1}{\partial u^1} & \frac{\partial x^2}{\partial u^1} & \cdots & \frac{\partial x^n}{\partial u^1} \\
\frac{\partial x^1}{\partial u^2} & \frac{\partial x^2}{\partial u^2} & \cdots & \frac{\partial x^n}{\partial u^2} \\
\vdots & \vdots & \ddots & \vdots \\
\frac{\partial x^1}{\partial u^n} & \frac{\partial x^2}{\partial u^n} & \cdots & \frac{\partial x^n}{\partial u^n}
\end{bmatrix}
\begin{bmatrix} T_1 \\ T_2 \\ \vdots \\ T_n \end{bmatrix}.
$$

Using Einstein summation notation we can write T's tensor components $\widetilde{T}_i$ in terms of T's tensor components T_j using the transformation rule

$$
\boxed{\text{Covariant Tensor Transformation Rule:} \qquad \widetilde{T}_i = \frac{\partial x^j}{\partial u^i} T_j.}
$$

Notice that the i in $\frac{\partial x^j}{\partial u^i}$ is regarded as a lower index of the whole term even though it is an upper index of the u, while the j is considered as an upper index of the term; the indices above the horizontal bar are considered upper and indices below the horizontal bar are considered lower.

This equation is what is given as the definition of a covariant tensor when tensors are introduced via transformation rules. That is, T is defined to be a covariant tensor if its components transform under changes of coordinates according to the rule $\widetilde{T}_i = \frac{\partial x^j}{\partial u^i} T_j$. Had you been given this as a definition for a covariant tensor would you have immediately recognized that is was just a one-form?

Turning our attention to the tangent spaces we know that

$$
\begin{aligned}
TM_{(x^1,x^2,\ldots,x^n)} &= \operatorname{span}\left\{\frac{\partial}{\partial x^1},\frac{\partial}{\partial x^2},\ldots,\frac{\partial}{\partial x^n}\right\}, \\
TM_{(u^1,u^2,\ldots,u^n)} &= \operatorname{span}\left\{\frac{\partial}{\partial u^1},\frac{\partial}{\partial u^2},\ldots,\frac{\partial}{\partial u^n}\right\}.
\end{aligned}
$$

We would like to see how we can write the $\frac{\partial}{\partial u^i}$ in terms of $\frac{\partial}{\partial x^j}$. Suppose we have a function $f : M \to \mathbb{R}$,

$$
f\left(x^1,x^2,\ldots,x^n\right) = f\Big(x^1\left(u^1,u^2,\ldots,u^n\right), x^2\left(u^1,u^2,\ldots,u^n\right),\ldots,x^n\left(u^1,u^2,\ldots,u^n\right)\Big).
$$

Using the chain rule gives us

$$\begin{aligned}
\frac{\partial f}{\partial u^1} &= \frac{\partial f}{\partial x^1}\frac{\partial x^1}{\partial u^1} + \frac{\partial f}{\partial x^2}\frac{\partial x^2}{\partial u^1} + \cdots + \frac{\partial f}{\partial x^n}\frac{\partial x^n}{\partial u^1} \\
&\Rightarrow \left(\frac{\partial}{\partial u^1}\right) f = \left(\frac{\partial x^1}{\partial u^1}\frac{\partial}{\partial x^1} + \frac{\partial x^2}{\partial u^1}\frac{\partial}{\partial x^2} + \cdots + \frac{\partial x^n}{\partial u^1}\frac{\partial}{\partial x^n}\right) f, \\
&\vdots \\
\frac{\partial f}{\partial u^n} &= \frac{\partial f}{\partial x^1}\frac{\partial x^1}{\partial u^n} + \frac{\partial f}{\partial x^2}\frac{\partial x^2}{\partial u^n} + \cdots + \frac{\partial f}{\partial x^n}\frac{\partial x^n}{\partial u^n} \\
&\Rightarrow \left(\frac{\partial}{\partial u^n}\right) f = \left(\frac{\partial x^1}{\partial u^n}\frac{\partial}{\partial x^1} + \frac{\partial x^2}{\partial u^n}\frac{\partial}{\partial x^2} + \cdots + \frac{\partial x^n}{\partial u^n}\frac{\partial}{\partial x^n}\right) f.
\end{aligned}$$

Leaving off the function f this system of identities can be rewritten as the matrix equation

$$\begin{bmatrix} \frac{\partial}{\partial u^1} \\ \frac{\partial}{\partial u^2} \\ \vdots \\ \frac{\partial}{\partial u^n} \end{bmatrix} = \begin{bmatrix} \frac{\partial x^1}{\partial u^1} & \frac{\partial x^2}{\partial u^1} & \cdots & \frac{\partial x^n}{\partial u^1} \\ \frac{\partial x^1}{\partial u^2} & \frac{\partial x^2}{\partial u^2} & \cdots & \frac{\partial x^n}{\partial u^2} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial x^1}{\partial u^n} & \frac{\partial x^2}{\partial u^n} & \cdots & \frac{\partial x^n}{\partial u^n} \end{bmatrix} \begin{bmatrix} \frac{\partial}{\partial x^1} \\ \frac{\partial}{\partial x^2} \\ \vdots \\ \frac{\partial}{\partial x^n} \end{bmatrix}.$$

We can write this system of equations in Einstein summation notation as

$$\boxed{\text{Vector Basis Transformation Rule:} \qquad \frac{\partial}{\partial u^i} = \frac{\partial x^j}{\partial u^i}\frac{\partial}{\partial x^j}.}$$

This is how basis vectors of TM transform. Compare this to what we had before with respect to covariant tensor components,

$$\widetilde{T}_i = \frac{\partial x^j}{\partial u^i} T_j.$$

Covariant tensors eat vectors and covariant tensor components transform like vectors basis elements.

When you think about it this is an odd and interesting point. Make sure you understand it. The components of a covariant tensor, a one-form, transform ***not*** like one-form basis elements do but instead as vector basis elements do.

Rank (1, 0)-Tensors (Rank-One Contravariant Tensors)
Now let us consider a $(1, 0)$-tensor, or a rank-one contravariant tensor. It is a linear mapping

$$T : T^*M \to \mathbb{R}.$$

What thing, when paired with a one-form, gives a number? A vector of course. The pairing between a vector v and a one-form α can be written in several different ways

$$\alpha\big(v\big) = \langle \alpha, v \rangle = v\big(\alpha\big).$$

Sometimes not a lot of attention is paid to the order in the angle brackets, so you may also occasionally see $\langle v, \alpha \rangle$ as well. But regardless of notation, the end point of this discussion is that a $(1, 0)$-tensor is really nothing more then a vector field. We will make the same assumptions as before, that M is an n-dimensional manifold and that we have a change of coordinates given by the mapping

$$M_{(x^1,x^2,\ldots,x^n)} \longrightarrow M_{(u^1,u^2,\ldots,u^n)}.$$

We also assume that this change of coordinates mapping is invertible. As before, the contravariant tensor, that is the vector field, T will be given by

$$\begin{aligned} T &= T^1 \frac{\partial}{\partial x^1} + T^2 \frac{\partial}{\partial x^2} + \cdots + T^n \frac{\partial}{\partial x^n} \\ &= \sum_{i=1}^{n} T^i \frac{\partial}{\partial x^i} \\ &\equiv T^i \frac{\partial}{\partial x^i}. \end{aligned}$$

The i in $\frac{\partial}{\partial x^i}$ is considered a lower index which means that the components of the contravariant tensor are indicated by upper indices, T^i, when using Einstein summation notation.

Like before we would like to write the tensor T in terms of the new coordinates, that is, we would like to find $\widetilde{T}^i$ such that

$$\begin{aligned} T &= \widetilde{T}^1 \frac{\partial}{\partial u^1} + \widetilde{T}^2 \frac{\partial}{\partial u^2} + \cdots + \widetilde{T}^n \frac{\partial}{\partial u^n} \\ &= \sum_{i=1}^{n} \widetilde{T}^i \frac{\partial}{\partial u^i} \\ &= \widetilde{T}^i \frac{\partial}{\partial u^i}. \end{aligned}$$

We want to know how the components of T transform under the coordinate change, so we want to know how to write the $\widetilde{T}^i$ in terms of the T^i.

Suppose we have a function $g : M \to \mathbb{R}$ which we can write as

$$g\big(u^1, u^2, \ldots, u^n\big) = g\Big(u^1\big(x^1, x^2, \ldots, x^n\big), u^2\big(x^1, x^2, \ldots, x^n\big), \ldots, u^n\big(x^1, x^2, \ldots, x^n\big)\Big).$$

Using the chain rule we find

$$\begin{aligned} \frac{\partial g}{\partial x^1} &= \frac{\partial g}{\partial u^1}\frac{\partial u^1}{\partial x^1} + \frac{\partial g}{\partial u^2}\frac{\partial u^2}{\partial x^1} + \cdots + \frac{\partial g}{\partial u^n}\frac{\partial u^n}{\partial x^1} \\ &\Rightarrow \left(\frac{\partial}{\partial x^1}\right) g = \left(\frac{\partial u^1}{\partial x^1}\frac{\partial}{\partial u^1} + \frac{\partial u^2}{\partial x^1}\frac{\partial}{\partial u^2} + \cdots + \frac{\partial u^n}{\partial x^1}\frac{\partial}{\partial u^n}\right) g, \\ &\vdots \\ \frac{\partial g}{\partial x^n} &= \frac{\partial g}{\partial u^1}\frac{\partial u^1}{\partial x^n} + \frac{\partial g}{\partial u^2}\frac{\partial u^2}{\partial x^n} + \cdots + \frac{\partial g}{\partial u^n}\frac{\partial u^n}{\partial x^n} \\ &\Rightarrow \left(\frac{\partial}{\partial x^n}\right) g = \left(\frac{\partial u^1}{\partial x^n}\frac{\partial}{\partial u^1} + \frac{\partial u^2}{\partial x^n}\frac{\partial}{\partial u^2} + \cdots + \frac{\partial u^n}{\partial x^n}\frac{\partial}{\partial u^n}\right) g. \end{aligned}$$

Leaving off the function g we can use these identities to rewrite the tensor T as

$$\begin{aligned}
T = & \; T^1 \frac{\partial}{\partial x^1} + T^2 \frac{\partial}{\partial x^2} + \cdots + T^n \frac{\partial}{\partial x^n} \\
= & \; T^1 \left(\frac{\partial u^1}{\partial x^1} \frac{\partial}{\partial u^1} + \frac{\partial u^2}{\partial x^1} \frac{\partial}{\partial u^2} + \cdots + \frac{\partial u^n}{\partial x^1} \frac{\partial}{\partial u^n} \right) \\
& + T^2 \left(\frac{\partial u^1}{\partial x^2} \frac{\partial}{\partial u^1} + \frac{\partial u^2}{\partial x^2} \frac{\partial}{\partial u^2} + \cdots + \frac{\partial u^n}{\partial x^2} \frac{\partial}{\partial u^n} \right) \\
& \vdots \\
& + T^n \left(\frac{\partial u^1}{\partial x^n} \frac{\partial}{\partial u^1} + \frac{\partial u^2}{\partial x^n} \frac{\partial}{\partial u^2} + \cdots + \frac{\partial u^n}{\partial x^n} \frac{\partial}{\partial u^n} \right) \\
= & \left(T^1 \frac{\partial u^1}{\partial x^1} + T^2 \frac{\partial u^1}{\partial x^2} + \cdots + T^n \frac{\partial u^1}{\partial x^n} \right) \frac{\partial}{\partial u^1} \\
& + \left(T^1 \frac{\partial u^2}{\partial x^1} + T^2 \frac{\partial u^2}{\partial x^2} + \cdots + T^n \frac{\partial u^2}{\partial x^n} \right) \frac{\partial}{\partial u^2} \\
& \vdots \\
& + \left(T^1 \frac{\partial u^n}{\partial x^1} + T^2 \frac{\partial u^n}{\partial x^2} + \cdots + T^n \frac{\partial u^n}{\partial x^n} \right) \frac{\partial}{\partial u^n} \\
= & \; \widetilde{T}^1 \frac{\partial}{\partial u^1} + \widetilde{T}^2 \frac{\partial}{\partial u^2} + \cdots + \widetilde{T}^n \frac{\partial}{\partial u^n}.
\end{aligned}$$

Thus we have the system of equations

$$\begin{aligned}
\widetilde{T}^1 &= T^1 \frac{\partial u^1}{\partial x^1} + T^2 \frac{\partial u^1}{\partial x^2} + \cdots + T^n \frac{\partial u^1}{\partial x^n} \\
\widetilde{T}^2 &= T^1 \frac{\partial u^2}{\partial x^1} + T^2 \frac{\partial u^2}{\partial x^2} + \cdots + T^n \frac{\partial u^2}{\partial x^n} \\
& \vdots \\
\widetilde{T}^n &= T^1 \frac{\partial u^n}{\partial x^1} + T^2 \frac{\partial u^n}{\partial x^2} + \cdots + T^n \frac{\partial u^n}{\partial x^n},
\end{aligned}$$

which can be rewritten as the matrix equation

$$\begin{bmatrix} \widetilde{T}^1 \\ \widetilde{T}^2 \\ \vdots \\ \widetilde{T}^n \end{bmatrix} = \begin{bmatrix} \frac{\partial u^1}{\partial x^1} & \frac{\partial u^1}{\partial x^2} & \cdots & \frac{\partial u^1}{\partial x^n} \\ \frac{\partial u^2}{\partial x^1} & \frac{\partial u^2}{\partial x^2} & \cdots & \frac{\partial u^2}{\partial x^n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial u^n}{\partial x^1} & \frac{\partial u^n}{\partial x^2} & \cdots & \frac{\partial u^n}{\partial x^n} \end{bmatrix} \begin{bmatrix} T^1 \\ T^2 \\ \vdots \\ T^n \end{bmatrix}.$$

Using Einstein summation notation this matrix equation can be written as

$$\boxed{\text{Contravariant Tensor Transformation Rule:} \qquad \widetilde{T}^i = \frac{\partial u^i}{\partial x^j} T^j.}$$

This equation shows how contravariant tensor components transform under a change of basis. Again, often contravariant tensors are simply defined as objects that transform according to this equation. Had you been introduced to contravariant tensors this way would it have been obvious to you that a contravariant tensor was really just a vector field?

Finally, we would like to see how basis elements of T^*M transform with this change of basis. Using exterior differentiation on the coordinate functions u^i we have that

$$\begin{aligned}
du^1 &= \frac{\partial u^1}{\partial x^1}dx^1 + \frac{\partial u^1}{\partial x^2}dx^2 + \cdots + \frac{\partial u^1}{\partial x^n}dx^n \\
du^2 &= \frac{\partial u^2}{\partial x^1}dx^1 + \frac{\partial u^2}{\partial x^2}dx^2 + \cdots + \frac{\partial u^2}{\partial x^n}dx^n \\
&\vdots \\
du^n &= \frac{\partial u^n}{\partial x^1}dx^1 + \frac{\partial u^n}{\partial x^2}dx^2 + \cdots + \frac{\partial u^n}{\partial x^n}dx^n,
\end{aligned}$$

which of course give the matrix equation

$$\begin{bmatrix} du^1 \\ du^2 \\ \vdots \\ du^n \end{bmatrix} = \begin{bmatrix} \frac{\partial u^1}{\partial x^1} & \frac{\partial u^1}{\partial x^2} & \cdots & \frac{\partial u^1}{\partial x^n} \\ \frac{\partial u^2}{\partial x^1} & \frac{\partial u^2}{\partial x^2} & \cdots & \frac{\partial u^2}{\partial x^n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial u^n}{\partial x^1} & \frac{\partial u^n}{\partial x^2} & \cdots & \frac{\partial u^n}{\partial x^n} \end{bmatrix} \begin{bmatrix} dx^1 \\ dx^2 \\ \vdots \\ dx^n \end{bmatrix}.$$

In Einstein summation notation this becomes

$$\boxed{\text{One-Form Basis Transformation Rule:} \qquad du^i = \frac{\partial u^i}{\partial x^j}dx^j.}$$

So this is the way that basis elements of T^*M transform. Compare this with how components of contravariant tensors transform

$$\widetilde{T}^i = \frac{\partial u^i}{\partial x^j}T^j.$$

Contravariant tensors eat one-forms and contravariant tensor components transform like one-form basis elements.

Let us compare this transformation rule with that of covariant tensors.

Covariant Tensors	Contravariant Tensors
$\widetilde{T}_i = \frac{\partial x^j}{\partial u^i}T_j$	$\widetilde{T}^i = \frac{\partial u^i}{\partial x^j}T^j$
Vector Basis Elements	One-Form Basis Elements
$\frac{\partial}{\partial u^i} = \frac{\partial x^j}{\partial u^i}\frac{\partial}{\partial x^j}$	$du^i = \frac{\partial u^i}{\partial x^j}dx^j$

A.3 Rank-Two Tensors

Before getting into general tensors we will take a closer look at rank-two tensors. There are three possibilities for rank-two tensors:

1. $(0, 2)$-Tensors (Rank-Two Covariant Tensor),
2. $(2, 0)$-Tensors (Rank-Two Contravariant Tensor),
3. $(1, 1)$-Tensors (Mixed-Rank Covariant-Contravariant Tensor).

Rank $(0, 2)$-Tensors (Rank-Two Covariant Tensors)

Once we get a firm grasp on how these three tensors work moving to the general case is little more than adding indices. First we will consider the $(0, 2)$-tensor, which is a rank two covariant tensor. It is a multilinear map

$$T : TM \times TM \longrightarrow \mathbb{R}.$$

So a $(0, 2)$-tensor takes as input two vectors and gives as output a number. We already know one thing that does exactly this, a two-form. After our success with $(0, 1)$-tensors being one-forms we might be tempted to guess that $(0, 2)$-tensors are two-forms. But if we did that we would be wrong. Two-forms are a subset of $(0, 2)$-tensors but not the whole set. There are many $(0, 2)$-tensors that are not two-forms. In fact, the most important of these is called the metric tensor, which we will talk about in a later section. Similarly, three-forms are a subset of $(0, 3)$-tensors, four-forms are a subset of $(0, 4)$-tensors, etc. Had you learned about tensors first instead of differential forms then differential forms would have been defined to be skew-symmetric covariant tensors.

Here we are getting very close to the abstract mathematical approach to defining tensors. Since that is beyond the scope of this book and we want to avoid that, we will not be mathematically rigorous, which will thereby necessitate some degree of vagueness in the following discussion. However, we hope that you will at least get a general idea of the issues involved.

We begin by defining the Cartesian product of two vector spaces V and W to be

$$V \times W = \Big\{ (v, w) \Big| v \in V \text{ and } w \in W \Big\}.$$

So $V \times W$ is the set of pairs of elements, the first of which is from V and the second of which is from W. If $(v_1, \ldots, v_n)$ is a basis of V and $(w_1, \ldots, w_m)$ is a basis of W then we can write

$$V \times W = \text{span}\Big\{ (v_i, 0), (0, w_j) \Big| 1 \le i \le n, 1 \le j \le m \Big\}.$$

Let us consider a $(0, 2)$-tensor T which is a mapping $T : TM \times TM \longrightarrow \mathbb{R}$. We might be tempted to say that T is an element of the vector space

$$T^*M \times T^*M = \text{span}\Big\{ (dx^i, 0), (0, dx^j) \Big| 1 \le i, j \le n \Big\}.$$

Unfortunately, this actually does not make much sense. Now is the time to remember that tensors are multilinear mappings. This means that the mapping T is linear in each "slot,"

$$T\Big(1^{\text{st}} \text{ slot }, 2^{\text{st}} \text{ slot }\Big).$$

In other words, suppose that $v, v_1, v_2, w, w_1, w_2 \in TM$ are vectors and $a, b \in \mathbb{R}$ are scalars. Then we have that

1. $T\big(av_1 + bv_2, w\big) = aT\big(v_1, w\big) + bT\big(v_2, w\big)$,
2. $T\big(v, aw_1 + bw_2\big) = aT\big(v, w_1\big) + bT\big(v, w_2\big)$.

We want to somehow get these multilinear mappings from the set $T^*M \times T^*M = \text{span}\big\{(dx^i, 0), (0, dx^j)\big\}$. We will completely hand-wave through this bit, but what is done is that something called a free vector space is constructed from this Cartesian product space and then the elements of this space are put into equivalence classes in a procedure similar to what was done when we discussed equivalence classes of curves on manifolds. A great deal of mathematical machinery is employed in this precise but abstract mathematical approach to defining tensors. If you did not follow the last couple paragraphs, don't worry, none of it is necessary for what follows.

The space made up of these equivalence classes is the tensor space $T^*M \otimes T^*M$, which can be thought of as the set of multilinear elements of the form $dx^i \otimes dx^j$, called the tensor product of dx^i and dx^j. Thus we have

$$T \in T^*M \otimes T^*M = \text{span}\left\{ dx^i \otimes dx^j \;\middle|\; 1 \le i, j \le n \right\}.$$

How do the elements of the form $dx^i \otimes dx^j$ work? If v and w are two vector fields on M then

$$dx^i \otimes dx^j(v, w) \equiv dx^i(v) dx^i(w).$$

So, what would $T \in T^*M \otimes T^*M$ actually look like?

For the moment let us suppose that M is a two dimensional manifold. Then we would have

$$T \in T^*M \otimes T^*M = \text{span}\left\{dx^1 \otimes dx^1,\ dx^1 \otimes dx^2,\ dx^2 \otimes dx^1,\ dx^2 \otimes dx^2\right\}$$

so that

$$T = T_{11} dx^1 \otimes dx^1 + T_{12} dx^1 \otimes dx^2 + T_{21} dx^2 \otimes dx^1 + T_{22} dx^2 \otimes dx^2.$$

Pay close attention to the way that the subscript indices work. In general we would write a $(0, 2)$-tensor, using Einstein summation notation, as

$$T = T_{ij} dx^i \otimes dx^j.$$

We could, if we wanted, write the tensor T as a matrix.

Like before suppose we had the change of basis mappings $M_{(x_1,x_2,\dots,x_n)} \longrightarrow M_{(u_1,u_2,\dots,u_n)}$ given by the n invertible functions $u_1(x_1, \dots, x_n), \dots, u_n(x_1, \dots, x_n)$. We know how to write

$$T = T_{ij} dx^i \otimes dx^j \in T^*M_{(x_1,x_2,\dots,x_n)} \otimes T^*M_{(x_1,x_2,\dots,x_n)}$$

and we want to write T as

$$T = \widetilde{T_{ij}} du^k \otimes du^l \in T^*M_{(u_1,u_2,\dots,u_n)} \otimes T^*M_{(u_1,u_2,\dots,u_n)}.$$

That is, we want to see how the components of T transform. Since T is a rank two covariant tensor, each index transforms covariantly and we have

$$\boxed{\text{Covariant Rank 2 Transformation Rule:} \qquad \widetilde{T_{kl}} = \frac{\partial x^i}{\partial u^k} \frac{\partial x^j}{\partial u^l} T_{ij}}$$

Finally, we point out again that the two-forms are the skew-symmetric elements of $T^*M \otimes T^*M$ they are a subset of the $(0, 2)$-tensors,

$$\underbrace{\bigwedge^2(M)}_{\text{two-forms}} \subsetneqq \underbrace{T^*M \otimes T^*M}_{(0,2)-\text{tensors}}.$$

What do we mean by that? Well, as we know, for a two-form $dx^i \wedge dx^i$ and vectors v and w, we have that

$$dx^i \wedge dx^i(v, w) = -dx^i \wedge dx^i(w, v).$$

The sign of our answer changes if we switch the order of our input vectors. This need not happen for a general tensor. For a general tensor there is no reason why we should expect the value $dx^i \otimes dx^j(v, w)$ to be related to the value $dx^i \otimes dx^j(w, v)$. They may not be related at all, much less related by a simple sign change. The tensors for which this happens, that is, the tensors for which

$$dx^i \otimes dx^j(v, w) = -dx^i \otimes dx^j(w, v)$$

are said to be skew-symmetric, or anti-symmetric. Tensors that have this property are a very special and unusual subset of the set of all tensors. The covariant tensors that have this property are so special that they have their own name, differential forms.

Rank (2, 0)-Tensors (Rank-Two Contravariant Tensors)
Now we consider the (2, 0)-tensor, which is a rank-two contravariant tensor. This is a multilinear map

$$T : T^*M \times T^*M \longrightarrow \mathbb{R}.$$

So a (2, 0)-tensor takes as input two one-forms and gives as output a number. Without repeating the whole discussion from above, something analogous happens. The (2, 0)-tensor is an element of the vector space

$$TM \otimes TM = \text{span}\left\{ \frac{\partial}{\partial x^i} \otimes \frac{\partial}{\partial x^j} \;\Big|\; 1 \leq i, j \leq n \right\},$$

which consists of the multilinear elements of the vector product space $TM \times TM$. Consider a two-dimensional manifold M. Then we have

$$T \in TM \otimes TM = \text{span}\left\{ \frac{\partial}{\partial x^1} \otimes \frac{\partial}{\partial x^1},\ \frac{\partial}{\partial x^1} \otimes \frac{\partial}{\partial x^2},\ \frac{\partial}{\partial x^2} \otimes \frac{\partial}{\partial x^1},\ \frac{\partial}{\partial x^2} \otimes \frac{\partial}{\partial x^2} \right\},$$

so that

$$T = T^{11}\frac{\partial}{\partial x^1} \otimes \frac{\partial}{\partial x^1} + T^{12}\frac{\partial}{\partial x^1} \otimes \frac{\partial}{\partial x^2} + T^{21}\frac{\partial}{\partial x^2} \otimes \frac{\partial}{\partial x^1} + T^{22}\frac{\partial}{\partial x^2} \otimes \frac{\partial}{\partial x^2}.$$

In general we would write a (2, 0)-tensor, in Einstein summation notation, as

$$T = T^{ij}\frac{\partial}{\partial x^i} \otimes \frac{\partial}{\partial x^j}.$$

Like before suppose we have an invertible change of basis $M_{(x_1,x_2,\ldots,x_n)} \longrightarrow M_{(u_1,u_2,\ldots,u_n)}$ and we know how to write

$$T = T^{ij}\frac{\partial}{\partial x^i} \otimes \frac{\partial}{\partial x^j} \in TM_{(x_1,x_2,\ldots,x_n)} \otimes TM_{(x_1,x_2,\ldots,x_n)}$$

and we want to write T as

$$T = \widetilde{T^{kl}}\frac{\partial}{\partial u^k} \otimes \frac{\partial}{\partial u^l} \in TM_{(u_1,u_2,\ldots,u_n)} \otimes TM_{(u_1,u_2,\ldots,u_n)}.$$

That is, we want to see how the components of T transform. Since T is a rank two contravariant tensor, each index transforms contravariantly and we have

$$\boxed{\text{Contravariant Rank 2 Transformation Rule:} \qquad \widetilde{T^{kl}} = \frac{\partial u^k}{\partial x^i}\frac{\partial u^l}{\partial x^j}T^{ij}.}$$

Rank (1, 1)-Tensors (Mixed Rank Tensors)
Now we will look at a (1, 1)-tensor, or a mixed-rank covariant-contravariant tensor. This tensor is a multilinear map

$$T : T^*M \times TM \longrightarrow \mathbb{R}.$$

A (1, 1)-tensor takes as input one vector and one one-form and gives as output a number. Again, the (1, 1)-tensor is an element of the vector space

$$TM \otimes T^*M = \text{span}\left\{ \frac{\partial}{\partial x^i} \otimes dx^j \;\Big|\; 1 \leq i, j \leq n \right\},$$

which consists of the multilinear elements of the vector product space $TM \times T^*M$. Considering a two dimensional manifold M we have that

$$T \in TM \otimes T^*M = \text{span}\left\{ \frac{\partial}{\partial x^1} \otimes dx^1,\; \frac{\partial}{\partial x^1} \otimes dx^2,\; \frac{\partial}{\partial x^2} \otimes dx^1,\; \frac{\partial}{\partial x^2} \otimes dx^2 \right\},$$

so that

$$T = T_1^1 \frac{\partial}{\partial x^1} \otimes dx^1 + T_2^1 \frac{\partial}{\partial x^1} \otimes dx^2 + T_1^2 \frac{\partial}{\partial x^2} \otimes dx^1 + T_2^2 \frac{\partial}{\partial x^2} \otimes dx^2.$$

In general we would write T, using Einstein summation notation, as

$$T = T_j^i \frac{\partial}{\partial x^i} \otimes dx^j.$$

Like before suppose we have an invertible change of basis $M_{(x_1,x_2,\ldots,x_n)} \longrightarrow M_{(u_1,u_2,\ldots,u_n)}$ and we know how to write

$$T = T_j^i \frac{\partial}{\partial x^i} \otimes dx^j \in TM_{(x_1,x_2,\ldots,x_n)} \otimes T^*M_{(x_1,x_2,\ldots,x_n)}$$

and we want to write T as

$$T = \widetilde{T_l^k} \frac{\partial}{\partial u^k} \otimes du^l \in TM_{(u_1,u_2,\ldots,u_n)} \otimes T^*M_{(u_1,u_2,\ldots,u_n)}.$$

That is, we want to see how the components of T transform. The upper index transforms contravariantly and the lower index transforms covariantly. Thus we have

$$\boxed{\text{Rank (1, 1)-Tensor Transformation Rule:} \qquad \widetilde{T_l^k} = \frac{\partial u^k}{\partial x^i} \frac{\partial x^j}{\partial u^l} T_j^i.}$$

A.4 General Tensors

Now we will finally take a look at a general (r, s)-tensor; it is a multilinear map

$$\mathscr{T} : \underbrace{T^*M \times \cdots \times T^*M}_{\substack{r \\ \text{contravariant} \\ \text{degree}}} \times \underbrace{TM \times \cdots \times TM}_{\substack{s \\ \text{covariant} \\ \text{degree}}} \longrightarrow \mathbb{R}.$$

This means that the general (r, s)-tensor is an element of the space

$$\begin{aligned} \mathscr{T} &\in \underbrace{TM \otimes \cdots \otimes TM}_{r} \otimes \underbrace{T^*M \otimes \cdots \otimes T^*M}_{s} \\ &= \text{span}\left\{ \frac{\partial}{\partial x^{i_1}} \otimes \cdots \otimes \frac{\partial}{\partial x^{i_r}} \otimes dx^{j_1} \otimes \cdots \otimes dx^{j_s} \Big| 1 \leq i_1, \ldots, i_r, j_1, \ldots, j_s \leq n \right\}. \end{aligned}$$

A general (r, s)-tensor is written

$$\begin{aligned}\mathscr{T} &= \mathscr{T}^{i_1\cdots i_r}_{j_1\cdots j_s}\frac{\partial}{\partial x^{i_1}}\otimes\cdots\otimes\frac{\partial}{\partial x^{i_r}}\otimes dx^{j_1}\otimes\cdots\otimes dx^{j_s}\\ &= \mathscr{T}^{i_1\cdots i_r}_{j_1\cdots j_s}\partial x^{i_1}\otimes\cdots\otimes\partial x^{i_r}\otimes dx^{j_1}\otimes\cdots\otimes dx^{j_s}.\end{aligned}$$

Given the same invertible change of basis $M_{(x_1,x_2,\ldots,x_n)} \longrightarrow M_{(u_1,u_2,\ldots,u_n)}$ we know how to write

$$\mathscr{T} = \widetilde{\mathscr{T}}^{k_1\cdots k_r}_{l_1\cdots l_s}\frac{\partial}{\partial u^{k_1}}\otimes\cdots\otimes\frac{\partial}{\partial u^{k_r}}\otimes du^{l_1}\otimes\cdots\otimes du^{l_s}.$$

Tensor $\mathscr{T}$'s components transform according to

Rank (r, w)-Tensor Transformation Rule: $\quad \widetilde{\mathscr{T}}^{k_1\cdots k_r}_{l_1\cdots l_s} = \dfrac{\partial u^{k_1}}{\partial x^{i_1}}\cdots\dfrac{\partial u^{k_r}}{\partial x^{i_r}}\dfrac{\partial x^{j_1}}{\partial u^{l_1}}\cdots\dfrac{\partial x^{j_s}}{\partial u^{l_s}}\mathscr{T}^{i_1\cdots i_r}_{j_1\cdots j_s}.$

Suppose that we have a mapping $\phi : M \to M$. The pullback of a rank $(0, t)$-tensor $\mathscr{T}$ at the point p is defined exactly as we defined the pullback of differential forms, by

$$\big(\phi^*\mathscr{T}_{\phi(p)}\big)_p(v_{1_p},\ldots,v_{t_p}) = \mathscr{T}_{\phi(p)}(\phi_* v_{1_p},\ldots,\phi_* v_{t_p}).$$

Now suppose we had two covariant tensors, a rank $(0, t)$-tensor $\mathscr{T}$ and a rank $(0, s)$-tensor $\mathscr{S}$. Then we have

$$\begin{aligned}&\Big(\phi^*\big(\mathscr{T}\otimes\mathscr{S}\big)_{\phi(p)}\Big)_p(v_{1_p},\ldots,v_{(t+s)_p})\\ &\quad= \big(\mathscr{T}\otimes\mathscr{S}\big)_{\phi(p)}(\phi_* v_{1_p},\ldots,\phi_* v_{(t+s)_p})\\ &\quad= \mathscr{T}_{\phi(p)}(\phi_* v_{1_p},\ldots,\phi_* v_{t_p})\mathscr{S}_{\phi(p)}(\phi_* v_{(t+1)_p},\ldots,\phi_* v_{(t+s)_p})\\ &\quad= \phi^*\mathscr{T}_{\phi(p)}(v_{1_p},\ldots,v_{t_p})\phi^*\mathscr{S}_{\phi(p)}(v_{(t+1)_p},\ldots,v_{(t+s)_p})\\ &\quad= \big(\phi^*\mathscr{T}_{\phi(p)}\otimes\phi^*\mathscr{S}_{\phi(p)}\big)(v_{1_p},\ldots,v_{(t+s)_p}).\end{aligned}$$

Thus we have shown for covariant tensors that

$$\phi^*\big(\mathscr{T}\otimes\mathscr{S}\big) = \phi^*\mathscr{T}\otimes\phi^*\mathscr{S}.$$

Let us now consider contravariant tensors. Even though the building blocks of contravariant tensors are vector fields, and we generally think of pushing-forward vector fields, we will still think in terms of pulling back contravariant tensors. The reason we do this is because we will want to pull-back mixed rank tensor fields as well, and in order to define the pullback of a mixed rank tensor field we have to know how to pullback a contravariant tensor field. It would be entirely possible to alter the definitions so that we always think of pushing forward tensor fields, but the general convention is to think of pulling back tensor fields.

We define the pullback of a rank $(t, 0)$-tensor by

$$\big(\phi^*\mathscr{T}_{\phi(p)}\big)_p(\alpha_{1_p},\ldots,\alpha_{t_p}) = \mathscr{T}_{\phi(p)}\big((\phi^{-1})^*\alpha_{1_p},\ldots,(\phi^{-1})^*\alpha_{t_p}\big).$$

Now suppose we had two contravariant tensors, a rank $(t, 0)$-tensor $\mathscr{T}$ and a rank $(s, 0)$-tensor $\mathscr{S}$. Then, using an identical line of reasoning, we have

$$\begin{aligned}&\Big(\phi^*\big(\mathscr{T}\otimes\mathscr{S}\big)_{\phi(p)}\Big)_p(\alpha_{1_p},\ldots,\alpha_{(t+s)_p})\\ &\quad= \big(\mathscr{T}\otimes\mathscr{S}\big)_{\phi(p)}\big((\phi^{-1})^*\alpha_{1_p},\ldots,(\phi^{-1})^*\alpha_{(t+s)_p}\big)\\ &\quad= \mathscr{T}_{\phi(p)}\big((\phi^{-1})^*\alpha_{1_p},\ldots,(\phi^{-1})^*\alpha_{t_p}\big)\mathscr{S}_{\phi(p)}\big((\phi^{-1})^*\alpha_{(t+1)_p},\ldots,(\phi^{-1})^*\alpha_{(t+s)_p}\big)\end{aligned}$$

$$= \phi^* \mathscr{T}_{\phi(p)}(\alpha_{1_p}, \ldots, \alpha_{t_p})\phi^* \mathscr{S}_{\phi(p)}(\alpha_{(t+1)_p}, \ldots, \alpha_{(t+s)_p})$$
$$= \big(\phi^* \mathscr{T}_{\phi(p)} \otimes \phi^* \mathscr{S}_{\phi(p)}\big)(\alpha_{1_p}, \ldots, \alpha_{(t+s)_p}).$$

Thus we have also shown for contravariant tensors that

$$\phi^*\big(\mathscr{T} \otimes \mathscr{S}\big) = \phi^*\mathscr{T} \otimes \phi^*\mathscr{S}.$$

Now we consider a mixed rank (s, t)-tensor $\mathscr{T}$. We define the pullback of a (s, t)-tensor as follows

$$\big(\phi^* \mathscr{T}_{\phi(p)}\big)_p(\alpha_{1_p}, \ldots, \alpha_{s_p}, v_{1_p}, \ldots, v_{t_p})$$
$$= \mathscr{T}_{\phi(p)}\big((\phi^{-1})^*\alpha_{1_p}, \ldots, (\phi^{-1})^*\alpha_{t_p}, \phi_* v_{1_p}, \ldots, \phi_* v_{t_p}\big).$$

Question A.1 Given a rank (q, r)-tensor $\mathscr{T}$ and a rank (s, t)-tensor $\mathscr{S}$, show that

$$\phi^*\big(\mathscr{T} \otimes \mathscr{S}\big) = \phi^*\mathscr{T} \otimes \phi^*\mathscr{S}.$$

Putting this all together, for any tensors $\mathscr{T}$ and $\mathscr{S}$ we have

$$\boxed{\phi^*\big(\mathscr{T} \otimes \mathscr{S}\big) = \phi^*\mathscr{T} \otimes \phi^*\mathscr{S}.}$$

A.5 Differential Forms as Skew-Symmetric Tensors

Quite often you will find books introduce tensors first and then introduce and define differential forms in terms of tensors. This is of course not the approach we have taken in this book, but we would like to devote a section to the definition of k-forms in terms of tensors for completeness sake. Also, some of the formulas look quite different from this perspective and it is important that you at least be exposed to these so that when you do see them in the future you will have some idea of where they came from.

Suppose we have a totally covariant tensor $\mathscr{T}$ of rank k, that is, a $(0, k)$-tensor. Thus, $\mathscr{T} \in \underbrace{T^*M \otimes \cdots \otimes T^*M}_{k}$. Since $\mathscr{T}$ is a tensor it is multilinear. But differential forms are skew-symmetric in addition to being multilinear. The tensor $\mathscr{T}$ is called anti-symmetric or skew-symmetric if it changes sign whenever any pair of its arguments are switched. That is,

$$\mathscr{T}\big(v_1, \ldots, v_i, \ldots, v_j, \ldots, v_k\big) = -\mathscr{T}\big(v_1, \ldots, v_j, \ldots, v_i, \ldots, v_k\big)$$

when v_i and v_j have switched their places and all other inputs have stated the same. Sometimes skew-symmetry is defined in terms of permutations π of $(1, 2, 3, \ldots, k)$. Thus, T is skew-symmetric if

$$\mathscr{T}\big(v_{\pi(1)}, v_{\pi(2)}, \ldots, v_{\pi(k)}\big) = \text{sgn}(\pi)\mathscr{T}\big(v_1, v_2, \ldots, v_k\big)$$

$$\boxed{\text{A } k\text{-form is a skew-symmetric rank } k \text{ covariant tensor.}}$$

Thus we have that the set of k-forms is a subset of the set of $(0, k)$-tensors,

$$\underbrace{\bigwedge^s(M)}_{k-\text{forms}} \subsetneq \underbrace{T^*M \otimes \cdots \otimes T^*M}_{(0,k)-\text{tensors}}.$$

Question A.2 Show that for a $(0, 2)$-tensor this implies that $\mathscr{T}_{ij} = -\mathscr{T}_{ji}$.

Question A.3 For $\mathscr{T}$ a $(0, 3)$-tensor show that the two definitions of skew-symmetry are equivalent. Then show the two definitions are equivalent for a $(0, 4)$-tensor.

In Sect. 3.3.3 on the general formulas for the wedgeproduct, we have the following formula for the wedgeproduct of a k-form α and an ℓ-form β,

$$(\alpha \wedge \beta)\big(v_1, \ldots, v_{k+\ell}\big) = \frac{1}{k!\ell!} \sum_{\sigma \in S_{k+\ell}} \operatorname{sgn}(\sigma)\alpha\big(v_{\sigma(1)}, \ldots, v_{\sigma(k)}\big)\beta\big(v_{\sigma(k+1)}, \ldots, v_{\sigma(k+\ell)}\big).$$

In that section we then proceeded to show that this definition was equivalent to the determinant based formula that finds volumes of a parallelepiped. We also said that an alternative definition of the wedgeproduct, given in terms of tensors, was

$$\alpha \wedge \beta = \frac{(k+\ell)!}{k!\ell!} \mathscr{A}(\alpha \otimes \beta),$$

where $\mathscr{A}$ was the skew-symmetrization operator. We said that we would explain that formula in the appendix on tensors.

We begin by defining the skew-symmetrization operator. The skew-symmetrization operator takes a p-tensor $\mathscr{T}$ and turns it into a skew-symmetric tensor $\mathscr{A}(\mathscr{T})$ according to the following formula,

$$\mathscr{A}(\mathscr{T})(v_1, \ldots, v_p) = \frac{1}{p!} \sum_{\pi \in S_p} \operatorname{sgn}(\pi)\ \mathscr{T}(v_{\pi(1)}, \ldots, v_{\pi(p)}),$$

where the sum is over all the permutations π of p elements. We divide by $p!$ because there are $p!$ different permutations of p elements, that is, $|S_p| = p!$, and hence there are $p!$ terms in the sum.

Question A.4 Show that $\mathscr{A}(\mathscr{T})$ is skew-symmetric when $\mathscr{T} = dx^1 \otimes dx^2$; when $\mathscr{T} = dx^1 \otimes dx^3 + dx^2 \otimes dx^3$; when $\mathscr{T} = dx^1 \otimes dx^2 \otimes dx^3$.

We now show that the definition of the wedgeproduct in terms of the skew-symmetrization operator is equivalent to the other definition from the previous chapter,

$$\begin{aligned}
\alpha \wedge \beta(v_1, \ldots, v_{k+\ell}) &= \frac{(k+\ell)!}{k!\ell!} \mathscr{A}(\alpha \otimes \beta)(v_1, \ldots, v_{k+\ell}) \\
&= \frac{(k+\ell)!}{k!\ell!} \frac{1}{(k+\ell)!} \sum_{\pi \in S_{k+\ell}} \operatorname{sgn}(\pi)\ (\alpha \otimes \beta)(v_{\pi(1)}, \ldots, v_{\pi(k+l)}) \\
&= \frac{1}{k!\ell!} \sum_{\pi \in S_{k+\ell}} \operatorname{sgn}(\pi)\ \alpha(v_{\pi(1)}, \ldots, v_{\pi(k)})\beta(v_{\pi(k+1)}, \ldots, v_{\pi(k+l)}).
\end{aligned}$$

Thus we see that this definition of the wedgeproduct is exactly equivalent to the other definitions we have used throughout this book. It should be noted that the various properties of the wedgeproduct can also be proved using this definition, though we will not attempt to do so here. In books that introduce tensors first, the above formula is often given as the definition of the wedgeproduct.

We finally turn to showing that the pullback of the wedgeproduct of two forms is the wedgeproduct of the pullbacks of the two forms. Another way of saying this is that pullbacks distribute across wedgeproducts. The proof of this fact proceeds quite nicely and cleanly when we use the tensor definition of the wedgeproduct. Dropping the base point from the notation, pullback of the sum of two covariant $(0, t)$-tensors is written as

$$\begin{aligned}
\phi^*\big(\mathscr{T} + \mathscr{S}\big)(v_1, \ldots, v_t) &= \big(\mathscr{T} + \mathscr{S}\big)(\phi_* v_1, \ldots, \phi_* v_t) \\
&= \mathscr{T}(\phi_* v_1, \ldots, \phi_* v_t) + \mathscr{S}(\phi_* v_1, \ldots, \phi_* v_t) \\
&= \phi^*\mathscr{T}(v_1, \ldots, v_t) + \phi^*\mathscr{S}(v_1, \ldots, v_t).
\end{aligned}$$

Hence we have $\phi^*(\mathscr{T} + \mathscr{S}) = \phi^*\mathscr{T} + \phi^*\mathscr{S}$, which means that pullbacks distribute over addition.

$$\begin{aligned}
\mathscr{A}(\phi^*\mathscr{T})(v_1, \ldots, v_p) &= \frac{1}{p!} \sum_{\pi \in S_p} \operatorname{sgn}(\pi)\ \phi^*\mathscr{T}(v_{\pi(1)}, \ldots, v_{\pi(p)}) \\
&= \frac{1}{p!}\phi^* \left(\sum_{\pi \in S_p} \operatorname{sgn}(\pi)\ \mathscr{T}(v_{\pi(1)}, \ldots, v_{\pi(p)}) \right) \\
&= \phi^* \left(\frac{1}{p!} \sum_{\pi \in S_p} \operatorname{sgn}(\pi)\ \phi^*\mathscr{T}(v_{\pi(1)}, \ldots, v_{\pi(p)}) \right) \\
&= \phi^*\mathscr{A}(\mathscr{T})(v_1, \ldots, v_p)
\end{aligned}$$

and so we have $\mathscr{A}\phi^* = \phi^*\mathscr{A}$.

Now suppose we have two skew-symmetric covariant rank $(0, k)$-tensors, in other words, two k-forms, α and β. Then using the identity $\phi^*(\mathscr{T} \otimes \mathscr{S}) = \phi^*\mathscr{T} \otimes \phi^*\mathscr{S}$ we derived at the end of the last section we have

$$\begin{aligned}
\phi^*(\alpha \wedge \beta) &= \phi^* \left(\frac{(k+\ell)!}{k!\ell!} \mathscr{A}(\alpha \otimes \beta) \right) \\
&= \frac{(k+\ell)!}{k!\ell!} \phi^*\mathscr{A}(\alpha \otimes \beta) \\
&= \frac{(k+\ell)!}{k!\ell!} \mathscr{A}\phi^*(\alpha \otimes \beta) \\
&= \frac{(k+\ell)!}{k!\ell!} \mathscr{A}(\phi^*\alpha \otimes \phi^*\beta) \\
&= \phi^*\alpha \wedge \phi^*\beta.
\end{aligned}$$

Hence we have the important identity

$$\boxed{\phi^*(\alpha \wedge \beta) = \phi^*\alpha \wedge \phi^*\beta.}$$

So we have shown that pullbacks distribute across wedgeproducts.

A.6 The Metric Tensor

In this section we finally get around to defining and introducing metrics on manifolds, something that we promised to do in the introduction to the chapter on manifolds. We have also alluded to metrics several times as well, particularly in the chapter on electromagnetism when we discussed special relativity. There we gave a very informal introduction to both the Euclidian metric and the Minkowski metric as matrices that were used in finding the inner product of two vectors. The concept of a metric is extremely important and very fundamental. It underlies many of the operations and concepts that you have worked with throughout both this book and your calculus classes. But there is a good possibility that before this book you have not even seen the word.

There are a couple good and practical reasons why you probably have never been formally introduced to metrics before. The first is that most of high school math and geometry, as well as calculus, is built on our comfortable familiar Euclidian spaces, and Euclidian space, particularly when it comes equipped with our nice comfortable familiar Cartesian coordinates, has the Euclidian metric seamlessly built in. The Euclidian metric is built into the formulas that you have always used for the distance between two points, or the length or norm of a vector, or the inner product between two vectors, or even into our ideas of what the shortest distance between two points is or our idea of what a straight line is. All of these things rely, implicitly, on the Euclidian metric.

Another reason is that as there are only so many new ideas and concepts that a human brain can assimilate at once, and when you are trying to learn calculus in $\mathbb{R}^n$ it is best that you concentrate on learning calculus. Side tracks into the theoretical underpinnings as to why things work would only muddy up the calculus concepts that you are trying to learn. And of course, unless you are going to go on to become a mathematician or a theoretical physicists, there is actually little need to go off and deal with other sorts of spaces or manifolds.

And finally, as you probably know, for a long time people thought Euclidian geometry was the only kind of geometry there was. Even after they were discovered, it took many decades for people to become comfortable that there were other kinds of "non-Euclidian" geometry out there, in particular hyperbolic geometry and elliptic geometry. And then later on differential geometry, and in particular Riemannian geometry were introduced. It actually took some time for mathematician to get all the underlying theoretical concepts, like metrics, sorted out. Thus the idea of metrics is several centuries more recent than the major ideas of calculus.

Now onto the definition of a metric. A ***metric*** *on the manifold M is a smooth, symmetric, non-degenerate, rank-two covariant tensor g*. Metric tensors are generally denoted with a lower-case g. We explain each of these terms in turn. Suppose we had two smooth vector fields v and w. Recall that in each coordinate neighborhood of M we have a coordinate system $(x_1, \ldots, x_n)$, which allows us to write the vectors as $v = v^i(x_1, \ldots, x_n)\partial_{x^i}$ and $w = w^i(x_1, \ldots, x_n)\partial_{x^j}$, where Einstein summation notation is used. The vector field v is called smooth if each of the functions $v^i(x_1, \ldots, x_n)$ is differentiable an infinite number of times with respect to the arguments $x_1, \ldots, x_n$.

The $(0, 2)$-tensor g is called **smooth** if for the smooth vector fields v and w then the real-valued function $g(v, w)$ is also differentiable an infinite number of times. Since g is a $(0, 2)$-tensor we can write $g = g_{ij}(x_1, \ldots, x_n)dx^i \otimes dx^i$, so another way to say that g is smooth is to say that the component functions $g_{ij}(x_1, \ldots, x_n)$ are infinitely differentiable in the arguments. We have already discussed skew-symmetry, so symmetry should be clear. The tensor g is symmetric if $g(v, w) = g(w, v)$ for all smooth vector fields v and w. Non-degeneracy is a little more complicated, and is discussed in a little more detail when the symplectic form is introduced in Sect. B.3, but we will state its definition here. Suppose at some point $p \in M$ that $g_p(v_p, w_p) = 0$ for any possible choice of vector v_p. If g is non-degenerate at p then that means the only way that this could happen is if the vector $w_p = 0$. g is called non-degenerate if it is non-degenerate at every point $p \in M$.

A manifold that has such a tensor g on it is called a **pseudo-Riemannian manifold** and the tensor g is called the metric or sometimes the **pseudo-Riemannian metric**. If the metric g also has one additional property, that $g(v, w) \geq 0$ for all vector fields v and w then it is called a **Riemannian metric** and the manifold is called a **Riemannian manifold**.

Question A.5 Is the Minkowski metric from Sect. 12.3 a Riemannian metric? Why or why not?

The metric tensor g gives an inner product on every vector space T_pM in the tangent bundle of M. The inner product of $v_p, w_p \in T_pM$ is given by $g(v_p, w_p)$. Most often the inner product of two vectors is denoted with $\langle \cdot, \cdot \rangle$ where

$$\langle v_p, w_p \rangle \equiv g(v_p, w_p).$$

For basis vectors we clearly have

$$\begin{aligned}\left\langle \frac{\partial}{\partial x^i}, \frac{\partial}{\partial x^j} \right\rangle &\equiv g\left(\frac{\partial}{\partial x^i}, \frac{\partial}{\partial x^j}\right) \\ &= g_{k\ell}dx^k \otimes dx^\ell \left(\frac{\partial}{\partial x^i}, \frac{\partial}{\partial x^j}\right) \\ &= g_{ij}.\end{aligned}$$

When $g_{ij} = \delta_{ij}$ then g is called the Euclidian metric, the associated inner product is called the Euclidian inner product, which is none other than the dot product that you are familiar with.

Since g is a rank-two tensor we can write g as a matrix. We start by noticing

$$\begin{aligned}g(v, w) &= g(v^i\partial_{x^i}, w^j\partial_{x^j}) \\ &= v^i w^j g(\partial_{x^i}, \partial_{x^j}) \\ &= v^i w^j g_{ij}.\end{aligned}$$

Another way of writing this would be as

$$
\begin{aligned}
g(v, w) &= [v^1, v^2, \cdots, v^n] \begin{bmatrix} g_{11} & g_{12} & \cdots & g_{1n} \\ g_{21} & g_{22} & \cdots & g_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ g_{n1} & g_{n2} & \cdots & g_{nn} \end{bmatrix} \begin{bmatrix} w^1 \\ w^2 \\ \vdots \\ w^n \end{bmatrix} \\
&= v^T [g_{ij}] w.
\end{aligned}
$$

Often you see metrics introduced or given simply as a matrix, as we did earlier when discussing the Euclidian and Minkowski metrics in the chapter on electromagnetism. This particularly happens when one wishes to avoid any discussion of tensors. In the case of the Euclidian metric, if $g_{ij} = \delta_{ij}$ then the matrix $[g_{ij}]$ is none other than the identity matrix, which is exactly the matrix which we had given for the Euclidian metric in the chapter on electromagnetism in Sect. 12.3. But this kind of representation of the metric implicitly depends on a particular coordinate system being used. In the case where the Euclidian metric was represented as the identity matrix the Cartesian coordinate system was being used. To find the matrix representation of a metric in another coordinate system one would have to perform a coordinate transformation of the metric tensor.

We can also use the metric to define what the length of a vector is. The length of a vector is also called the norm of the vector, and is most often denoted by $\| \cdot \|$. The length of a vector v_p at a point p is defined to be

$$
\|v_p\| \equiv \sqrt{|g(v_p, v_p)|},
$$

where the $| \cdot |$ in the square root is simply the absolute value, which is necessary if g is a pseudo-Riemannian metric instead of a Riemannian metric.

Question A.6 Show that the Euclidian metric gives the usual dot product of two vectors, $g(v, w) = v \cdot w$. In particular, show that the matrix which defined the dot product in Chap. 5 is exactly the Euclidian metric on $\mathbb{R}^3$ defined above by having $g_{ij} = \delta_{ij}$.

Question A.7 Recall that the Minkowski metric is given by the matrix

$$
\begin{bmatrix} -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}.
$$

Show that the absolute value in the definition of the norm of a vector is necessary for this metric.

When a manifold has a metric defined on it we can also use this metric to find the distance between two points on the manifold, at least in theory. Suppose we have two points p and q that are both in the same coordinate patch of M and they are connected by a curve $\gamma : [a, b] \subset \mathbb{R} \to M$ where $\gamma(a) = p$ and $\gamma(b) = q$. The curve $\gamma(t)$ has tangent velocity vectors $\dot{\gamma}(t)$ along the curve. To ensure the tangent velocity vectors actually exist at the endpoints the curve needs to be extended a tiny amount ϵ to $(a - \epsilon, b + \epsilon) \subset \mathbb{R}$. The length of the curve γ from p to q is defined to be

$$
L(\gamma) = \int_a^b \sqrt{\left| g\big(\dot{\gamma}(t), \dot{\gamma}(t)\big) \right|}\, dt.
$$

We will not actually try to solve any equations of this nature, though in all probability you did solve a few similar to this in your multi-variable calculus course. It can be shown that the length of the curve, $L(\gamma)$, is independent of the parametrization we use to define the curve, though we will not show that here. We also point out that this is an example of an integral where the integrand is not a differential form. As you are of course aware of by this point, differential forms are very natural things to integrate because they handle changes in variable so well, but they are not the only objects that can be integrated - this example is a case in point.

We then define the distance between the two points p and q as

$$d(p,q) = \inf_{\gamma} L(\gamma),$$

where γ is any piecewise continuous curve that connects p and q. If you don't know what infimum is, it is basically the lower limit of lengths of all possible curves that connect p and q. This is of course a nice theoretical definition for the distance between two points, but in general one certainly wouldn't want to compute the distance between any two points this way. This procedure turns out to give exactly the distance formula that you are used to from Euclidian geometry.

However, the main point to understand is that distances between points on a manifold rely on the fact that there is a metric defined on the manifold. If a manifold does not have a metric on it the very idea distances between points simply does not make sense.

A.7 Lie Derivatives of Tensor Fields

We now have all the pieces in place to define a new kind of derivative called the Lie derivative, named after the mathematician Sophus Lie. The Lie derivative does exactly what we think of derivatives doing, measuring how something changes in a particular direction. In a sense, the Lie derivative conforms a little more naturally to our calculus ideas of what derivatives do than exterior derivatives, even though exterior derivatives are such natural generalizations of directional derivatives. The idea that there can be different ways measuring how objects change, and hence different ways to define differentiation, seems very odd at first. But it should come as no surprise. After all, how the derivative of a function $f : \mathbb{R}^n \to \mathbb{R}^m$ is defined is somewhat different from how the derivative of a function $f : \mathbb{R} \to \mathbb{R}$ is defined. And the more complicated a mathematical object is then it stands to reason that the more ways there are of measuring how that object changes.

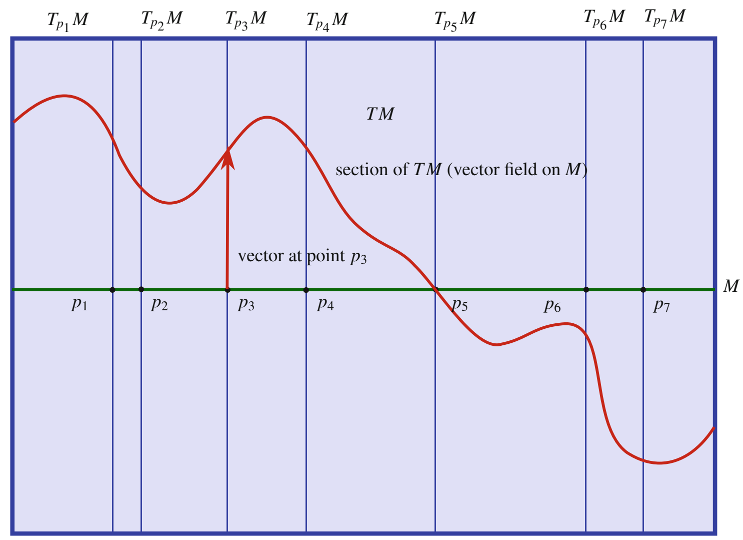
The most obvious difference is that Lie derivatives apply to any tensor field at all, whereas the exterior derivative only applies to differential forms. Of course, since differential forms are skew-symmetric covariant tensors then one can also take the Lie derivative of a differential form as well. Hence, after this section we will know two different concepts of differentiation for differential forms. As a book on differential forms we believe it is important that you at least be introduced to the idea of Lie derivatives since it is one of the ways differential forms can be differentiated.

Our introduction to Lie derivatives will be follow a standard path and be very similar to introduction that you will see in other books. Because tensors are mappings

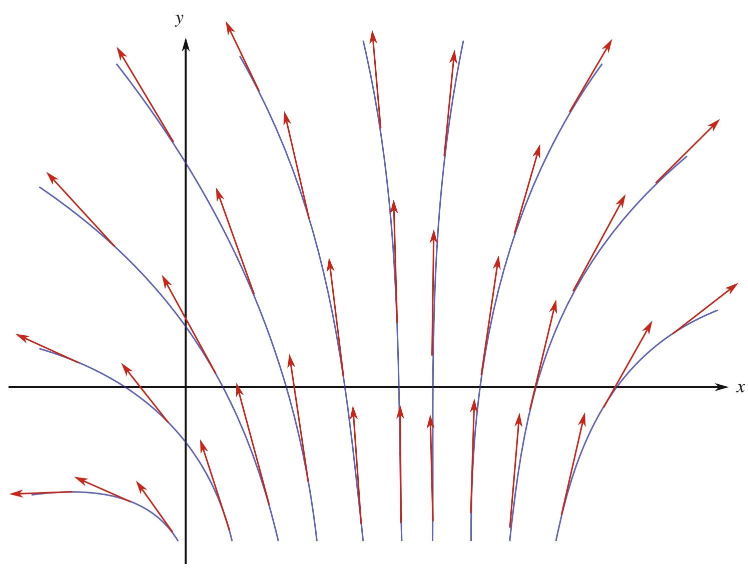
$$T : \underbrace{T^*M \times \cdots \times T^*M}_{\substack{r \\ \text{contravariant} \\ \text{degree}}} \times \underbrace{TM \times \cdots \times TM}_{\substack{s \\ \text{covariant} \\ \text{degree}}} \longrightarrow \mathbb{R}$$

and thus are made up of building blocks of the form $\frac{\partial}{\partial x^{i_1}} \otimes \cdots \otimes \frac{\partial}{\partial x^{i_r}} \otimes dx^{j_1} \otimes \cdots \otimes dx^{j_s}$ we will first consider the Lie derivative of vectors, and in particular vectors of the form ∂_x. We will then consider lie derivatives of one-forms, and in particular elements of the form dx. It is then conceptually easy, though notationally cumbersome, to extend the concept to general tensors. We then consider the Lie derivative of a function and then will derive some identities that involve the Lie derivative. Finally we will compare the Lie derivative to the exterior derivative. For the remainder of this section we will always assume we are in a single coordinate chart on a manifold. It is not difficult, though again somewhat cumbersome, to patch together coordinate charts, though we will not do that here. Furthermore, due to the amount of material we will cover in this section we will be substantially more terse than we have been throughout this book so far, but hopefully it will serve as a first introduction.

We begin by considering a smooth integrable vector field v. What we mean by this is that the vector field has nice integral curves. But what are integral curves? Of course the vector field is simply a section of the tangent bundle TM, see Fig. A.1, but smooth integrable vector fields are a special kind of vector field. Suppose that the manifold coordinates of the chart we are in are given by $(x^1, \ldots, x^n)$. Thus we can write the vector field as $v = \sum v^i(x^1, \ldots, x^n) \frac{\partial}{\partial x^i} = v^i \partial_{x^i}$ where the v^i are real-valued functions on M. The smooth integrable vector field v is a vector field for which there exists functions $\gamma : \mathbb{R} \to M$ of a parameter t, usually thought of as time, such that the velocity vectors $\dot{\gamma}$ of the curves $\gamma(t) = \big(\gamma^1(t), \ldots, \gamma^n(t)\big)$ are



A smooth section v in the tangent bundle TM.



The vector field v shown along with its integral curves.

Fig. A.1 An integrable section v of the tangent bundle TM (top) is simply a vector field v on the manifold M that has integral curves γ on M (bottom)

given by the vector field v. This means that at each point $p \in M$ we have

$$\left.\frac{d\ \gamma(t)}{dt}\right|_p = v_p.$$

Since $\gamma(t) = \big(\gamma^1(t), \ldots, \gamma^n(t)\big)$ we can write this equation as the system of n differential equations,

$$\frac{d}{dt}\gamma^i = v^i.$$

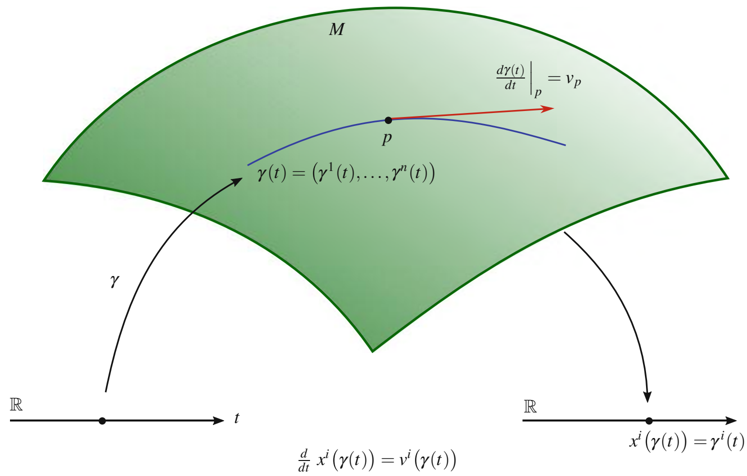


Fig. A.2 The manifold M shown with one of the integral curves γ which is obtained by solving the differential equation $\frac{d}{dt}x^i\Big(\gamma(t)\Big) = v^i\Big(\gamma(t)\Big)$. In essence, the integral curves γ are obtained by integrating the vector field v. Even though only one curve is shown here, the whole manifold M is filled with these curves

Thinking of γ as a function onto the manifold, $\gamma : \mathbb{R} \to M$, and x^i as coordinate functions on the manifold $x^i : M \to \mathbb{R}$ then we could also write this system of n differential equations as

$$\frac{d}{dt}x^i\Big(\gamma(t)\Big) = v^i\Big(\gamma(t)\Big),$$

where the $v^i : M \to \mathbb{R}$ are the real-valued component functions of the vector v. Solving this set of n differential equations gives the n integral functions γ^i which make up the integral curve γ. See Fig. A.2. Thus this particular idea of integration of a vector field on a manifold boils down to differential equations.

Integral curves γ are just that, curves on the manifold M, but we can also think of them as a family of mappings from the manifold M to itself, or to another copy of M. We begin by fixing some time t_0 as our zero time. Then for each time t we have a mapping γ_t that sends $\gamma(t_0)$ to $\gamma(t_0 + t)$,

$$\begin{aligned} M &\xrightarrow{\gamma_t} M \\ p = \gamma(t_0) &\longmapsto \gamma_t(p) = \gamma(t_0 + t). \end{aligned}$$

Each different time t gives a different mapping γ_t, thereby giving us a whole family of mappings, as shown in Fig. A.3. As soon as we have a mapping $\gamma_t : M \to M$ between manifolds we have the push-forward mapping

$$T_p\gamma_t \equiv \gamma_{t*} : T_pM \to T_{\gamma_t(p)}M,$$

which allows us to push-forward vectors, and the pullback mapping

$$T_p^*\gamma_t \equiv \gamma_t^* : T_{\gamma_t(p)}^*M \to T_p^*M,$$

which allows us to pull-back one-forms. We generally drop the point p from the notation and infer it from context. Doing so we would simply write $T\gamma_t \equiv \gamma_{t*}$ and $T^*\gamma_t \equiv \gamma_t^*$. See Fig. A.4.

Of course the mapping γ_t is invertible. Suppose we have $\gamma_t(p) = q$. Then we can define $\gamma_t^{-1}(q) = p$. Let us look closely at the definitions, $p = \gamma(t_0)$ and $q = \gamma(t_0 + t)$ for some t_0. We can run time backwards just as well as forward, so if we defined $\tilde{t}_0 = t_0 + t$ then we would have $\gamma(\tilde{t}_0) = q$ and $\gamma(\tilde{t}_0 - t) = p$. But this is exactly the same as saying $\gamma_{-t}(q) = p$, which is exactly what we want γ_t^{-1} to do. Hence, we have $\gamma_t^{-1} = \gamma_{-t}$. Therefore we also can define the push-forward and pullback of $\gamma_t^{-1} = \gamma_{-t}$. We only need the push-forward,

$$T_q\gamma_t^{-1} \equiv T_q\gamma_{-t} \equiv \gamma_{-t*} : T_qM \to T_{\gamma_{-t}(q)}M,$$

where of course $q = \gamma_t(p)$. Dropping the point q from the notation we write $T\gamma_t^{-1} \equiv T\gamma_{-t}$ and are left to infer the base point from context.

Lie Derivatives of Vector Fields

Now we are ready for the definition of the Lie derivative. We begin by giving the definition of the Lie derivative of a vector field w in the direction of v, though perhaps it is more accurate to call it the Lie derivative of w along the flow of the vector field v, at the point p,

$$\boxed{\text{Lie Derivative of Vector Field:} \qquad (\mathscr{L}_v w)_p = \lim_{t\to 0} \frac{T\gamma_{-t} \cdot w_{\gamma_t(p)} - w_p}{t}.}$$

You will of course come across a fairly wide variety of notations for the Lie derivative of a vector field, such as

$$\begin{aligned}(\mathscr{L}_v w)_p &= \lim_{t\to 0} \frac{\left(\gamma_t^{-1}\right)_* w_{\gamma_t(p)} - w_p}{t} \\ &= \lim_{t\to 0} \frac{\left(\gamma_{-t}\right)_* w_{\gamma_t(p)} - w_p}{t}\end{aligned}$$

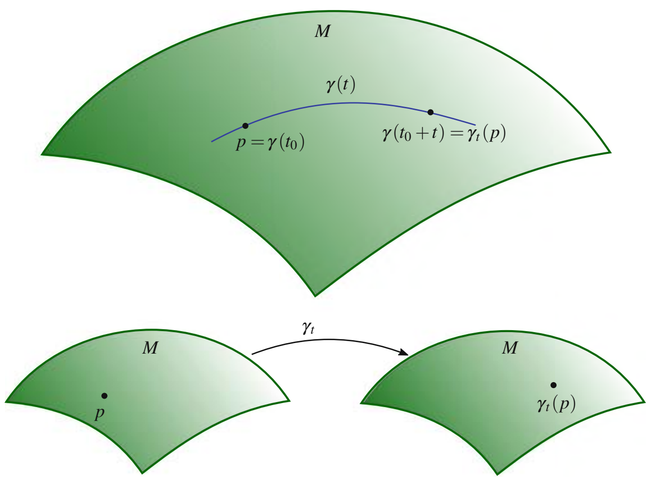


Fig. A.3 The integral curves γ obtained by integrating the vector field v can be used to define a whole family of mappings $\gamma_t : M \to M$, one for each value of t. The mapping γ_t sends each point $p = \gamma(t_0)$ of the manifold to $\gamma_t(p) = \gamma(t_0 + t)$

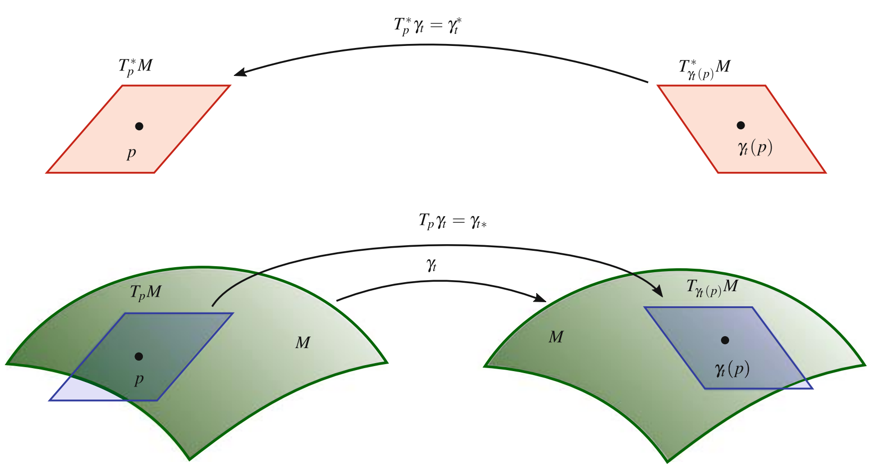


Fig. A.4 The mapping γ_t can be used to push-forward vectors with $T_p\gamma_t \equiv \gamma_{t*}$ and pullback one-forms with $T_p^*\gamma_t \equiv \gamma_t^*$

$$
\begin{aligned}
&= \lim_{t\to 0} \frac{\left(\gamma_t^{-1}\right)_* w_{\gamma_t(p)} - \left(\gamma_0^{-1}\right)_* w_{\gamma_0(p)}}{t} \\
&\equiv \frac{d}{dt}\left(\left(\gamma_t^{-1}\right)_* w_{\gamma_t(p)}\right)\Big|_0 \\
&= \frac{d}{dt}\left(\left(\gamma_{-t}\right)_* w_{\gamma_t(p)}\right)\Big|_0 .
\end{aligned}
$$

Question A.8 Show that these different notations all mean the same thing.

What is happening here? We want to understand how the vector field w is changing at the point p. In order to do this we need to first choose a direction v_p. Actually, we need just a little more than that, we need a little piece of an integral curve of a vector field v defined close to the point p. See Fig. A.5. With this we can measure how the vector field w is changing along the integral curve of v. We do this by pulling back the vector $w_{\gamma_t(p)}$ to the point p by $T\gamma_{-t}$. Thus both w_p and $T\gamma_{-t} \cdot w_{\gamma_t(p)}$ are both vectors at the point p, that is, both vectors in T_pM, so we can subtract one from the other, as shown graphically in Fig. A.6. This becomes the numerator of the difference quotient, with t being the denominator. Dividing a vector by a number t simply means dividing each component of the vector by t. Taking the limit as $t \to 0$ gives us the Lie derivative of w along v at the point p. A picture of this is shown in Fig. A.5. Thus we have that the Lie derivative is in fact another vector field, one that we have encountered already. More will be said about this in a few pages.

Lie Derivatives of One-Forms

The Lie derivative of a differential form α in the direction of v at the point p is defined in exactly the same way, only we have to use the pullback instead of push-forward,

$$
\boxed{\text{Lie Derivative of One-Form:} \qquad (\mathscr{L}_v\alpha)_p = \lim_{t\to 0} \frac{T^*\gamma_t \cdot \alpha_{\gamma_t(p)} - \alpha_p}{t},}
$$

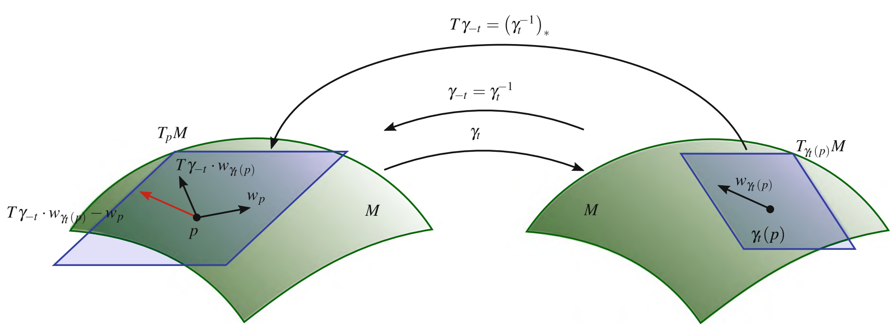


Fig. A.5 To find the Lie derivative of a vector field w in the direction v we first find the integral curves γ of v. With γ we have a whole family of maps γ_t that simply move points p of M along the flow γ by time t as shown in Fig. A.4. We can then push-forward a vector $w_{\gamma_t(p)}$ by $T\gamma_{-t}$ to the point p and subtract w_p from it. This becomes the numerator of the difference quotient for the Lie derivative

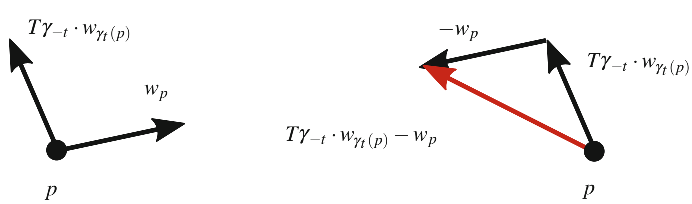


Fig. A.6 Here we graphically show how the vector w_p is subtracted from the vector $T\gamma_{-t} \cdot w_{\gamma_t(p)}$ to give the vector $T\gamma_{-t} \cdot w_{\gamma_t(p)} - w_p$

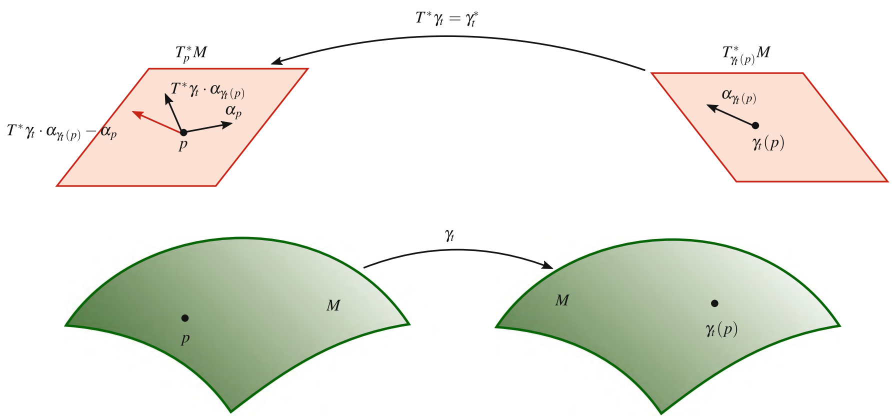


Fig. A.7 To find the Lie derivative of a one-form $w\alpha$ in the direction v we first find the integral curves γ of v. With γ we have a whole family of maps γ_t that simply move points p of M along the flow γ by time t as shown in Fig. A.4. We can then pullback a one-form $\alpha_{\gamma_t(p)}$ by $T^*\gamma_t$ to the point p and subtract α_p from it. This becomes the numerator of the difference quotient for the Lie derivative

which of course can also be written as

$$\begin{aligned}(\mathscr{L}_v\alpha)_p &= \lim_{t\to 0} \frac{\gamma_t^*\alpha_{\gamma_t(p)} - \alpha_p}{t}\\ &= \lim_{t\to 0} \frac{\gamma_t^*\alpha_{\gamma_t(p)} - \gamma_0^*\alpha_{\gamma_0(p)}}{t}\\ &\equiv \frac{d}{dt}\left((\gamma_t)^*\alpha_{\gamma_t(p)}\right)\Big|_0 .\end{aligned}$$

Notice that the Lie derivative of the one-form α is defined in an completely analogous way to the Lie derivative of the vector field w, see Fig. A.7.

Question A.9 Show that the above notations are all equivalent.

Lie Derivatives of Functions

What would the Lie derivative of a function be? Since we have previously viewed functions as zero-forms we could also view them as rank-zero covariant tensors and use the formula for the Lie derivative of a one-form to get

$$(\mathscr{L}_v f)_p = \lim_{t\to 0} \frac{T^*\gamma_t \cdot f_{\gamma_t(p)} - f_p}{t}.$$

But $f_{\gamma_t(p)}$ is simply $f\big(\gamma_t(p)\big) = f(q)$, or the numerical value of f evaluated at $\gamma_t(p) = q$. The pull-back of a number by γ_t^* is simply that number, so we have $\gamma_t^* f_{\gamma_t(p)} = f\big(\gamma_t(p)\big) = f(q)$. Similarly, f_p is simply the function f evaluated at point p, or $f(p)$. Since $p = \gamma(t_0)$ and $q = \gamma(t_0 + t)$, putting everything together we can write

$$\begin{aligned}(\mathscr{L}_v f)_p &= \lim_{t\to 0} \frac{f\big(\gamma(t_0+t)\big) - f\big(\gamma(t)\big)}{t}\\ &= \frac{\partial f\big(\gamma(t)\big)}{\partial t}\bigg|_{t_0}\\ &= \frac{\partial f}{\partial \gamma^i}\bigg|_{\gamma(t_0)} \frac{\partial \gamma^i}{\partial t}\bigg|_{t_0} .\end{aligned}$$

The last equality is simply the chain rule. The integral curve γ can be written as $\gamma(t) = \big(\gamma^1(t), \ldots, \gamma^n(t)\big)$, but what is γ^i? It is simply the x^i coordinate of the curve, so we could relabel and write the curve in more standard notation as $\gamma(t) = \big(x^1(t), \ldots, x^n(t)\big)$. Rewriting this curve in more standard notation simply allows us to recognize the end result easier, which can be written as

$$(\mathcal{L}_v f)_p = \left.\frac{\partial f}{\partial x^i}\right|_{\gamma(t_0)} \left.\frac{\partial x^i}{\partial t}\right|_{t_0}.$$

Next recall how we got the x^i coordinate γ^i in the first place, we got it from the vector v by solving the set of differential equations

$$\frac{d}{dt}\gamma^i = v^i,$$

thereby giving us

$$\begin{aligned}(\mathcal{L}_v f)_p &= \frac{\partial f}{\partial x^i}\, v^i \\ &= v^i\, \frac{\partial f}{\partial x^i} \\ &= \left(v^i \frac{\partial}{\partial x^i}\right) f \\ &= v_p[f],\end{aligned}$$

which is simply our directional derivative of f in the direction v. So for functions there is no difference between Lie derivatives and our familiar directional derivative. Thus we have obtained

$$\boxed{\text{Lie Derivative of Function:} \qquad (\mathcal{L}_v f)_p = v_p[f].}$$

Lie Derivatives of (r, s)-Tensors

Now we turn to the Lie derivative of a (r, s)-tensor $\mathcal{T}$. We will define the pull-back of the tensor $\mathcal{T}$ by γ_t with the following

$$\big(\gamma_t^* \mathcal{T}\big)_p\big(\alpha_1, \ldots, \alpha_r, v_1, \ldots, v_s\big) = \mathcal{T}_{\gamma_t(p)}\big(\gamma_{-t}^* \alpha_1, \ldots, \gamma_{-t}^* \alpha_r, \gamma_{t*} v_1, \ldots, \gamma_{t*} v_s\big).$$

Compare this to our original definition of the pullback of a one-form. Defining the pull-pack of a tensor this way we define the Lie derivative of the tensor $\mathcal{T}$ with the same formula that we used to define the Lie derivative of a one-form,

$$\boxed{\text{Lie Derivative of Tensor:} \qquad (\mathcal{L}_v \mathcal{T})_p = \lim_{t\to 0} \frac{\gamma_t^* \mathcal{T}_{\gamma_t(p)} - \mathcal{T}_p}{t}.}$$

Of course we can also write the Lie derivative of a general tensor field as

$$\begin{aligned}(\mathcal{L}_v \mathcal{T})_p &= \lim_{t\to 0} \frac{\gamma_t^* \mathcal{T}_{\gamma_t(p)} - \gamma_0^* \mathcal{T}_{\gamma_0(p)}}{t} \\ &= \frac{d}{dt}\left.\left(\gamma_t^* \mathcal{T}_{\gamma_t(p)}\right)\right|_0.\end{aligned}$$

Some Lie Derivative Identities

Now we turn to showing a number of important identities that involve the Lie derivative. We move thorough these identities rather quickly, leaving some as exercises. The first identity is that regardless of the kind of tensors $\mathcal{S}$ and $\mathcal{T}$ are, the Lie derivative is linear. That is, for $a, b \in \mathbb{R}$, we have

$$\boxed{\mathcal{L}_v\big(a\mathcal{S} + b\mathcal{T}\big) = a\mathcal{L}_v\mathcal{S} + b\mathcal{L}_v\mathcal{T}.}$$

Question A.10 Using the definition of the Lie derivative of a tensor, show that $\mathcal{L}_v\big(a\mathcal{S} + b\mathcal{T}\big) = a\mathcal{L}_v\mathcal{S} + b\mathcal{L}_v\mathcal{T}$.

Question A.11 Using the definition of the Lie derivative of a tensor, show that $\phi^*\big(\mathscr{T} \otimes \mathscr{S}\big) = \phi^*\mathscr{T} \otimes \phi^*\mathscr{S}$

Turning to the Lie derivative of the tensor product of two tensors, and dropping the base points from our notation and relying on the identity $\phi^*\big(\mathscr{T} \otimes \mathscr{S}\big) = \phi^*\mathscr{T} \otimes \phi^*\mathscr{S}$, we find that

$$\begin{aligned}
\mathscr{L}_v(\mathscr{S} \otimes \mathscr{T}) &= \lim_{t\to 0} \frac{\gamma_t^*(\mathscr{S} \otimes \mathscr{T}) - \mathscr{S} \otimes \mathscr{T}}{t} \\
&= \lim_{t\to 0} \frac{\gamma_t^*\mathscr{S} \otimes \gamma_t^*\mathscr{T} - \mathscr{S} \otimes \mathscr{T}}{t} \\
&= \lim_{t\to 0} \frac{\big(\mathscr{S} + (\gamma_t^*\mathscr{S} - \mathscr{S})\big) \otimes \big(\mathscr{T} + (\gamma_t^*\mathscr{T} - \mathscr{T})\big) - \mathscr{S} \otimes \mathscr{T}}{t} \\
&= \lim_{t\to 0} \frac{(\gamma_t^*\mathscr{S} - \mathscr{S}) \otimes \mathscr{T} + \mathscr{S} \otimes (\gamma_t^*\mathscr{T} - \mathscr{T}) + (\gamma_t^*\mathscr{S} - \mathscr{S}) \otimes (\gamma_t^*\mathscr{T} - \mathscr{T})}{t} \\
&= \lim_{t\to 0}\left(\frac{\gamma_t^*\mathscr{S} - \mathscr{S}}{t} \otimes \mathscr{T}\right) + \lim_{t\to 0}\left(\mathscr{S} \otimes \frac{\gamma_t^*\mathscr{T} - \mathscr{T}}{t}\right) + \lim_{t\to 0}\left(\frac{\gamma_t^*\mathscr{S} - \mathscr{S}}{t} \otimes (\gamma_t^*\mathscr{T} - \mathscr{T})\right) \\
&= \mathscr{L}_v\mathscr{S} \otimes \mathscr{T} + \mathscr{S} \otimes \mathscr{L}_v\mathscr{T} + \mathscr{L}_v\mathscr{S} \otimes 0 \\
&= \mathscr{L}_v\mathscr{S} \otimes \mathscr{T} + \mathscr{S} \otimes \mathscr{L}_v\mathscr{T}.
\end{aligned}$$

It is clear that as $t \to 0$ we have $(\gamma_t^*\mathscr{T} - \mathscr{T}) \to (\mathscr{T} - \mathscr{T}) = 0$ in the next to last line.

Question A.12 In the fifth line of the above string of equalities we used that

$$\lim_{t\to 0} \frac{(\gamma_t^*\mathscr{S} - \mathscr{S}) \otimes (\gamma_t^*\mathscr{T} - \mathscr{T})}{t} = \lim_{t\to 0}\left(\frac{\gamma_t^*\mathscr{S} - \mathscr{S}}{t} \otimes (\gamma_t^*\mathscr{T} - \mathscr{T})\right).$$

Show that this is true. Also show that

$$\lim_{t\to 0} \frac{(\gamma_t^*\mathscr{S} - \mathscr{S}) \otimes (\gamma_t^*\mathscr{T} - \mathscr{T})}{t} = \lim_{t\to 0}\left((\gamma_t^*\mathscr{S} - \mathscr{S}) \otimes \frac{\gamma_t^*\mathscr{T} - \mathscr{T}}{t}\right).$$

Explain that the final result in the proof of the above identity does not depend on which of the equalities in this question is used.

Putting everything together we have shown the identity

$$\boxed{\mathscr{L}_v(\mathscr{S} \otimes \mathscr{T}) = \mathscr{L}_v\mathscr{S} \otimes \mathscr{T} + \mathscr{S} \otimes \mathscr{L}_v\mathscr{T}.}$$

Similarly, relying on the identity $\phi^*(\alpha \wedge \beta) = \phi^*\alpha \wedge \phi^*\beta$ from Sect. 6.7 we have

$$\begin{aligned}
\mathscr{L}_v(\alpha \wedge \beta) &= \lim_{t\to 0} \frac{\gamma_t^*(\alpha \wedge \beta) - \alpha \wedge \beta}{t} \\
&= \lim_{t\to 0} \frac{\gamma_t^*\alpha \wedge \gamma_t^*\beta - \alpha \wedge \beta}{t} \\
&= \lim_{t\to 0} \frac{\big(\alpha + (\gamma_t^*\alpha - \alpha)\big) \wedge \big(\beta + (\gamma_t^*\beta - \beta)\big) - \alpha \wedge \beta}{t} \\
&= \lim_{t\to 0} \frac{(\gamma_t^*\alpha - \alpha) \wedge \beta + \alpha \wedge (\gamma_t^*\beta - \beta) + (\gamma_t^*\alpha - \alpha) \wedge (\gamma_t^*\beta - \beta)}{t} \\
&= \lim_{t\to 0}\left(\frac{\gamma_t^*\alpha - \alpha}{t} \wedge \beta\right) + \lim_{t\to 0}\left(\alpha \wedge \frac{\gamma_t^*\beta - \beta}{t}\right) + \lim_{t\to 0}\left(\frac{\gamma_t^*\alpha - \alpha}{t} \wedge (\gamma_t^*\beta - \beta)\right) \\
&= \mathscr{L}_v\alpha \wedge \beta + \alpha \wedge \mathscr{L}_v\beta + \mathscr{L}_v\alpha \wedge 0 \\
&= \mathscr{L}_v\alpha \wedge \beta + \alpha \wedge \mathscr{L}_v\beta.
\end{aligned}$$

Again it is clear that as $t \to 0$ we have $(\gamma_t^*\beta - \beta) \to (\beta - \beta) = 0$ in the next to last line.

Question A.13 The above proof used that

$$\lim_{t\to 0} \frac{(\gamma_t^*\alpha - \alpha) \wedge (\gamma_t^*\beta - \beta)}{t} = \lim_{t\to 0} \left(\frac{\gamma_t^*\alpha - \alpha}{t} \wedge (\gamma_t^*\beta - \beta) \right).$$

Show this is true. Also, show that

$$\begin{aligned}\lim_{t\to 0} \frac{(\gamma_t^*\alpha - \alpha) \wedge (\gamma_t^*\beta - \beta)}{t} &= \lim_{t\to 0} \left(\frac{\gamma_t^*\alpha - \alpha}{t} \wedge (\gamma_t^*\beta - \beta) \right) \\ &= \lim_{t\to 0} \left((\gamma_t^*\alpha - \alpha) \wedge \frac{\gamma_t^*\beta - \beta}{t} \right).\end{aligned}$$

Explain that the final result in the proof of the above identity does not depend on which of the equalities in this question are used.

Putting everything together we have shown the identity

$$\boxed{\mathscr{L}_v(\alpha \wedge \beta) = \mathscr{L}_v\alpha \wedge \beta + \alpha \wedge \mathscr{L}_v\beta.}$$

The next identity involves functions f and vector fields v and w. It is

$$\boxed{\mathscr{L}_v(fw) = f\mathscr{L}_v w + v[f]w.}$$

Question A.14 Using the definitions and identities already given and proved, show that $\mathscr{L}_v(fw) = f\mathscr{L}_v w + v[f]w$.

Another important identity is

$$\boxed{d\big(\mathscr{L}_v\alpha\big) = \mathscr{L}_v\big(d\alpha\big).}$$

This follows from the commutativity of pull-backs and exterior differentiation, $\phi^* d = d\phi^*$.

Question A.15 Using the definition of the Lie derivative for a covariant tensor along with the fact that $\phi^* d = d\phi^*$, show that $d\big(\mathscr{L}_v\alpha\big) = \mathscr{L}_v\big(d\alpha\big)$.

Cartan's Homotopy Formula

We now prove a very important identity called either **Cartan's homotopy formula** or sometimes **Cartan's magic formula**. Given a k-form α and a vector field v, and recalling the definition of the interior product ι from Sect. 3.4, Cartan's magic formula is

$$\mathscr{L}_v\alpha = \iota_v(d\alpha) + d(\iota_v\alpha).$$

While there are several ways to show this identity we will use induction. First we consider our base case for a zero-form $\alpha = f$. The interior product $\iota_v f$ is not defined for a zero-form hence the identity is simply

$$\mathscr{L}_v f = \iota_v(df).$$

But this is obviously true, as long as you remember all the different ways of writing the same thing. Recall, that we found the Lie derivative of a function is simply the directional derivative of the function, that is, $\mathscr{L}_v f = v[f]$. Also recall from early in the book that $v[f] \equiv df(v)$ and of course $df(v) = \iota_v df$ by definition of the interior product. Hence we get the following string of equalities that are little more than different ways of writing the same thing,

$$\mathscr{L}_v f = v[f] = df(v) = \iota_v df.$$

This classic example of "proof by notation" establishes our base case. Also notice that if the function f was a coordinate function x^i then we would have $\mathscr{L}_v x^i = \iota_v dx^i$. We will need this fact in a moment.

Now assume that the identity $\mathscr{L}_v \alpha = \iota_v(d\alpha) + d(\iota_v \alpha)$ holds for all $(k-1)$-forms α. Suppose we have a k-form $\omega = f dx^1 \wedge \cdots \wedge dx^k$, which is clearly a base term for a general k-form. We can write $\omega = dx^1 \wedge \alpha$ where $\alpha = f dx^2 \wedge \cdots \wedge dx^k$, a $(k-1)$-form. Recalling one of the identities we just proved, the left hand side of Cartan's magic formula becomes

$$\begin{aligned}\mathscr{L}_v \omega &= \mathscr{L}_v (dx^1 \wedge \alpha) \\ &= \mathscr{L}_v dx^1 \wedge \alpha + dx^1 \wedge \mathscr{L}_v \alpha.\end{aligned}$$

Before turning to the right hand side of Cartan's magic formula we recall a couple of identities we will need. The first is an identity of the interior product proved in Sect. 3.4. If α is a k-form and β is another differential form, then

$$\begin{aligned}\iota_v(\alpha \wedge \beta) &= (\iota_v \alpha) \wedge \beta + (-1)^k \alpha \wedge (\iota_v \beta) \\ &= \iota_v \alpha \wedge \beta + (-1)^k \alpha \wedge \iota_v \beta.\end{aligned}$$

The second is an identity of the exterior derivative. This identity was either derived as a consequence of how we defined exterior derivative, as in Sect. 4.2, or it was viewed as an axiom for exterior differentiation that allowed us to find the formula for the exterior derivative, as in Sect. 4.3. If α is a k-form and β is another differential form, then

$$\begin{aligned}d(\alpha \wedge \beta) &= (d\alpha) \wedge \beta + (-1)^k \alpha \wedge (d\beta) \\ &= d\alpha \wedge \beta + (-1)^k \alpha \wedge d\beta.\end{aligned}$$

Now we consider the right hand side of Cartan's magic formula,

$$\begin{aligned}\iota_v(d\omega) + d(\iota_v \omega) &= \iota_v d\big(dx^1 \wedge \alpha\big) + d\iota_v\big(dx^1 \wedge \alpha\big) \\ &= \iota_v\Big(\underbrace{ddx^1}_{=0} \wedge \alpha + (-1)^1 dx^1 \wedge d\alpha\Big) + d\Big(\iota_v dx^1 \wedge \alpha + (-1)^1 dx^1 \wedge i_v \alpha\Big) \\ &= -\iota_v\big(dx^1 \wedge d\alpha\big) + d\big(\iota_v dx^1 \wedge \alpha\big) - d\big(dx^1 \wedge \iota_v \alpha\big) \\ &= -\cancel{\iota_v dx^1 \wedge d\alpha} + dx^1 \wedge \iota_v d\alpha \\ &\qquad + d\underbrace{\iota_v dx^1}_{\mathscr{L}_v x^1} \wedge \alpha + (-1)^0 \cancel{\iota_v dx^1 \wedge d\alpha} \\ &\qquad\quad - \underbrace{ddx^1}_{=0} \wedge \iota_v \alpha - (-1)^1 dx^1 \wedge d\iota_v \alpha \\ &= d\big(\mathscr{L}_v x^1\big) \wedge \alpha + dx^1 \wedge \iota_v d\alpha + dx^1 \wedge d\iota_v \alpha \\ &= \mathscr{L}_v dx^1 \wedge \alpha + dx^1 \wedge \Big(\underbrace{\iota_v d\alpha + d\iota_v \alpha}_{\substack{=\mathscr{L}_v \alpha \\ \text{induction hypothesis}}} \Big) \\ &= \mathscr{L}_v dx^1 \wedge \alpha + dx^1 \wedge \mathscr{L}_v \alpha.\end{aligned}$$

So both the left and the right hand side of Cartan's magic formula for $\omega = dx^1 \wedge \alpha$ are equal to the same thing, namely $\mathscr{L}_v dx^1 \wedge \alpha + dx^1 \wedge \mathscr{L}_v \alpha$, and are therefore equal to each other.

Question A.16 Use linearity of exterior differentiation, the Lie derivative, and the interior product, to show that Cartan's magic formula applies to every differential form of the form $\omega = \sum f_{i_1 \cdots i_k} dx^{i_1} \wedge \cdots \wedge dx^{i_k}$.

Thus we have shown for any differential form α that

$$\boxed{\text{Cartan's Homotopy Formula:} \qquad \mathscr{L}_v \alpha = \iota_v(d\alpha) + d(\iota_v \alpha).}$$

This identity is very often simply written as $\mathscr{L}_v = \iota_v d + d\iota_v$.

Lie Derivative Equals Commutator

We now turn our attention to showing that the Lie derivative of the vector field w in the direction v is equal to the commutator of v and w. In other words, we will show the identity

$$\boxed{\mathscr{L}_v w = [v, w] = vw - wv.}$$

We first encountered the commutator of v and w, also called the Lie-brackets of v and w, in Sect. 4.4.2 where we attempted to explain what $[v, w]$ was from a somewhat geometrical perspective. The proof of this identity is almost too slick, amounting to little more than some clever playing with definitions and notation.

First let us consider the situation where we have a map $\phi : M \to N$ and a vector w_p on M and a real-valued function $f : N \to \mathbb{R}$. We can use ϕ to push-forward the vector w_p on M to the vector $T_p\phi \cdot w_p$ on N. We can then find the directional derivative of f in the direction of $T_p\phi \cdot w_p$ at the point $\phi(p)$. This is a numerical value. By definition, the push-forward of a vector by a map ϕ acting on a function f gives the same numerical value as the vector acting on $f \circ \phi$ at the point p, that is,

$$w_p[f \circ \phi] = \big(T_p\phi \cdot w_p\big)_{\phi(p)}[f].$$

We reiterate, as numerical values these are equal. But as we know, the derivative of a function is another function, the numerical value is only obtained when we evaluate this derived function at a point. If w were actually a vector field on M then $w[f \circ \phi]$ is a function on M. Similarly, we have $\big(T\phi \cdot w\big)[f]$ is a function on N. But clearly, since these two functions are not even on the same manifold

$$w[f \circ \phi] \neq \big(T\phi \cdot w\big)[f].$$

How do we fix this? Since $\big(T\phi \cdot w\big)[f]$ is a function on N, which is nothing more than a zero-form, we can pull it back to M. Recalling how the pull-backs of zero-forms are defined we have

$$T^*\phi \cdot \Big(\big(T\phi \cdot w\big)[f]\Big) = \Big(\big(T\phi \cdot w\big)[f]\Big) \circ \phi,$$

which is now a function on M that gives the same numerical values as $w[f \circ \phi]$ does, which we can also rewrite as $w[T^*\phi \cdot f]$. See Fig. A.8. Thus we end up with the rather slick equality of functions on M,

$$w[T^*\phi \cdot f] = T^*\phi \cdot \Big(\big(T\phi \cdot w\big)[f]\Big).$$

This procedure illustrates the general idea that we need to use.

Now we consider the specific situation that now concerns us. Suppose we have an integrable vector field v on M with integral curves γ parameterized by time t. Then for each time t we get a mapping $\gamma(t) : M \to \widetilde{M}$, where we simply have the range manifold $\widetilde{M} = M$. We can run time backwards to get the inverse map $(\gamma_t)^{-1} = \gamma_{-t} : \widetilde{M} \to M$. Now suppose we have a vector field w on $M = \widetilde{M}$ and a real-valued mapping $f : M \to \mathbb{R}$, see Fig. A.9. We have the following numerical values being equal,

$$\big(T_{\gamma_t(p)}\gamma_{-1} \cdot w_{\gamma_t(p)}\big)_p[f] = w_{\gamma_t(p)}[f \circ \gamma_{-t}].$$

But if we take away the base point we no longer have numerical values but functions on each side of this equality; on left hand side $\big(T\gamma_{-1} \cdot w\big)[f]$ is a function on M and the right hand side $w[f \circ \gamma_{-t}]$ as a function on $\widetilde{M}$. To get an equality between functions we can pull-back the right hand side to M we get an equality of function on M,

$$\big(T\gamma_{-1} \cdot w\big)[f] = \big(w[f \circ \gamma_{-t}]\big) \circ \gamma_t,$$

which can also be rewritten as

$$\big(T\gamma_{-1} \cdot w\big)[f] = T^*\gamma_t \cdot \big(w[T^*\gamma_{-t} \cdot f]\big).$$

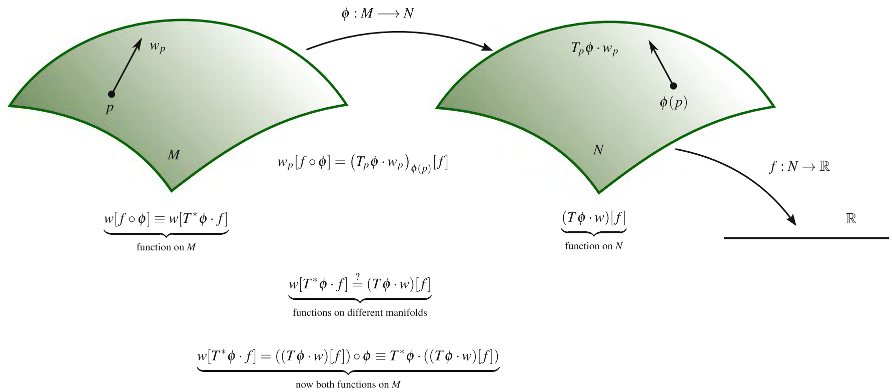


Fig. A.8 From one of the questions in Sect. 10.3 we know that $w_p[f \circ \phi] = \big(T_p\phi \cdot w_p\big)_{\phi(p)}[f]$. But this is an equality of values. By eliminating the base point the right hand side becomes $w[f \circ \phi]$, the directional derivative of $f \circ \phi$ in the w direction, which is itself a function on M. Similarly, the left hand side becomes $(T\phi \cdot w)[f]$, the directional derivative of f in the $T\phi \cdot w$ direction, which is itself a function on N. Thus $v[f \circ \phi]$ and $T\phi \cdot w[f]$ can not be equal as functions. But this can be fixed by pulling-back the zero-form (function) $(T\phi \cdot w)[f]$ by $T^*\phi$. The pull-back of a zero-form was also defined in Sect. 10.3

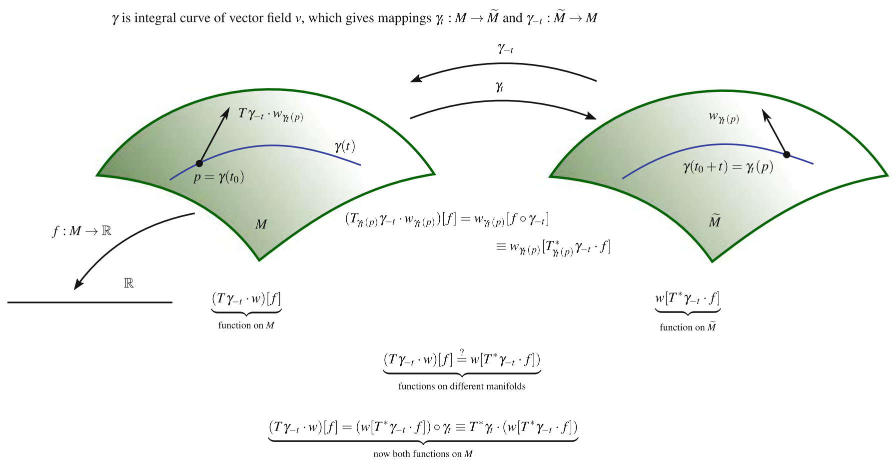


Fig. A.9 Here a procedure similar to that done in Fig. A.8 allows us to equate two functions on M, obtaining $T\gamma_{-t} \cdot w[f] = T^*\gamma_t \cdot (w[T^*\gamma_{-t} \cdot f])$. This is the particular case we need in the proof of $\mathscr{L}_v w = [v, w]$

This is the exact identity we need.

Now we are finally ready for the proof of $\mathscr{L}_v w = [v, w] = vw - wv$. We begin by writing the definition of the Lie derivative of a vector field w in the direction v,

$$(\mathscr{L}_v w)_p = \lim_{t \to 0} \frac{T\gamma_{-t} \cdot w_{\gamma_t(p)} - w_p}{t}.$$

Since we are interested in a general formula independent of base point p we will use this definition without the base point. If you wanted to put the base points back in you would need to think carefully about where the vectors w are located at. Also, since $\mathscr{L}_v w$ is itself a vector field we can use it to take the directional derivative of a function f, so we have

$$\begin{aligned}
(\mathscr{L}_v w)[f] &= \left(\lim_{t \to 0} \frac{T\gamma_{-t} \cdot w - w}{t} \right)[f] \\
&= \lim_{t \to 0} \frac{(T\gamma_{-t} \cdot w)[f] - w[f]}{t} \\
&= \lim_{t \to 0} \frac{T^*\gamma_t \cdot \big(w[T^*\gamma_{-t} \cdot f]\big) - w[f]}{t}
\end{aligned}$$

$$
\begin{aligned}
&= \lim_{t\to 0} \frac{T^*\gamma_t \cdot \big(w[T^*\gamma_{-t}\cdot f]\big) - T^*\gamma_t \cdot \big(w[f]\big) + T^*\gamma_t \cdot \big(w[f]\big) - w[f]}{t} \\
&= \lim_{t\to 0} \frac{T^*\gamma_t \cdot \big(w[T^*\gamma_{-t}\cdot f]\big) - T^*\gamma_t \cdot \big(w[f]\big)}{t} + \lim_{t\to 0} \frac{T^*\gamma_t \cdot \big(w[f]\big) - w[f]}{t} \\
&= \lim_{t\to 0} T^*\gamma_t \cdot \left(w\left[\frac{T^*\gamma_{-t}\cdot f - f}{t} \right]\right) + \lim_{t\to 0} \frac{T^*\gamma_t \cdot \big(w[f]\big) - w[f]}{t} \\
&= \lim_{t\to 0} T^*\gamma_t \cdot \left(w\left[\frac{f\circ\gamma_{-t} - f}{t} \right]\right) + \lim_{t\to 0} \frac{\big(w[f]\big)\circ\gamma_t - w[f]}{t} \\
&= T^*\gamma_0 \cdot \left(w\left[-v[f]\right]\right) + v\big[w[f]\big] \\
&= -w\big[v[f]\big] + v\big[w[f]\big] \\
&= [v,w]f.
\end{aligned}
$$

Writing without the function f gives us $\mathscr{L}_v w = [v,w] = vw - wv$.

Question A.17 Show that

$$
\frac{T^*\gamma_t \cdot \big(w[T^*\gamma_{-t}\cdot f]\big) - T^*\gamma_t \cdot \big(w[f]\big)}{t} = T^*\gamma_t \cdot \left(w\left[\frac{T^*\gamma_{-t}\cdot f - f}{t} \right]\right).
$$

Question A.18 Recalling that γ is the integral curve of the vector field v show that

$$
\lim_{t\to 0} \frac{\big(w[f]\big)\circ\gamma_t - w[f]}{t} = v\big[w[f]\big].
$$

Question A.19 Recalling that γ is the integral curve of the vector field v show that

$$
\lim_{t\to 0} \frac{f\circ\gamma_{-t} - f}{t} = -v[f].
$$

Here we will simply take a moment to write a formula for the commutator. Suppose $v = \sum v^i \frac{\partial}{\partial x^i} = v^i \frac{\partial}{\partial x^i}$ and $w = \sum w^j \frac{\partial}{\partial x^j} = w^j \frac{\partial}{\partial x^j}$. The Lie derivative $\mathscr{L}_v w$ is itself a vector field and as such it can act on a function f; that is, we can find $\mathscr{L}_v w[f]$, the directional derivative of f in the direction $\mathscr{L}_v w$. What the identity means is that

$$
\begin{aligned}
\mathscr{L}_v w[f] &= v[w[f]] - w[v[f]] \\
&= v\left[w^j \frac{\partial}{\partial x^j}[f] \right] - w\left[v^i \frac{\partial}{\partial x^i}[f] \right] \\
&= v\left[w^j \frac{\partial f}{\partial x^j} \right] - w\left[v^i \frac{\partial f}{\partial x^i} \right] \\
&= v^i \frac{\partial}{\partial x^i}\left[w^j \frac{\partial f}{\partial x^j} \right] - w^j \frac{\partial}{\partial x^j}\left[v^i \frac{\partial f}{\partial x^i} \right] \\
&= v^i \frac{\partial w^j}{\partial x^i}\frac{\partial f}{\partial x^j} + v^i w^j \frac{\partial^2 f}{\partial x^i \partial x^j} - w^j \frac{\partial v^i}{\partial x^j}\frac{\partial f}{\partial x^i} - w^j v^i \frac{\partial^2 f}{\partial x^j \partial x^i} \\
&= v^i \frac{\partial w^j}{\partial x^i}\frac{\partial f}{\partial x^j} - w^j \frac{\partial v^i}{\partial x^j}\frac{\partial f}{\partial x^i} \\
&= \left(v^i \frac{\partial w^j}{\partial x^i}\frac{\partial}{\partial x^j} - w^j \frac{\partial v^i}{\partial x^j}\frac{\partial}{\partial x^i} \right) f.
\end{aligned}
$$

Thus we could also write

$$\boxed{\mathscr{L}_v w = v^i \frac{\partial w^j}{\partial x^i}\frac{\partial}{\partial x^j} - w^j \frac{\partial v^i}{\partial x^j}\frac{\partial}{\partial x^i}.}$$

Global Formula for Lie Derivative of Differential Forms

Now we turn our attention to the global formula for Lie derivatives of differential forms. This formula will be used in the proof of the global formula for exterior differentiation. The global formula for Lie derivatives of differential forms is given by

$$\big(\mathscr{L}_v\alpha\big)(w_1,\dots,w_k) = v\big[\alpha(w_1,\dots,w_k)\big] - \sum_{i=1}^{k} \alpha(w_1,\dots,[v,w_i],\dots,w_k).$$

We will actually carry out the proof for the two-form case and leave the general case as an exercise. Other than being notationally more complex, the general case is similar to the two-form case. Letting γ be the integral curves of v, α a two-form, and w_1, w_2 vector fields, we note that $\alpha(w_1,w_2)$ is a function. Thus we have

$$\begin{aligned}
\big(\mathscr{L}_v\alpha(w_1,w_2)\big)_p &= \lim_{t\to 0} \frac{T^*\gamma_t\big(\alpha(w_1,w_2)\big)_{\gamma_t(p)} - \big(\alpha(w_1,w_2)\big)_p}{t} \\
&= \lim_{t\to 0} \frac{\alpha_{\gamma_t(p)}\left(w_{1_{\gamma_t(p)}}, w_{2_{\gamma_t(p)}}\right) - \alpha_p\left(w_{1_p}, w_{2_p}\right)}{t} \\
&= \lim_{t\to 0} \frac{\alpha_{\gamma_t(p)}\left(w_{1_{\gamma_t(p)}}, w_{2_{\gamma_t(p)}}\right) - \alpha_p\left(T\gamma_{-t}\cdot w_{1_{\gamma_t(p)}}, T\gamma_{-t}\cdot w_{2_{\gamma_t(p)}}\right)}{t} \\
&\quad + \lim_{t\to 0} \frac{\alpha_p\left(T\gamma_{-t}\cdot w_{1_{\gamma_t(p)}}, T\gamma_{-t}\cdot w_{2_{\gamma_t(p)}}\right) - \alpha_p\left(w_{1_p}, T\gamma_{-t}\cdot w_{2_{\gamma_t(p)}}\right)}{t} \\
&\quad + \lim_{t\to 0} \frac{\alpha_p\left(w_{1_p}, T\gamma_{-t}\cdot w_{2_{\gamma_t(p)}}\right) - \alpha_p\left(w_{1_p}, w_{2_p}\right)}{t}.
\end{aligned}$$

The first term following the last equality gives us

$$\begin{aligned}
&\lim_{t\to 0} \frac{\alpha_{\gamma_t(p)}\left(w_{1_{\gamma_t(p)}}, w_{2_{\gamma_t(p)}}\right) - \alpha_p\left(T\gamma_{-t}\cdot w_{1_{\gamma_t(p)}}, T\gamma_{-t}\cdot w_{2_{\gamma_t(p)}}\right)}{t} \\
&\quad = \lim_{t\to 0} \frac{T^*\gamma_t\cdot\alpha_{\gamma_t(p)}\left(T\gamma_{-t}\cdot w_{1_{\gamma_t(p)}}, T\gamma_{-t}\cdot w_{2_{\gamma_t(p)}}\right) - \alpha_p\left(T\gamma_{-t}\cdot w_{1_{\gamma_t(p)}}, T\gamma_{-t}\cdot w_{2_{\gamma_t(p)}}\right)}{t} \\
&\quad = \lim_{t\to 0} \left(\frac{T^*\gamma_t\cdot\alpha_{\gamma_t(p)} - \alpha_p}{t}\right)\left(T\gamma_{-t}\cdot w_{1_{\gamma_t(p)}}, T\gamma_{-t}\cdot w_{2_{\gamma_t(p)}}\right) \\
&\quad = \big(\mathscr{L}_v\alpha\big)_p(w_{1_p}, w_{2_p}).
\end{aligned}$$

The second term following the last equality gives us

$$\begin{aligned}
&\lim_{t\to 0} \frac{\alpha_p\left(T\gamma_{-t}\cdot w_{1_{\gamma_t(p)}}, T\gamma_{-t}\cdot w_{2_{\gamma_t(p)}}\right) - \alpha_p\left(w_{1_p}, T\gamma_{-t}\cdot w_{2_{\gamma_t(p)}}\right)}{t} \\
&\quad = \lim_{t\to 0} \alpha_p\left(\frac{T\gamma_{-t}\cdot w_{1_{\gamma_t(p)}} - w_{1_p}}{t}, T\gamma_{-t}\cdot w_{2_{\gamma_t(p)}}\right) \\
&\quad = \alpha_p\left((\mathscr{L}_v w_1)_p, w_{2_p}\right).
\end{aligned}$$

The third term following the last equality is very similar, giving us

$$\begin{aligned}
\lim_{t\to 0} \frac{\alpha_p\left(w_{1_p}, T\gamma_{-t}\cdot w_{2_{\gamma_t(p)}}\right) - \alpha_p\left(w_{1_p}, w_{2_p}\right)}{t}
&= \lim_{t\to 0} \alpha_p\left(w_{1_p}, \frac{T\gamma_{-t}\cdot w_{2_{\gamma_t(p)}} - w_{2_p}}{t}\right) \\
&= \alpha_p\left(w_{1_p}, (\mathscr{L}_v w_2)_p\right).
\end{aligned}$$

Thus we have, leaving off the base point,

$$\mathscr{L}_v\alpha(w_1, w_2) = \big(\mathscr{L}_v\alpha\big)(w_1, w_2) + \alpha(\mathscr{L}_v w_1, w_2) + \alpha(w_1, \mathscr{L}_v w_2).$$

This can be rewritten as

$$\big(\mathscr{L}_v\alpha\big)(w_1, w_2) = \mathscr{L}_v \underbrace{\alpha(w_1, w_2)}_{\text{a function}} - \alpha(\mathscr{L}_v w_1, w_2) - \alpha(w_1, \mathscr{L}_v w_2).$$

Recalling the definition of the Lie derivative of a function and the Lie derivative commutator identity this can again be rewritten as

$$\big(\mathscr{L}_v\alpha\big)(w_1, w_2) = v\big[\alpha(w_1, w_2)\big] - \alpha([v, w_1], w_2) - \alpha(w_1, [v, w_2]).$$

This is the global formula for the Lie derivative of a two-form.

Question A.20 Prove the global formula for the Lie derivative of a general k-form.

Global Lie Derivative Formula for differential forms:

$$\big(\mathscr{L}_v\alpha\big)(w_1, \ldots, w_k) = v\big[\alpha(w_1, \ldots, w_k)\big] - \sum_{i=1}^{k} \alpha(w_1, \ldots, [v, w_i], \ldots, w_k).$$

The global Lie derivative formula for a general rank (r, s)-tensor $\mathscr{T}$ is given by

Global Lie Derivative Formula for (r, s)-tensors:

$$\begin{aligned}
\big(\mathscr{L}_v\mathscr{T}\big)(\alpha_1, \ldots, \alpha_r, w_1, \ldots, w_s) &= v\big[\mathscr{T}(\alpha_1, \ldots, \alpha_r, w_1, \ldots, w_s)\big] \\
&\quad - \sum_{i=1}^{r} \mathscr{T}(\alpha_1, \ldots, \mathscr{L}_v\alpha_i, \ldots, \alpha_r, w_1, \ldots, w_s) \\
&\quad - \sum_{i=1}^{s} \mathscr{T}(\alpha_1, \ldots, \alpha_r, v_1, \ldots, \mathscr{L}_v w_i, \ldots, w_s).
\end{aligned}$$

Question A.21 Prove the global Lie derivative formula for a general rank (r, s)-tensor $\mathscr{T}$.

Global Formula for Exterior Differentiation

There are actually quite a number of other important identities and formulas that involve Lie derivatives, and if this were a full-fledged course in differential geometry they would need to be presented and proved. However, we will forego them. We have now covered Lie derivatives in sufficient detail that you have a basic understanding of what they are. We now will close this appendix with exploring the relationship between Lie derivatives and exterior differentiation. Since exterior derivatives only apply to differential forms, or skew-symmetric covariant tensors, then it is obvious that Lie derivatives are, in a sense, more general then exterior derivatives since you can take the Lie derivatives of any kind of tensors.

Let's take another look at Cartan's magic formula,

$$\mathscr{L}_v\alpha = \iota_v(d\alpha) + d(\iota_v\alpha).$$

The Lie derivatives of the k-form α along the flow of v is equal to the interior product by v of the exterior derivative of α plus the exterior derivative of the interior product by v of α. In other words, Cartan's magic formula allows us to write the Lie derivative of α in terms of the exterior derivatives of α. Thus we start to see that there really is a relationship between the two kinds of derivatives.

The question is, can we go the other way and write the exterior derivative in terms of the Lie derivative? The answer is of course yes, but it is in terms of the Lie derivative of vector fields, not k-forms. In fact, this formula is also called the global formula for exterior differentiation and was given in Sect. 4.4.2. The proof of the global formula for exterior differentiation for a general k-form requires induction. We begin by showing the formula for a one-form α,

$$\begin{aligned} d\alpha(v, w) &= v[\alpha(w)] - w[\alpha(v)] - \alpha\big([v, w]\big) \\ &= v\alpha(w) - w\alpha(v) - \alpha\big([v, w]\big). \end{aligned}$$

Question A.22 Show that both sides of the above equation are linear. To do this you need to show that if $\alpha = \sum f_i dx^i$ then $d\big(\sum f_i dx^i\big)(v, w) = \sum d\big(f_i dx^i\big)(v, w)$, $v\big[\big(\sum f_i dx^i\big)(w)\big] = \sum v\big[f_i dx^i(w)\big]$, and $\big(\sum f_i dx^i\big)\big([v, w]\big) = \sum \big(f_i dx^i\big)\big([v, w]\big)$.

It is enough to assume we are in a single coordinate patch and so a general one-form is written as $\alpha = \sum f_i dx^i$. Since both sides of the above equation are linear we only need to prove the equation for a single term $f_i dx^i$. But remember that x^i is simply the coordinate function. For simplicity's sake we will prove the equation for the one-form $\alpha = f dg$, where both f and g are functions on the manifold. Thus we have

$$\begin{aligned} d\alpha(v, w) &= d(f dg)(v, w) \\ &= \big(df \wedge dg\big)(v, w) \\ &= df(v)dg(w) - df(w)dg(v) \\ &= v[f] \cdot w[g] - w[f] \cdot v[g] \\ &= (vf)(wg) - (wf)(vg). \end{aligned}$$

Next we look at the term $v\alpha(w)$,

$$\begin{aligned} v\alpha(w) &= v\big(f dg\big)(w) \\ &= v\big(f dg(w)\big) \\ &= v\big(f \cdot \underbrace{w[g]}_{\text{a function}}\ \big) \\ &= v[f] \cdot w[g] + f \cdot v\big[w[g]\big] \qquad \text{Product Rule} \\ &= (vf)(wg) + f \cdot vwg. \end{aligned}$$

Similarly we have $w\alpha(v) = (wf)(vg) + f \cdot wvg$. Since $[v, w] = vw - wv$ we also have

$$\begin{aligned} \alpha\big([v, w]\big) &= (f dg)\big([v, w]\big) \\ &= f \cdot [v, w][g] \\ &= f \cdot \big(vw - wv\big)[g] \\ &= f \cdot vw[g] - f \cdot wv[g] \\ &= f \cdot vwg - f \cdot wvg. \end{aligned}$$

Putting everything together we have

$$\begin{aligned} v\alpha(w) - w\alpha(v) - \alpha\big([v,w]\big) &= (vf)(wg) + f\cdot vwg \\ &\quad - (wf)(vg) - f\cdot wvg \\ &\qquad - f\cdot vwg + f\cdot wvg \\ &= (vf)(wg) - (wf)(vg) \\ &= d\alpha(v,w), \end{aligned}$$

which is the identity we wanted to show. This identity is the base case of the global formula for exterior derivatives. Notice that we could of course also write this using the Lie derivative as

$$d\alpha(v,w) = v\alpha(w) - w\alpha(v) - \alpha\big(\mathscr{L}_v w\big).$$

Of course, if we let $g = x^i$, the coordinate function, and then use the linearity that you proved in the last question we can see that this formula applies for all one-forms $\alpha = \sum f_i dx^i$, and hence we have completed the proof of our base case. Notice that this formula for the exterior derivative of a one-form does not depend at all on which coordinate system we use, thus it is called a coordinate independent formula.

We now turn to proving the global formula for the exterior derivative of a k-form α. The global formula is

Global Exterior Derivative Formula:

$$\begin{aligned} (d\alpha)(v_0,\ldots,v_k) = \sum_{i=0}^{k} &(-1)^i \; v_i\big[\alpha(v_0,\ldots,\widehat{v_i},\ldots,v_k)\big] \\ + \sum_{0\le i<j\le k} &(-1)^{i+j}\; \alpha\big([v_i,v_j],v_0,\ldots,\widehat{v_i},\ldots,\widehat{v_j},\ldots,v_k\big). \end{aligned}$$

Question A.23 Show that for a one-form α the global formula above simplifies to our base case formula.

We have already proved the base case for a one-form. Our induction hypothesis is that this formula is true for all $(k-1)$-forms. We will use the induction hypothesis, along with the global Lie derivative formula of a k-form, to show that this formula is true for a k-form α. Using Cartan's magic formula we have

$$\begin{aligned} d\alpha(v_0,v_1,\ldots,v_k) &= \big(\iota_{v_0} d\alpha\big)(v_1,\ldots,v_k) \\ &= \big(\mathscr{L}_{v_0}\alpha - d\iota_{v_0}\alpha\big)(v_1,\ldots,v_k) \\ &= \big(\mathscr{L}_{v_0}\alpha\big)(v_1,\ldots,v_k) - \big(d\iota_{v_0}\alpha\big)(v_1,\ldots,v_k). \end{aligned}$$

The first term is given by the global Lie derivative formula,

$$\begin{aligned} \big(\mathscr{L}_{v_0}\alpha\big)(v_1,\ldots,v_k) &= v_0\big[\alpha(v_1,\ldots,v_k)\big] - \sum_{i=1}^{k} \alpha(v_1,\ldots,[v_0,v_i],\ldots,v_k) \\ &= v_0\big[\alpha(v_1,\ldots,v_k)\big] - \sum_{i=1}^{k} (-1)^{i-1}\alpha([v_0,v_i],v_1,\ldots,\widehat{v_i},\ldots,v_k) \end{aligned}$$

and the second term is given by the induction hypothesis. That means we assume the global exterior derivative formula is true for the $(k-1)$-form $\iota_{v_0}\alpha$,

$$\begin{aligned} \big(d\iota_{v_0}\alpha\big)(v_1,\ldots,v_k) = \sum_{i=1}^{k} &(-1)^{i-1}\; v_i\big[\iota_{v_0}\alpha(v_1,\ldots,\widehat{v_i},\ldots,v_k)\big] \\ + \sum_{1\le i<j\le k} &(-1)^{i+j-1}\; \iota_{v_0}\alpha\big([v_i,v_j],v_1,\ldots,\widehat{v_i},\ldots,\widehat{v_j},\ldots,v_k\big) \end{aligned}$$

$$
\begin{aligned}
&= \sum_{i=1}^{k} (-1)^{i-1}\, v_i\big[\alpha(v_0, v_1, \ldots, \widehat{v_i}, \ldots, v_k)\big] \\
&\qquad + \sum_{1\le i<j\le k} (-1)^{i+j-1}\, (-1)\alpha\big([v_i, v_j], v_0, v_1, \ldots, \widehat{v_i}, \ldots, \widehat{v_j}, \ldots, v_k\big).
\end{aligned}
$$

Combining we end up with the following mess,

$$
\begin{aligned}
d\alpha(v_0, v_1, \ldots, v_k) &= \big(\mathscr{L}_{v_0}\alpha\big)(v_1, \ldots, v_k) - \big(d\iota_{v_0}\alpha\big)(v_1, \ldots, v_k) \\
&= v_0\big[\alpha(v_1, \ldots, v_k)\big] - \sum_{i=1}^{k} (-1)^{i-1}\, v_i\big[\alpha(v_0, v_1, \ldots, \widehat{v_i}, \ldots, v_k)\big] \\
&\qquad + \sum_{j=1}^{k} (-1)^{0+j}\alpha([v_0, v_i], v_1, \ldots, \widehat{v_i}, \ldots, v_k) \\
&\qquad - \sum_{1\le i<j\le k} (-1)^{i+j}\, \alpha\big([v_i, v_j], v_0, \ldots, \widehat{v_i}, \ldots, \widehat{v_j}, \ldots, v_k\big).
\end{aligned}
$$

The first two terms sum to the first term of the global exterior derivative formula while the third and forth terms sum to the second term of the global exterior derivative formula. Pay attention to the signs and how the dummy indices are used and make sure you understand each step. Thus we have shown the global exterior derivative formula. Sometimes the definition of the exterior derivative is given in terms of this formula. The reason this is sometimes done is that this formula is totally independent of what basis you are using. Notice that the commutator $[v_i, v_j]$ of two vector fields v_i and v_j can be introduced and explained without referring to the Lie derivative.

A.8 Summary and References

A.8.1 Summary

A tensor on a manifold is a multilinear map

$$
\begin{aligned}
T : \underbrace{T^*M \times \cdots \times T^*M}_{\substack{r \\ \text{contravariant} \\ \text{degree}}} \times \underbrace{TM \times \cdots \times TM}_{\substack{s \\ \text{covariant} \\ \text{degree}}} &\longrightarrow \mathbb{R} \\
T\big(\alpha_1, \ldots, \alpha_r, v_1, \ldots, v_s\big) &\longrightarrow \mathbb{R}.
\end{aligned}
$$

That is, the map T eats r one-forms and s vectors and produces a real number.

Covariant tensors eat vectors and covariant tensor components transform like vector basis elements.

Contravariant tensors eat one-forms and contravariant tensor components transform like one-form basis elements.

Let us compare this transformation rule with that of covariant tensors.

Covariant Tensors	Contravariant Tensors
$\widetilde{T}_i = \frac{\partial x^j}{\partial u^i} T_j$	$\widetilde{T}^i = \frac{\partial u^i}{\partial x^j} T^j$
Vector Basis Elements	One-Form Basis Elements
$\frac{\partial}{\partial u^i} = \frac{\partial x^j}{\partial u^i} \frac{\partial}{\partial x^j}$	$du^i = \frac{\partial u^i}{\partial x^j} dx^j$

A general rank (r, w)-tensor $\mathscr{T}$'s components transform according to

$$\boxed{\text{Rank } (r, w)\text{-Tensor Transformation Rule:} \qquad \widetilde{\mathscr{T}}^{k_1 \cdots k_r}_{l_1 \cdots l_s} = \frac{\partial u^{k_1}}{\partial x^{i_1}} \cdots \frac{\partial u^{k_r}}{\partial x^{i_r}} \frac{\partial x^{j_1}}{\partial u^{l_1}} \cdots \frac{\partial x^{j_s}}{\partial u^{l_s}} \mathscr{T}^{i_1 \cdots i_r}_{j_1 \cdots j_s}.}$$

The tensor $\mathscr{T}$ is called anti-symmetric or skew-symmetric if it changes sign whenever any pair of its arguments are switched. That is,

$$\mathscr{T}\big(v_1, \ldots, v_i, \ldots, v_j, \ldots, v_k\big) = -\mathscr{T}\big(v_1, \ldots, v_j, \ldots, v_i, \ldots, v_k\big)$$

when v_i and v_j have switched their places and all other inputs have stayed the same.

$$\boxed{\text{A } k\text{-form is a skew-symmetric rank-}k\text{ covariant tensor.}}$$

Thus we have that the set of k-forms is a subset of the set of $(0, k)$-tensors. Using the tensor definition of differential forms it is easy to prove an extremely important identity, that pullbacks distribute across wedgeproducts,

$$\boxed{\phi^*(\alpha \wedge \beta) = \phi^*\alpha \wedge \phi^*\beta.}$$

A metric on the manifold M is a smooth, symmetric, non-degenerate, rank-two covariant tensor g. Metric tensors are generally denoted with a lower-case g. A manifold that has such a metric on it is called a pseudo-Riemannian manifold. If the metric g also has one additional property, that $g(v, w) \geq 0$ for all vector fields v and w then it is called a Riemannian metric and the manifold is called a Riemannian manifold. Metrics are used to define an inner-product on the manifold, which in term is used to define the lengths of curves on a manifold. The length of the curve γ from p to q is defined to be

$$L(\gamma) = \int_a^b \sqrt{\big|g\big(\dot{\gamma}(t), \dot{\gamma}(t)\big)\big|}\, dt.$$

With this we can then define the distance between the two points p and q as

$$d(p, q) = \inf_{\gamma} L(\gamma),$$

where γ is any piecewise continuous curve that connects p and q. Thus a metric is essential for many of the concepts we take for granted on Euclidian manifolds $\mathbb{R}^n$.

Another concept of differentiation on a manifold called the Lie derivative was introduced. The lie derivative depends on there existing integral curves γ for a vector field v. The Lie derivative can be applied to vector fields, differential forms, functions, and tensors.

$$\boxed{\text{Lie Derivative of Vector Field:} \qquad (\mathscr{L}_v w)_p = \lim_{t \to 0} \frac{T\gamma_{-t} \cdot w_{\gamma_t(p)} - w_p}{t}.}$$

$$\boxed{\text{Lie Derivative of One-Form:} \qquad (\mathscr{L}_v \alpha)_p = \lim_{t \to 0} \frac{T^*\gamma_t \cdot \alpha_{\gamma_t(p)} - \alpha_p}{t},}$$

The Lie derivative of a function was found to be exactly equivalent to directional derivative,

$$\boxed{\text{Lie Derivative of Function:} \qquad (\mathscr{L}_v f)_p = v_p[f].}$$

We will define the pull-back of the tensor $\mathscr{T}$ by γ_t with the following

$$\big(\gamma_t^* \mathscr{T}\big)_p\big(\alpha_1, \ldots, \alpha_r, v_1, \ldots, v_s\big) = \mathscr{T}_{\gamma_t(p)}\big(\gamma_{-t}^* \alpha_1, \ldots, \gamma_{-t}^* \alpha_r, \gamma_{t*} v_1, \ldots, \gamma_{t*} v_s\big).$$

Defining the pull-pack of a tensor this way we define the Lie derivative of the tensor $\mathscr{T}$ with the same formula that we used to define the Lie derivative of a one-form,

$$\boxed{\text{Lie Derivative of Tensor:} \qquad (\mathscr{L}_v \mathscr{T})_p = \lim_{t \to 0} \frac{\gamma_t^* \mathscr{T}_{\gamma_t(p)} - \mathscr{T}_p}{t}.}$$

Notice that since k-forms are a special kind of tensor this definition covers differential k-forms as well.

The following identities describe how the Lie derivative act over addition of tensors, tensor products, and wedgeproducts,

$$\boxed{\mathscr{L}_v\big(a\mathscr{S} + b\mathscr{T}\big) = a\mathscr{L}_v\mathscr{S} + b\mathscr{L}_v\mathscr{T},}$$

$$\boxed{\mathscr{L}_v(\mathscr{S} \otimes \mathscr{T}) = \mathscr{L}_v\mathscr{S} \otimes \mathscr{T} + \mathscr{S} \otimes \mathscr{L}_v\mathscr{T},}$$

$$\boxed{\mathscr{L}_v(\alpha \wedge \beta) = \mathscr{L}_v\alpha \wedge \beta + \alpha \wedge \mathscr{L}_v\beta.}$$

The next identity involves the Lie derivative of the product of functions f and vector fields w,

$$\boxed{\mathscr{L}_v(fw) = f\mathscr{L}_v w + v[f]w.}$$

The exterior derivative of the lie derivative of a differential form α is given by

$$\boxed{d\big(\mathscr{L}_v\alpha\big) = \mathscr{L}_v\big(d\alpha\big).}$$

Next is one of the most important identities. It shows the relationship between the Lie derivative, the exterior derivative, and the interior product,

$$\boxed{\text{Cartan's Homotopy Formula:} \qquad \mathscr{L}_v\alpha = \iota_v(d\alpha) + d(\iota_v\alpha).}$$

This identity is very often simply written as $\mathscr{L}_v = \iota_v d + d\iota_v$. Finally the Lie derivative of a vector field can be written as the commutator of the two vector fields,

$$\boxed{\mathscr{L}_v w = [v, w] = vw - wv.}$$

Global formulas are formulas that do not use coordinates. They are sometimes called coordinate-free formulas. The global formula for the Lie derivative of differential forms is given by

$$\boxed{\begin{array}{l}\text{Global Lie Derivative Formula for differential forms:}\\ \qquad \big(\mathscr{L}_v\alpha\big)(w_1, \ldots, w_k) = v\big[\alpha(w_1, \ldots, w_k)\big] - \displaystyle\sum_{i=1}^{k} \alpha(w_1, \ldots, [v, w_i], \ldots, w_k).\end{array}}$$

The global Lie derivative formula for a general rank (r, s)-tensor $\mathscr{T}$ is given by

Global Lie Derivative Formula for differential forms:

$$\begin{aligned}\big(\mathscr{L}_v \mathscr{T}\big)(\alpha_1, \ldots, \alpha_r, w_1, \ldots, w_s) = v\big[\mathscr{T}(\alpha_1, \ldots, \alpha_r, w_1, \ldots, w_s)\big] \\ - \sum_{i=1}^{r} \mathscr{T}(\alpha_1, \ldots, \mathscr{L}_v \alpha_i, \ldots, \alpha_r, w_1, \ldots, w_s) \\ - \sum_{i=1}^{s} \mathscr{T}(\alpha_1, \ldots, \alpha_r, v_1, \ldots, \mathscr{L}_v w_i, \ldots, w_s).\end{aligned}$$

The global formula for the exterior derivative of a k-form α is given by

Global Exterior Derivative Formula:

$$\begin{aligned}(d\alpha)(v_0, \ldots, v_k) = \sum_{i=0}^{k} (-1)^i \; v_i\big[\alpha(v_0, \ldots, \widehat{v_i}, \ldots, v_k)\big] \\ + \sum_{0 \le i < j \le k} (-1)^{i+j} \; \alpha\big([v_i, v_j], v_0, \ldots, \widehat{v_i}, \ldots, \widehat{v_j}, \ldots, v_k\big).\end{aligned}$$

A.8.2 References and Further Reading

Many books on advanced calculus, manifolds, and physics begin with tensors and then introduce differential forms as a particular kind of tensor. For example, Martin [33] and Bushop and Goldberg [6] take this approach. Several additional very good references for tensors which we relied on are Tu [46], Kay [29], Renteln [37], and Abraham, Marsden, and Ratiu [1]. The very short book by Domingos [15] is also a nice introduction.

Appendix B
Some Applications of Differential Forms

In this appendix we will attempt to broadly show some of the fields of mathematics and physics where differential forms are used. After reading a book this size on differential forms it would be nice for the reader to have some idea of what some of the applications are, other than electromagnetism. None of these sections aim for anything approaching completeness. Instead, they are meant to broaden the reader's horizons, to give just a taste of each subject, and hopefully to whet the reader's appetite for more.

Section one introduces de Rham cohomology in a very general way. In a sense this is one of the "hard core" mathematical applications of differential forms. Basically, differential forms can be used to study the topology of manifolds. In section two we look at some examples of de Rham cohomology groups and see how they can be used to learn about the global topology of the manifold. In section three the idea of a symplectic manifold is introduced, as is the most common symplectic manifold around, the cotangent bundle of any manifold. Symplectic manifolds lead to symplectic geometry. Section four covers the Darbaux theorem, which is a fundamental result in symplectic geometry. And then section five discusses a physics application of symplectic geometry, namely geometric mechanics.

B.1 Introduction to de Rham Cohomology

In this next section we will give a very brief introduction of a field of mathematics called de Rham cohomology. In essence, de Rham cohomology creates a link between differential forms on a manifold M and the global topology of that manifold. This is an incredible and rich area of mathematics and we will not be able to do any more than give an introductory glimpse of it. We will be using differential forms, and in particular the relation between closed and exact forms, to determine global topological information about manifolds. This basically allows us to use what are in essence calculus-like computations to determine global topological properties. But now what do we mean by the words "global" and "topological"? We will not attempt to give rigorous definition but will try to give you a feel for what is meant.

You may roughly think of **topology** as the study of how manifolds, or subsets of $\mathbb{R}^n$, are connected and how many "holes" the manifold or subset has. And what do we mean by global, as opposed to local, properties of a manifold? Recall our intuitive idea of a manifold is something that is locally Euclidian. A little more precisely, an n-dimensional manifold is covered with overlapping patches U_i which are the image of mappings $\phi_i : \mathbb{R}^n \to U_i$. For example, in Fig. B.1 the manifold $M = \mathbb{R}^2 - \{(0,0)\}$, which is the plane with the origin point missing, is shown with four coordinate patches. Every point $p \in M$ is contained in some open neighborhood which is essentially identical to an open neighborhood of $\mathbb{R}^2$.

So the local information of $M = \mathbb{R}^2 - \{(0,0)\}$ is the same as the local information in $\mathbb{R}^2$. But clearly in M we have one point, the origin, missing. Since there is only one point missing you may be tempted to think, using the common usage of the word local, that the missing point is a local property, but it is not. The missing origin is, in fact, a global property; it says something about the whole manifold, namely that the manifold has a hole in it. The fact that M has a hole in it is a global topological property.

We can use certain differential forms defined on a manifold to detect global properties like holes. When it comes to detecting global properties certain forms are interesting and certain forms are not interesting at all. The interesting forms turn out to be those whose exterior derivative is zero, that is, the closed forms.

J. P. Fortney, *A Visual Introduction to Differential Forms and Calculus on Manifolds*,
https://doi.org/10.1007/978-3-319-96992-3

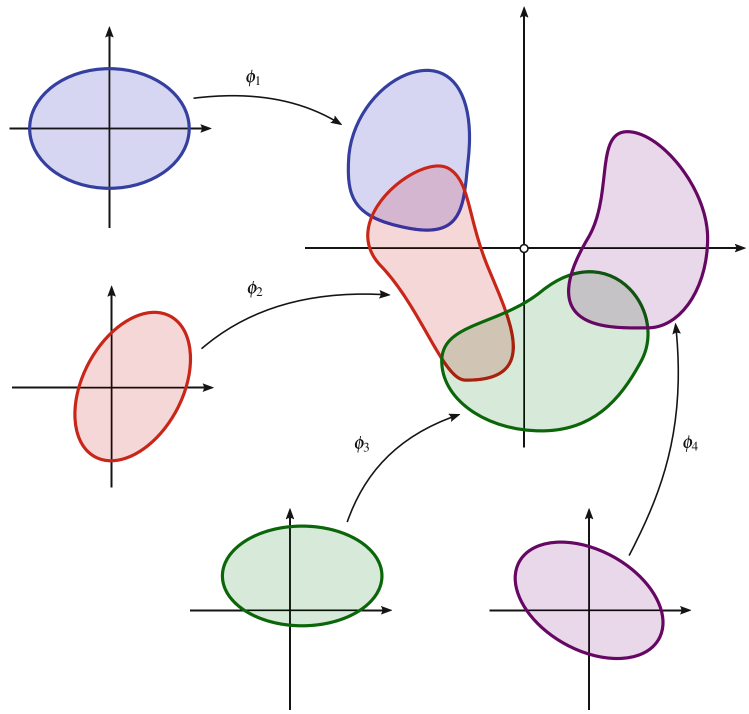


Fig. B.1 The manifold $M = \mathbb{R}^2 - \{(0,0)\}$ shown with four coordinate patches

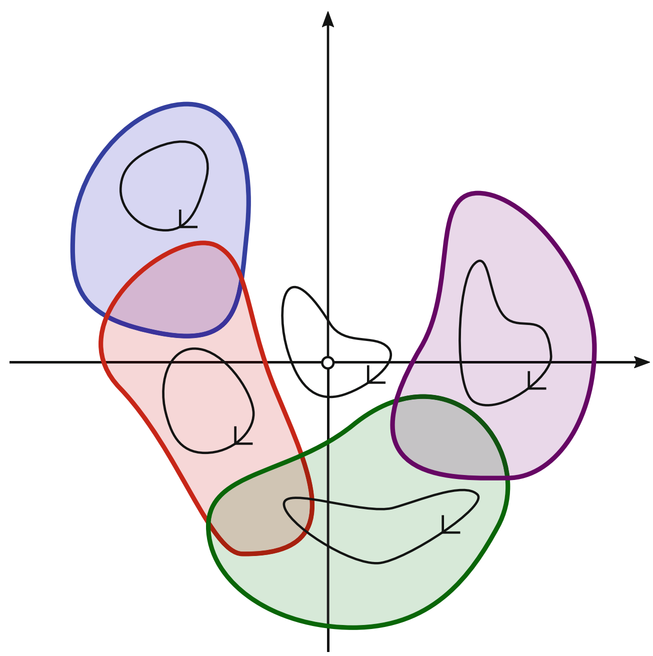


Fig. B.2 The manifold $M = \mathbb{R}^2 - \{(0,0)\}$ is shown with five small closed curves. For of these curves are contained in a single coordinate patch, which the fifth curve encloses the hole at the origin. This curve can not be drawn entirely inside a single coordinate patch

For the moment we will set aside the comment that closed forms give global information and instead try to illustrate to you that closed forms do not give local information. Suppose α is a closed form. Looking locally at M, that is, inside some patch U_i, we integrate α over some small closed curve C that lies entirely inside the patch U_i; that is, over a "local" curve. See Fig. B.2.

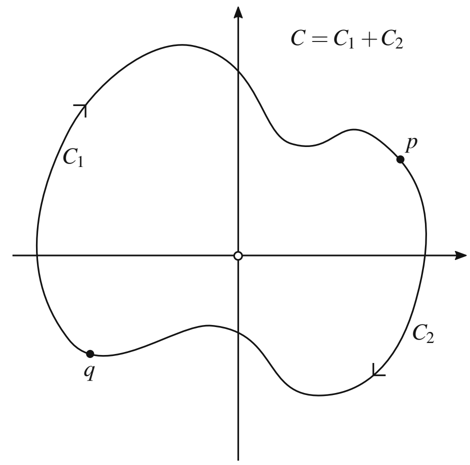


Fig. B.3 The closed curve C broken into two curves C_1 and C_2

Using Stokes' theorem on the local curve C, which is inside a single coordinate patch and thus the boundary of some domain D, we have

$$\int_{C=\partial D} \alpha = \int_D d\alpha = \int_D 0 = 0.$$

Since the integral of α over a local curve is zero this tells us nothing about the manifold locally. So closed forms do not give us any local information about the manifold. Of course by now you may already be guessing how the interesting "global" nature of closed forms arises. What if we integrate a closed α on a curve that goes around the missing point at the origin? First of all, this curve is no longer contained entirely in a single patch U_i. If we had $\int_C \alpha \neq 0$ then we would know that $C \neq \partial D$ for some domain D and that there must be something strange going on inside the curve C. The very fact that there exists a curve C such that $\int_C \alpha \neq 0$ for a closed α tells us some global topological information about M. If C is not the boundary for some region D then C must enclose a hole of some sort.

Now suppose that α were exact, that is, there exists a β such that $\alpha = d\beta$. We will see that exact forms are not interesting. For any closed curve C, whether or not C encloses a hole, by choosing two points p and q on the curve C we can split the curve into two parts, C_1, which goes from q to p and C_2, which goes from p to q, as is shown in Fig. B.3. Since C is now made up of the two curves C_1 and C_2 together. We often write $C = C_1 + C_2$. We then have

$$\begin{aligned}
\int_C \alpha &= \int_{C_1+C_2} \alpha \\
&= \int_{C_1} \alpha + \int_{C_2} \alpha \\
&= \int_{C_1} d\beta + \int_{C_2} d\beta \\
&= \int_{\partial C_1} \beta + \int_{\partial C_2} \beta \\
&= \int_{\{p\}-\{q\}} \beta + \int_{\{q\}-\{p\}} \beta \\
&= \int_{\{p\}-\{q\}} \beta - \int_{\{p\}-\{q\}} \beta \\
&= 0.
\end{aligned}$$

In other words, for an exact form α we have $\int_C \alpha = 0$ whether or not C encloses a hole. In other worlds, exact forms are not interesting, at least when it comes to telling us something about holes.

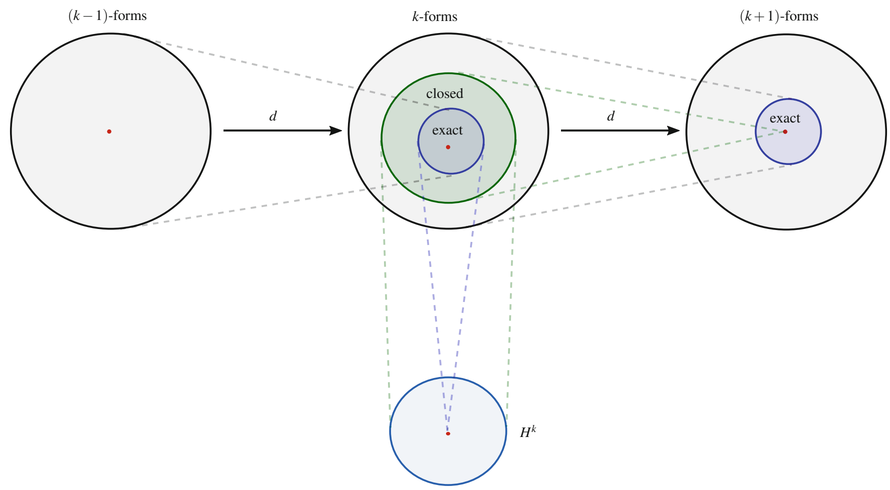


Fig. B.4 We want to measure how different the set of closed k-forms is from the set of exact k-forms. We do this by means of the equivalence class $H^k_{dR}(M)$ (Bachman [4], 2006, p.118)

If we were in the case where our manifold were $\mathbb{R}^n$ then by the Poincaré lemma every closed form is exact. That means that the set of closed forms is the same as the set of exact forms, so then there is no curve C such that for closed α we have $\int_C \alpha \neq 0$. But that should not surprise us too much since we know that there are no holes in $\mathbb{R}^n$.

We are basically interested in the case where the set of closed forms is different than the set of exact forms. We want some way to measure how different these two sets are. In Fig. B.4 we use our Venn-like picture to aid us in visualizing the situation.

What we are interested in is some way to measure how far away from exact a set of closed forms are. We do this by means of what is called an **equivalence relation**. An equivalence relation is simply a way of talking a set of objects and creating another set of objects by defining certain objects in the original set to be "the same." In fact, we have already seen this concept in action when we used equivalence classes of curves to define tangent vectors in Sect. 10.2. There we defined two curves as equivalent, or "the same", if certain criteria were met. Here, two closed k-forms, α and β on a manifold M are considered to be the same, or equivalent, if the difference between them is exact. In other words, we will consider α and β to be equivalent if $\alpha - \beta$ is an exact k-form on M. To be more precise, suppose that α and β are two closed k-forms on M. Then we say that α is equivalent to β, which we write as $\alpha \sim \beta$, if

$$\alpha - \beta = d\omega$$

for some $(k-1)$-form ω on M. We denote the equivalence class of α by

$$[\alpha] = \left\{ \beta \in \bigwedge^k(M) \;\middle|\; \alpha \sim \beta \right\}.$$

When we do this we end up with a set whose elements are the equivalence classes of closed forms that differ by an exact form. This set is denoted $H^k_{dR}(M)$

$$H^k_{dR}(M) = \frac{\left\{ \alpha \in \bigwedge^k(M) \;\middle|\; d\alpha = 0 \right\}}{\left\{ d\omega \in \bigwedge^k(M) \;\middle|\; \omega \in \bigwedge^{k-1}(M) \right\}}.$$

When read out loud the horizonal line in the middle is pronounced modulo, or mod for short. The right hand side would be read "the closed forms mod the exact forms." This $H^k_{dR}(M)$ is called the kth **de Rham cohomology group**. What this vector space $H^k_{dR}(M)$ is actually doing is telling us something about the connectedness and number of holes the manifold M has.

We will not, in this book, explain what the word group means, but if you have not had an abstract algebra class yet then be aware that the word group has a very precise meaning in mathematics and is not to be used lightly or imprecisely. Unfortunately, or perhaps fortunately, for us the Rham cohomology group is actually also a vector space, which was defined in Sect. 1.1. It turns out that a vector space is indeed an abelian group with a scalar multiplication defined on it. That is, a vector space is an abelian group with some additional structure defined on it; the additional structure is scalar multiplication. Scalar multiplication just means that you can multiply the elements of the space by a scalar, which is simply another word for a real number $c \in \mathbb{R}$. So, calling the de Rham cohomology group a group, while not inaccurate, is not as accurate as it could be either. The de Rham cohomology group is a vector space. But as happens in language, once a name gets attached to an object and everyone gets used to that name it is basically impossible to change it, so any attempt to change the terminology to the de Rham cohomology vector space would be pointless.

But why do we care about these de Rham cohomology groups? It turns out that the de Rham cohomology groups on a manifold tell us something about certain global topological properties of the manifold. In the next section we will look at a few examples to see how this works.

B.2 de Rham Cohomology: A Few Simple Examples

Now we will compute a few de Rham cohomology groups and try to explain what they mean. While most of the detailed computations in his section have been left as exercises for you to try to do if you are so inclined, don't let yourself get bogged down in details. Many of these exercises may be quite difficult. And the reality is that de Rham cohomology has a great deal of advanced mathematical machinery that is actually used in doing these sorts of computations that is far beyond the purview of this book. This is only an extremely short introduction to a vast subject. For the moment it is sufficient to simply walk away with a general feel for the fact that computations involving differential forms can actually tell you something interesting about a manifold's global topology.

Suppose that M is a manifold. For convenience' sake we will denote the vector space of closed k-forms on M by

$$Z^k(M) = \left\{ \alpha \in \bigwedge^k(M) \;\middle|\; d\alpha = 0 \right\}$$

and the set of exact k-forms on M by

$$B^k(M) = \left\{ d\omega \in \bigwedge^k(M) \;\middle|\; \omega \in \bigwedge^{k-1}(M) \right\}.$$

This makes it easier to discuss the vector spaces of closed and exact forms. Thus we have

$$H^k_{dR}(M) = Z^k(M)/B^k(M).$$

The Manifold $\mathbb{R}$

We start with the manifold $M = \mathbb{R}$. To find $H^0_{dR}(\mathbb{R})$ we need to find both $Z^0(\mathbb{R})$ and $B^0(\mathbb{R})$. We first turn our attention to $Z^0(\mathbb{R})$, the set of closed zero-forms on $\mathbb{R}$. What are the zero-forms on $\mathbb{R}$? They are just the functions f on $\mathbb{R}$. Thus we have

$$Z^0(\mathbb{R}) = \left\{ f : \mathbb{R} \to \mathbb{R} \;\middle|\; df = 0 \right\}.$$

To find these we take the exterior derivative of f, $df = \frac{\partial f}{\partial x^1} dx^1$ which of course implies that $\frac{\partial f}{\partial x^1} = 0$. For what real-valued functions is this true? Those functions that do not change at all, in other words, the constant functions $f(x^1) = c$ for some constant $c \in \mathbb{R}$. Thus $Z^0(\mathbb{R})$ is simply the set of all constant functions. Since any real number can be a constant we have that $Z^0(\mathbb{R})$ is isomorphic to $\mathbb{R}$, written as $Z^0(\mathbb{R}) \simeq \mathbb{R}$. Isomorphisms of vector spaces were discussed in Sect. 3.1. Now we consider $B^0(\mathbb{R})$, the set of exact 0-forms on $\mathbb{R}$. But since there are no such thing as (-1)-forms then no zero-form can be exact so we say that $B^0(\mathbb{R}) = \{0\}$. This gives

$$H^0_{dR}(\mathbb{R}) = Z^0(\mathbb{R})/B^0(\mathbb{R}) = \mathbb{R}/\{0\} = \mathbb{R}.$$

To find $H^1_{dR}(\mathbb{R})$ we turn our attention to $Z^1(\mathbb{R})$ and $B^1(\mathbb{R})$. But here things become easy because of the Poincaré lemma which says that every closed form on $\mathbb{R}^n$ is exact. This of course means that $Z^1(\mathbb{R}) = B^1(\mathbb{R})$ and hence

$$H^1_{dR}(\mathbb{R}) = \{0\} \equiv 0.$$

In fact, by the Poincaré lemma we have that $Z^k(\mathbb{R}) = B^k(\mathbb{R})$ for all $k > 0$ so we actually have

$$H^k_{dR}(\mathbb{R}) = \{0\} \equiv 0.$$

for every $k > 0$.

The Manifold $\mathbb{R}^n$

Question B.1 For the manifold $\mathbb{R}^n$ show that

$$H^k_{dR}(\mathbb{R}^n) = \begin{cases} \mathbb{R} & \text{if } k = 0, \\ 0 & \text{if } k \geq 1. \end{cases}$$

The Manifold $\mathbb{R} - \{(0)\}$

Now we turn our attention to the manifold $\mathbb{R} - \big\{(0)\big\} = (-\infty, 0) \cup (0, \infty)$, which is the real line $\mathbb{R}$ with the origin removed. We have not spent any time discussing manifolds of this nature, but it is perfectly possible to have a manifold with components that are not connected to each other. Consider $\mathbb{R} - \big\{(0)\big\}$; we say that this manifold has two connected components. Implicit in this is that these two connected components are not connected to each other. In topology there are several different ways to define what **connected** mean, but here we will stick with the simplest and most intuitive idea, that of **path-connectedness**. A space is called path-connected if, for any two points in that space, it is possible to find a path between those two points. Consider the left side of $\mathbb{R} - \big\{(0)\big\}$, which is simply the negative numbers $(-\infty, 0)$. It is possible to find a path in $(-\infty, 0)$ that connects any two negative numbers. Similarly, the right side of $\mathbb{R} - \big\{(0)\big\}$, which is the set of positive numbers $(0, \infty)$, is path-connected as well. However, any potential path that would connect a negative number to a positive number would have to go through the origin, which has been removed. Hence $(-\infty, 0)$ and $(0, \infty)$ are not connected to each other and so we say the manifold $\mathbb{R} - \big\{(0)\big\}$ has two connected components.

Of course it is possible to define both functions and differential forms on a manifold like $\mathbb{R} - \big\{(0)\big\}$ that has multiple connected components. At times it makes sense to write a function f on $\mathbb{R} - \big\{(0)\big\}$ as

$$f = \begin{cases} f(x) & x \in (-\infty, 0), \\ \tilde{f}(x) & x \in (0, \infty). \end{cases}$$

In order to find $Z^0\Big(\mathbb{R} - \big\{(0)\big\}\Big)$ we want to find all closed functions on $\mathbb{R} - \big\{(0)\big\}$. By definition of closed we need to have $df = 0$, which means that

$$0 = df = \begin{cases} \frac{\partial f(x)}{\partial x} dx & x \in (-\infty, 0) \\ \frac{\partial \tilde{f}(x)}{\partial x} dx & x \in (0, \infty), \end{cases}$$

which leads to

$$\frac{\partial f(x)}{\partial x} = 0 \qquad \text{and} \qquad \frac{\partial \tilde{f}(x)}{\partial x} = 0,$$

which only occurs if both $f(x)$ and $\tilde{f}(x)$ are constant functions. But there is no reason to expect that the function f takes on the same constant on each component.

Question B.2 Find df for the following functions,

(i) $f = \begin{cases} 7 & x \in (-\infty, 0) \\ -3 & x \in (0, \infty), \end{cases}$

(ii) $f = \begin{cases} -6 & x \in (-\infty, 0) \\ 5 & x \in (0, \infty), \end{cases}$

(iii) $f = \begin{cases} \frac{14}{3} & x \in (-\infty, 0) \\ -6\pi & x \in (0, \infty). \end{cases}$

Question B.3 Explain why each closed zero-form on $\mathbb{R} - \big\{(0)\big\}$ can be exactly specified by two real numbers, and thus that $Z^0\Big(\mathbb{R} - \big\{(0)\big\}\Big) \simeq \mathbb{R}^2$.

Since no zero-form can be exact we have $B^0(\mathbb{R} - \{(0)\}) = \{0\}$. This means that

$$\begin{aligned} H^0_{dR}\Big(\mathbb{R} - \big\{(0)\big\}\Big) &= \frac{Z^0\Big(\mathbb{R} - \big\{(0)\big\}\Big)}{B^0\Big(\mathbb{R} - \big\{(0)\big\}\Big)} \\ &= \mathbb{R}^2\big/\{0\} \\ &= \mathbb{R}^2. \end{aligned}$$

The Manifold $\mathbb{R} - \{p\} - \{q\}$

Suppose $p, q \in \mathbb{R}$ and $p < q$. Consider the manifold $\mathbb{R}$ with the points p and q removed. As before we can write a function f on $\mathbb{R} - \{p\} - \{q\}$ as

$$f = \begin{cases} f_1(x) & x \in (-\infty, p), \\ f_2(x) & x \in (p, q), \\ f_3(x) & x \in (q, \infty). \end{cases}$$

Finding the closed zero-forms on $\mathbb{R} - \{p\} - \{q\}$ reduces to finding functions f_1, f_2, f_3 such that $\frac{\partial f_1(x)}{\partial x} = 0$, $\frac{\partial f_2(x)}{\partial x} = 0$, and $\frac{\partial f_3(x)}{\partial x} = 0$, which are clearly the constant functions. And of course on each component of $\mathbb{R} - \{p\} - \{q\}$ the constant may be a different value.

Question B.4 Show that $Z^0\big(\mathbb{R} - \{p\} - \{q\}\big) \simeq \mathbb{R}^3$ and hence that $H^0_{dR}\big(\mathbb{R} - \{p\} - \{q\}\big) = \mathbb{R}^3$.

Manifolds with m Connected Components

Question B.5 Let $p_1, p_2, \ldots, p_m \in \mathbb{R}$ be distinct points and let $M = \mathbb{R} - \{p_1\} - \{p_2\} - \cdots - \{p_m\}$. Show that $H^0_{dR}(M) = \mathbb{R}^m$.

Question B.6 Argue that for a general manifold M with m connected components that $H^0_{dR}(M) = \mathbb{R}^m$.

Thus we can see that the zeroth de Rham cohomology group actually tells us how many connected components our manifold has.

Connected n-Dimensional Manifolds

If M is a connected n-dimensional manifold, then by arguments already familiar to us we have

$$H^0_{dR}(M) = \mathbb{R}.$$

Question B.7 Find $Z^k(M)$ and $B^k(M)$ for $k > n$. Show that $H^k_{dR}(M) = 0$.

The Manifold S^1

The manifold S^1 is simply a circle, which we can consider to be the unit circle, see Fig. 2.8. That is, we will consider

$$S^1 = \left\{ (x, y) \;\middle|\; x^2 + y^2 = 1 \right\} \subset \mathbb{R}^2.$$

This manifold is connected, one-dimensional, and has a "hole" in it. Since it is connected and one-dimensional we already know that

$$H^0_{dR}(M) = \mathbb{R},$$

and

$$H_{dR}^{k}(M) = 0 \ \text{ for } k \geq 2.$$

Now we actually want to find what $H_{dR}^{1}(M)$ is. This will take some work. Let us begin by considering the one-form

$$\omega = \frac{-y}{x^2 + y^2}\, dx + \frac{x}{x^2 + y^2}\, dz.$$

It is easy to show that $d\omega = 0$ and hence that ω is closed.

Question B.8 Show that $d\omega = 0$.

Our next goal is to show that ω is not exact. We shall do this by assuming that ω is exact and then reason our way to a contradiction, thereby showing that ω can not be exact. Since ω is a one-form defined on S^1 and we are assuming it is exact then there must be some zero-form f_0 such that $\omega = df_0$. Now suppose that there is some other zero-form f_1 such that $\omega = df_1$ as well. This of course means that $df_1 = df_0$ and hence $f_1 = f_0 + f_2$ for some closed zero-form f_2. But since S^1 is a connected manifold then we know that the closed zero-forms on S^1 are the constants, and hence $f_1 = f_0 + c$ for some constant $c \in \mathbb{R}$. So, if we can find even a single zero-form f_0 such that $df_0 = \omega$ then we automatically know every possible zero-form f_1 such that $df_1 = \omega$.

We begin by simply presenting a candidate function for f_0. Let f_0 be the polar coordinate θ. From trigonometry we have that $\tan\theta = \frac{y}{x}$ for $x \neq 0$ and $\cot\theta = \frac{x}{y}$ for $y \neq 0$.

Question B.9 Using $\tan\theta = \frac{y}{x}$ show that $d\theta = \omega$ when $x \neq 0$. You will need to recall how to take derivatives of trigonometric functions and use the identity $\sec^2\theta = 1 + \tan^2\theta$.

Question B.10 Using $\cot\theta = \frac{x}{y}$ show that $d\theta = \omega$ when $y \neq 0$. You will need to recall how to take derivatives of trigonometric functions and use the identity $\csc^2\theta = 1 + \cot^2\theta$.

Clearly θ indeed works since $d\theta = \omega$. Of course the real problem, and the contradiction, lies with the actual "function" θ.

Question B.11 Using trigonometry argue that θ takes on multiple values at the point $(x, y) = (1, 1)$.

The "function" θ takes on multiple values at any point on S^1 and so it actually is not a function at all since it is not well-defined. But by our reasoning we know that any function f_1 that satisfies $df_1 = \omega$ must be of the form $\theta + c$ hence there is no function that satisfies $df_1 = \omega$ which means that ω must not be exact.

Question B.12 Suppose we had the manifold $\mathbb{R}^2 - \big\{x, 0) \;\big|\; x \leq 0\big\}$, that is, the plane with the non-positive portion of the x-axis removed. Show that the one-form ω is actually exact on $\mathbb{R}^2 - \big\{(x, 0) \;\big|\; x \leq 0\big\}$. Similarly, if S^1 is the unit circle, show that on the manifold $S^1 - \big\{(-1, 0)\big\}$ the one-form ω is exact.

Question B.13 Suppose that α is a closed one-form on S^1. Define $r = \frac{1}{2\pi}\int_{S^1} \alpha$. While a rigorous proof would be difficult, give a plausibility argument that $\psi = \alpha - r\omega$ is an exact one-form on S^1. Use this to then argue that $H_{dR}^{1}(M)$ is a one-dimensional vector space with basis being the equivalence class $[\omega]$.

So for the manifold S^1 we have

$$\begin{aligned}
H_{dR}^{0}(S^1) &= \mathbb{R}, \\
H_{dR}^{1}(S^1) &= \mathbb{R}, \\
H_{dR}^{k}(S^1) &= 0, \quad k > 1.
\end{aligned}$$

Putting It All Together

Now we are ready to put together the ideas from the last section and the examples from this section and try to understand, at least in a very general and somewhat imprecise way, what topological properties of the manifold the de Rham cohomology groups tell us. It is clear that the dimension of the vector space $H_{dR}^{0}(M)$ tells us how many connected components the manifold M has.

Things can start to get interesting when we look at $H^1_{dR}(M)$. For our Euclidian spaces $\mathbb{R}^n$ we had $H^k_{dR}(M) = 0$ for all $k > 0$ but for the unit sphere we had $H^1_{dR}(S^1) = \mathbb{R}$. What is the difference between these manifolds? S^1 has a "hole" in it while the Euclidian spaces do not. The first de Rham cohomology group captures that information. Similarly, the first de Rham cohomology group would capture the "hole" in the manifold $\mathbb{R}^2 - \{(0,0)\}$. But notice, S^1 is a one-dimensional manifold but $\mathbb{R}^2 - \{(0,0)\}$ is a two-dimensional manifold.

Consider the de Rham cohomology groups of some other manifolds, the two-sphere S^2 gives

$$
\begin{aligned}
H^0_{dR}(S^2) &= \mathbb{R}, \\
H^1_{dR}(S^2) &= 0, \\
H^2_{dR}(S^2) &= \mathbb{R}, \\
H^k_{dR}(S^2) &= 0, \quad k > 2,
\end{aligned}
$$

while the n-sphere S^n gives

$$
\begin{aligned}
H^0_{dR}(S^n) &= \mathbb{R}, \\
H^k_{dR}(S^n) &= 0, \quad 0 < k < n, \\
H^n_{dR}(S^n) &= \mathbb{R}, \\
H^\ell_{dR}(S^n) &= 0, \quad \ell > n,
\end{aligned}
$$

and the two-torus T^2 gives

$$
\begin{aligned}
H^0_{dR}(T^2) &= \mathbb{R}, \\
H^1_{dR}(T^2) &= \mathbb{R}^2, \\
H^2_{dR}(T^2) &= \mathbb{R}, \\
H^k_{dR}(T^2) &= 0, \quad k > 2.
\end{aligned}
$$

Question B.14 Explain the de Rham cohomology groups of the n-sphere in terms of the dimension of the n-sphere and the "holes" it has.

Question B.15 Explain the de Rham cohomology groups of the 2-torus in terms of the dimension of the 2-torus and the "holes" it has.

Of course, we have managed to give only the briefest of introductions to de Rham cohomology, and mainly from a very naive perspective at that. Even though the de Rham cohomology group is a vector space defined in terms of closed and exact differential forms, there is a vast amount of mathematical machinery that can be employed in computing these vector spaces. One doesn't actually muck around playing with differential forms as we did in this section for a few of the easiest manifolds. As you could see, even when we got to a manifold as simple as S^1 computing the de Rham cohomology groups this way was becoming difficult and tedious.

B.3 Symplectic Manifolds and the Connonical Symplectic Form

Here we will introduce the concept of a symplectic form. Any manifold which has a symplectic form defined on it is called a symplectic manifold. It turns out that in physics, the Hamiltonian formulation of classical mechanics can formulated abstractly in terms of symplectic manifolds. Hence sometimes you will see symplectic manifolds referred to as Hamiltonian manifolds. In fact, the concept of symplectic manifolds can be generalized to get what are called Poisson manifolds, which can be used to solve a wide class of physics problems. The field that studies mechanics problems in terms of the geometry of manifolds is called geometric mechanics.

It turns out that the cotangent bundle T^*M of a manifold M is itself always a symplectic manifold. That is, T^*M itself is a symplectic manifold. A symplectic form on T^*M always exists and is called the canonical symplectic form. The canonical symplectic form is an extraordinarily interesting, and not so difficult to understand, example of a differential form. Many of the most interesting naturally occurring differential forms in mathematics require more background material to understand then we are able to include in this book, but seeing how the canonical symplectic form arises will give you a taste of how differential forms can naturally arise in mathematics.

A symplectic manifold is a manifold M that has a closed non-degenerate two-form ω defined everywhere on it. We already know that closed means that the exterior derivative of ω is zero; that is, $d\omega = 0$. Now, what does non-degenerate mean?

Let us first consider the situation at a point $p \in M$. At that point the two-form, $\omega_p : T_pM \times T_pM \to \mathbb{R}$, is called **non-degenerate** if the fact that $\omega(v_p, w_p) = 0$ is true for every possible vector v_p means that we must have $w_p = 0$. In other words, if you had $\omega(v_p, w_p) = 0$, no matter what your vector v_p was, then that means that you must have w_p being the zero vector. The two-form is said to be non-degenerate on M if this is true at every point $p \in M$.

Let us consider an example. In a certain sense this example we will look at is the only example there is. Consider a manifold M that has coordinates (on some coordinate patch) given by $(x_1, \ldots, x_n, y_1, \ldots, y_n)$. It is clear from the set of coordinates that M must be even dimensional. Then the two-form

$$\begin{aligned}
\omega &= \sum_{i=1}^{n} dx_i \wedge dy_i \\
&= dx_1 \wedge dy_1 + dx_2 \wedge dy_2 + \cdots + dx_n \wedge dy_n
\end{aligned}$$

is a symplectic form on M. Admittedly, how we have labeled M's coordinates may seem a little strange. After all, what is wrong with $(x_1, \ldots, x_{2n})$? Then we would have simply written $\omega = \sum_{i=1}^{n} dx_i \wedge dx_{n+i}$. But the real reason we have divided up our coordinates into two sets, $x_1, \ldots, x_n$ and $y_1, \ldots, y_n$ has more to do with the kinds of manifolds symplectic forms arise on and the notation that is often used in applications. This will be clearer in a few pages. For now we will just say that these two sets of coordinates are somehow distinct.

First of all, let us check that $\omega = \sum_{i=1}^{n} dx_i \wedge dy_i$ is closed.

$$\begin{aligned}
d\omega &= d\left(\sum_{i=1}^{n} dx_i \wedge dy_i\right) \\
&= \sum_{i=1}^{n} d\left(dx_i \wedge dy_i\right) \\
&= \sum_{i=1}^{n} \left(\sum_{j=1}^{n} \underbrace{\frac{\partial 1}{\partial x_j}}_{=0} dx_j \wedge dx_i \wedge dy_i + \sum_{k=1}^{n} \underbrace{\frac{\partial 1}{\partial y_k}}_{=0} dy_k \wedge dx_i \wedge dy_i \right) \\
&= 0.
\end{aligned}$$

Next we check to see if ω is non-degenerate. We will leave off the point $p \in M$ form our notation for simplicity's sake. Suppose we have the following two vectors

$$v = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_{2n} \end{bmatrix} \quad \text{and} \quad w = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_{2n} \end{bmatrix}.$$

Then we get

$$\begin{aligned}\omega(v, w) &= \sum_{i=1}^{n} (dx_i \wedge dy_i)(v, w) \\ &= \sum_{i=1}^{n} (dx_i(v)dy_i(w) - dx_i(w)dy_i(v)) \\ &= \sum_{i=1}^{n} (a_i b_{n+i} - a_{n+i} b_i).\end{aligned}$$

If $\sum_{i=1}^{n}(a_i b_{n+i} - a_{n+i} b_i) = 0$ for every possible vector v; that is, for all possible values of $a_1, \ldots, a_{2n}$, then we must have that $b_1 = \cdots = b_{2n} = 0$. In other words, the only way we have $\omega(v, w) = 0$ for every single possible choice of v is if w is the zero vector. And of course since this is true at every point $p \in M$ then ω is non-degenerate on M.

Question B.16 Show that $\omega(v, w) = \sum_{i=1}^{n}(a_i b_{n+i} - a_{n+i} b_i) = 0$ for every possible vector v implies that w is the zero vector, that is, that $b_1 = \cdots = b_{2n} = 0$.

Now we will turn our attention to the most common example of a symplectic manifold, the cotangent bundle T^*M of a manifold M. An attempt to draw the manifold T^*M is made in Fig. B.5. Suppose we have a manifold M with coordinates $(x_1, \ldots, x_n)$. For each point on p on the manifold M the cotangent space at p, T_p^*M, is an n-dimensional vector space with the basis $(dx_1)_p, \ldots, (dx_n)_p$.

What we will do is consider the whole cotangent bundle T^*M as our manifold. We could, if we wanted, write $\mathfrak{M} = T^*M$ to make this clear. Though we will not actually do this, it is important to keep in mind that now we are considering the whole cotangent bundle as a manifold and not just the base space M. What are the coordinate functions of this new manifold T^*M? Remember, that in essence the coordinates give an "address" to each point on the manifold. Also remember that we tend to

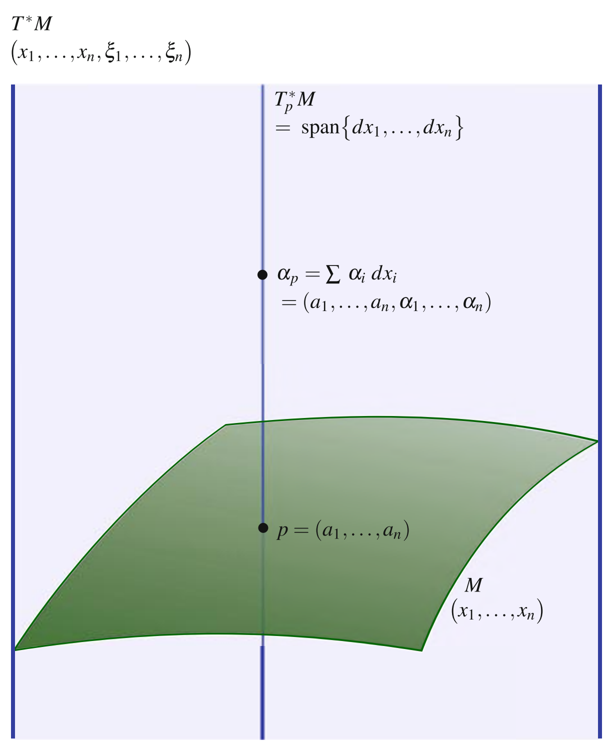


Fig. B.5 The cotangent bundle T^*M. The coordinate functions of the manifold M are given as $(x_1, \ldots, x_n)$ and the coordinate functions of the manifold T^*M are given as $(x_1, \ldots, x_n, \xi_1, \ldots, \xi_n)$. The point $p \in M$ is denoted by $(a_1, \ldots, a_n)$ and the point $\alpha_p \in T^*M$ is denoted by $(a_1, \ldots, a_n, \alpha_1, \ldots, \alpha_n)$. Notice, since the coordinate functions of M are $(x_1, \ldots, x_n)$ the cotangent spaces T_p^*M are given by the span$\{dx_1, \ldots, dx_n\}$

use the word coordinates interchangeably with coordinate functions resulting in equations that look like

$$x_1(p) = x_1, \quad x_2(p) = x_2, \quad \ldots, \quad x_n(p) = x_n$$

where the x_i on the left hand side is a coordinate function and the x_i on the right hand side is a numerical "address" for the point p using a particular coordinate-patch map.

In order to figure out what the coordinates on the manifold T^*M are we need to first ask ourselves what a point in T^*M actually is. But a point in T^*M is just the same thing as an element of T^*M, and we know what those are, differential one-forms at some point p on M. Consider a one-form $\alpha_p = \sum \alpha_i (dx_i)_p \in T_p^*M$ where $x_1(p) = a_1, \ldots, x_n(p) = a_n$ and α_i are either numbers or functions on M. (We will assume in the following discussion that they are numbers. If they are functions then at any given point we can find the numerical value simply by evaluating it at the point, that is, $\alpha_i(p) = \alpha_i$.) Writing $p = (a_1, \ldots, a_n)$ we write the basis elements of T_p^*M as $(dx_i)_p = (dx_i)_{(a_1,\ldots,a_n)}$. Putting this together we can write

$$\alpha_{(a_1,\ldots,a_n)} = \sum_{i=1}^{n} \alpha_i (dx_i)_{(a_1,\ldots,a_n)}.$$

This one-form is completely determined by the values $a_1, \ldots, a_n$ and the values $\alpha_1, \ldots, \alpha_n$. Thus we can define the "address" of the point α_p as $(a_1, \ldots, a_n, \alpha_1, \ldots, \alpha_n)$. Clearly the coordinate functions for the first half of this address are the same as the coordinate functions on M, namely $x_1, \ldots, x_n$. In other words,

$$x_1(\alpha_p) = a_1, \ x_2(\alpha_p) = a_2, \ \ldots, \ x_n(\alpha_p) = a_n.$$

What about the second half? What is the coordinate function that give the values α_i as the output? Sometimes you will see these coordinate functions denoted as dx_i. Thus, we would have

$$dx_1(\alpha_p) = \alpha_1, \ dx_2(\alpha_p) = \alpha_2, \ \ldots, \ dx_n(\alpha_p) = \alpha_n.$$

Of course, now the dx_i are certainly not doing what we have come to expect dx_i to do, which is to eat vectors and spit out numbers. Instead the dx_is are eating one-forms at a point and spitting out the coefficient corresponding to the dx_i in that one-form. So here dx_i are coordinate functions and not differential one-forms. Using this notation we would say the coordinate functions, or the coordinates, on T^*M are $(x_1, \ldots, x_n, dx_1, \ldots, dx_n)$.

As an example, consider the manifold $M = \mathbb{R}^3$ and the one-form

$$\alpha_{(3,2,-1)} = 5dx_1 - 4dx_2 + 6dx_3$$

at the point $(3, 2, -1) \in \mathbb{R}^3$. Then the coordinate functions, or coordinates, of $T^*\mathbb{R}^3$ are $(x_1, x_2, x_3, dx_1, dx_2, dx_3)$ or, if you prefer, (x, y, z, dx, dy, dz). We then have

$$\begin{aligned} x_1(\alpha_{(3,2,-1)}) &= 3, & x_2(\alpha_{(3,2,-1)}) &= 2, & x_3(\alpha_{(3,2,-1)}) &= -1 \\ dx_1(\alpha_{(3,2,-1)}) &= 5, & dx_2(\alpha_{(3,2,-1)}) &= -4, & dx_3(\alpha_{(3,2,-1)}) &= 6. \end{aligned}$$

Hence $\alpha_{(3,2,-1)}$ can be written "in coordinates" as $(3, 2, -1, 5, -4, 6)$.

Given how much we have emphasized differential one-forms as things that eat vectors and spit out numbers it seems very strange to now think of them as coordinate functions on a manifold T^*M that eat points of the manifold, which happen to themselves be one-forms, and spit out real number coefficients. How would we recognize when we are thinking of dx_i as a one-form and when we want to think of it as a coordinate function on the manifold T^*M? One could just say "recognize it from context," which is very often the case, but that seems a little unfair. Instead, what we will do is relabel the coordinate function version of dx_i as ξ_i. That is, we will use $\xi_1 = dx_1, \ldots, \xi_n = dx_n$ only when we are thinking of $dx_1, \ldots, dx_n$ as coordinate functions on T^*M. So, for our example we would have

$$\begin{aligned} x_1(\alpha_{(3,2,-1)}) &= 3, & x_2(\alpha_{(3,2,-1)}) &= 2, & x_3(\alpha_{(3,2,-1)}) &= -1 \\ \xi_1(\alpha_{(3,2,-1)}) &= 5, & \xi_2(\alpha_{(3,2,-1)}) &= -4, & \xi_3(\alpha_{(3,2,-1)}) &= 6. \end{aligned}$$

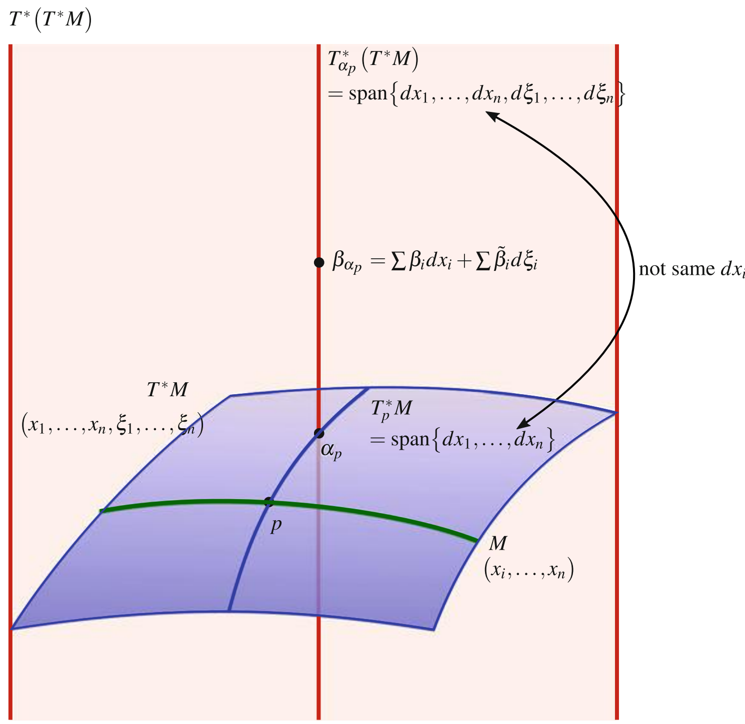


Fig. B.6 The cotangent bundle $T^*(T^*M)$. The coordinate functions of the manifold M are given by $(x_1, \ldots, x_n)$ and the coordinate functions of the manifold T^*M are given by $(x_1, \ldots, x_n, \xi_1, \ldots, \xi_i)$. The coordinate functions of the manifold $T^*(T^*M)$ are not given. $T_p^*M = \text{span}\{dx_1, \ldots, dx_n\}$ and $T^*_{\alpha_p}(T^*M) = \text{span}\{dx_1, \ldots, dx_n, d\xi_1, \ldots, d\xi_n\}$, but be careful, the dx_i for these two spaces are not the same

We attempt to summarize all of this in Fig. B.5.

Now, just as we defined one-forms on the manifold M we can define one-forms on the manifold T^*M. A one-form on the manifold T^*M is actually an element of the cotangent bundle of T^*M, which we will denote by $T^*(T^*M)$. An attempt to draw the cotangent bundle of T^*M is made in Fig. B.6. Since our coordinate functions on T^*M are now being written as $(x_1, \ldots, x_n, \xi_1, \ldots, \xi_n)$ our one-form basis elements for the manifold T^*M are $dx_1, \ldots, dx_n, d\xi_1, \ldots, d\xi_n$. These are our basis elements for the space $T^*(T^*M)$. At a particular point $\alpha_p \in T^*M$ we have

$$(dx_1)_{\alpha_p}, \ldots, (dx_n)_{\alpha_p}, (d\xi_1)_{\alpha_p}, \ldots, (d\xi_n)_{\alpha_p} \in T^*_{\alpha_p}(T^*M).$$

Notice that the dx_i that are one-forms on the manifold T^*M look just like the dx_i that are one-forms on the manifold M.

We will go ahead and say that you need to learn to recognize from the context when dx_i is a one-form on the manifold T^*M and when it is a one-form on the manifold M. Otherwise our notation would get overwhelmingly complicated. This is also another reason we switched notation to ξ_i when we wanted to think of dx_i as a coordinate functions on the manifold T^*M. Having the same notation, dx_i, potentially mean three different things is just too much.

Now that we have started to get a handle on the idea of the cotangent bundle being a manifold in its own right, and have introduced the idea of the cotangent bundle of the cotangent bundle, let's consider what a one-form on T^*M, that is, an element of $T^*(T^*M)$, would be like. It would have the general form

$$\beta = \sum_{i=1}^{n} \beta_i dx_i + \sum_{i=1}^{n} \widetilde{\beta}_i d\xi_i,$$

where β_i and $\widetilde{\beta}_i$, for $i = 1, \ldots, n$, are either numbers or functions on T^*M, $\beta_i : T^*M \to \mathbb{R}$ and $\widetilde{\beta}_i : T^*M \to \mathbb{R}$. But, this is exactly what are coordinate functions x_i and ξ_i are, functions on T^*M. That is, $x_i : T^*M \to \mathbb{R}$ and $\xi_i : T^*M \to \mathbb{R}$. If we let $\beta_i = \xi_i$ and $\widetilde{\beta}_i = 0$ then we get a very special one-form on T^*M called the **Liouville one-form** which we will denote by an α,

$$\boxed{\text{Liouville One-Form:} \qquad \alpha = \sum_{i=1}^{n} \xi_i dx_i.}$$

It is important to keep everything straight in your head, the $\xi_i : T^*M \to \mathbb{R}$ is a coordinate function on the manifold T^*M and the dx_i is a one-form on the manifold T^*M, which means that $dx_i \in T^*\big(T^*M\big)$. The Liouville one-form is also known as the **tautological one-form** or the **canonical one-form**. This one-form always exists on the manifold T^*M.

The Liouville one-form can be used to come up with a two-form on the manifold T^*M. We do this by taking the exterior derivative of the Liouville one-form,

$$
\begin{aligned}
d\alpha &= \sum_{i=1}^{n} d\xi_i \wedge dx_i \\
&= -\sum_{i=1}^{n} dx_i \wedge d\xi_i .
\end{aligned}
$$

Notice that now dx_i and $d\xi_i$ are both one-forms on the manifold T^*M, that is, $dx_i, d\xi_i \in T^*\big(T^*M\big)$. Notice that this is exactly a symplectic form on the manifold T^*M. We define the **canonical symplectic form** on T^*M as

$$
\boxed{\text{Canonical Symplectic Form:} \qquad \omega \equiv -d\alpha = \textstyle\sum_{i=1}^{n} dx_i \wedge d\xi_i}
$$

where α is the Liouville one-form on T^*M. The only real reason the negative sign is there is so the order of our two-form elements is in the same order as we like to write the coordinates.

There is another way that the canonical symplectic form is very often introduced. Suppose we have the projection mapping π that projects the cotangent bundle to the underlying manifold.

$$
\begin{array}{c}
T^*M \\
\big\downarrow \pi \\
M
\end{array}
$$

This mapping takes any element $\xi_p \in T^*M$, which is a one-form at some point $p \in M$, and projects it to the base point, $\pi(\xi_p) = p$. As we have discussed already, just as a manifold M has a tangent space TM, the manifold T^*M has a tangent space $T\big(T^*M\big)$. Suppose v_{ξ_p} is a vector on the manifold T^*M with the base point $\xi_p \in T^*M$. As with all mappings between manifolds, the projection mapping $\pi : T^*M \to M$ induces a push-forward mapping between tangent bundles

$$
T\big(T^*M\big) \xrightarrow{\ \pi_*\ } TM
$$

that pushes forward the vector $v_{\xi_p} \in T_{\xi_p}\big(T^*M\big)$ to a vector in TM at the point p, that is, $\pi_*(v_{\xi_p}) \in T_pM$.

Just as we defined the projection $\pi : T^*M \to M$, we can define the projection

$$
\begin{array}{c}
T\big(T^*M\big) \\
\big\downarrow \Pi \\
T^*M
\end{array}
$$

by $\Pi(v_{\xi_p}) = \xi_p$. You should be aware that often all projection maps are simply denoted by π and you need to figure out from context which space is being projected to which manifold. However, here we differentiate between the two projections we are using; $pi : T^*M \to M$ and $\Pi : T(T^*M) \to T^*M$.

Similarly, the manifold T^*M also has a cotangent bundle $T^*\big(T^*M\big)$ whose elements are the one-forms on T^*M that eat vectors $v_{\xi_p} \in T\big(T^*M\big)$. We define a one-form α in $T^*\big(T^*M\big)$ by

$$
\boxed{\begin{aligned}
\text{Liouville One-Form:} \qquad \alpha(v_{\xi_p}) &\equiv \Pi(v_{\xi_p})\big(\pi_*(v_{\xi_p})\big) \\
&= \Big\langle \Pi(v_{\xi_p}), \pi_*(v_{\xi_p}) \Big\rangle,
\end{aligned}}
$$

$$
\begin{tikzcd}
 & & v_{\xi_p} \in T\left(T^*M\right) \arrow[dll, "\Pi"'] \arrow[drr, "\pi_*"] & & \\
\Pi(v_{\xi_p}) \in T^*M \arrow[rr, Rightarrow] \arrow[drr, "\pi"'] & & \alpha(v_{\xi_p}) \equiv \left\langle \Pi(v_{\xi_p}), \pi_*(v_{\xi_p}) \right\rangle & & \pi_*(v_{\xi_p}) \in TM \arrow[ll, Rightarrow] \\
 & & \pi(\xi_p) \in M & &
\end{tikzcd}
$$

Fig. B.7 The Liouville one-form α is defined intrinsically in terms of the vector v_{ξ_p}. The projection mapping $\pi : T^*M \to M$, defined by $\pi(\xi_p) = p$, induces the push-forward mapping $\pi_* : T(T^*M) \to TM$. The mapping Π is also a projection mapping $\Pi : T(T^*M) \to T^*M$ defined by $\Pi(v_{\xi_p}) = \xi_p$. The Liouville one-form α eating the vector v_{ξ_p} is defined by the canonical pairing of the one-form $\Pi(v_{\xi_p})$ with the vector $\pi_*(v_{\xi_p})$. Notice that at no point in this definition were the actual manifold coordinates employed. The canonical symplectic form is then simply the negative of the exterior derivative of the Liouville one-form, $\omega = -d\alpha$

where $\langle \cdot, \cdot \rangle$ is simply another way to write the canonical pairing between one-forms and vectors. This α is none other than the Liouville one-form. We illustrate this definition of the Liouville one-form in Fig. B.7. The canonical symplectic form is simply the negative of the exterior derivative of the Liouville one-form, $\omega = -d\alpha$.

But what an odd-looking definition - basically is seems like everything is being defined in terms of one vector, the vector $v_{\xi_p} \in T\big(T^*M\big)$. But notice something interesting, at no point anywhere in this definition did we use the actual coordinates from any of our manifolds. We have been able to define the Liouville one-form intrinsically, that is, without any reference to the manifold coordinates. We have mentioned this before. Mathematicians like intrinsic, coordinate independent, definitions because these definitions often make it easier to see the underlying mathematical structure and also often make it easier to prove things. However, when it comes down to actual computations, one usually needs formulas that involve coordinates.

Suppose we had coordinates on M of $(x_1, \dots, x_n)$ and we had the one-form $\xi_p = \sum_{i=1}^{n} \xi_i dx_i \in T^*M$ giving us coordinates on T^*M of $(x_1, \dots, x_n, \xi_1, \dots, \xi_n)$. We often abbreviate these coordinates to (x_i) and (x_i, ξ_i) respectively. Using these coordinates the projection mapping is $\pi(x_1, \dots, x_n, \xi_1, \dots, \xi_n) = (x_1, \dots, x_n)$ or $\pi(x_i, \xi_i) = x_i$.

Question B.17 Using the coordinates described above, show that the definition of the Liouville one-form as $\alpha(v_{\xi_p}) \equiv \Pi(v_{\xi_p})\big(\pi_*(v_{\xi_p})\big)$ gives the formula $\alpha = \sum_{i=1}^{n} \xi_i dx_i$ and hence that $-d\alpha$ gives the formula for the canonical symplectic form.

B.4 The Darboux Theorem

This is a book on differential forms, not manifold theory, so we have only been considering "nice" manifolds, that is, differentiable manifolds. A differentiable manifold M is a manifold that is locally similar enough to $\mathbb{R}^n$ that we can do calculus on it. That means the manifold has an atlas with coordinate charts (U_i, ϕ_i), where $\phi_i : U_i \subset \mathbb{R}^n \to M$ is smooth, ϕ_i^{-1} exists, and all $\phi_i^{-1} \circ \phi_j$ are differentiable mappings.

Symplectic manifolds are a special case of differentiable manifolds. We have already said that a symplectic manifold M has a symplectic form defined everywhere on it, where a symplectic form is a closed non-degenerate two-form. We have also looked at an example of a symplectic form, $\sum_{i=1}^{n} dx_i \wedge dy_i$, where $(x_1, \dots, x_n, y_1, \dots, y_n)$ are the coordinate functions. Since this form is clearly defined everywhere on $\mathbb{R}^{2n}$ we can see that $\mathbb{R}^{2n}$ is always a symplectic manifold.

Darboux's theorem states that in a neighborhood around any point on a symplectic manifold (M, ω) it is possible to choose coordinates $(x_1, \dots, x_n, y_1, \dots, y_n)$ such that we can write ω as

$$\omega = \sum_{i=1}^{n} dx_i \wedge dy_i.$$

In essence, what Darboux's theorem is implying is that all symplectic forms ω are essentially the same. This is nice since it means it is always possible to choose coordinates that allow us to write the symplectic form as $\sum_{i=1}^{n} dx_i \wedge dy_i$. To prove the Darboux theorem we will use what is often called **Moser's trick** or **Moser's lemma**. Moser's lemma has several versions, but we will stick to the simplest version which serves our purpose here.

Lemma B.1 (Moser's Lemma) *Let* $\{\omega_t\}$ *for* $0 \le t \le 1$ *be a family of symplectic forms on the manifold* M*. Then for every point* $p \in M$ *there exists a neighborhood* U *of* p *and a function* $g_t : U \to U$ *such that* g_0^* *is the identity function and* $g_t^* \omega_t = \omega_0$ *for all* $0 \le t \le 1$.

Our goal is of course to find the mappings $g_t : U \to U$. But consider Fig. A.1, which shows the integral curves γ_t of a vector field v. Another way of saying this is that

$$\frac{d}{dt}\gamma_t(p) = v\big(\gamma_t(p)\big).$$

So if we know what the vector field v is we can integrate v to get the integral curves γ_t, which are a function of time. Furthermore, at each moment of time γ_t is a map from the manifold to itself, that is, $\gamma_t : M \to M$. This is essentially the strategy used in Moser's lemma; a vector field v is found which then gives integral curves $g_t : U \to U$ that satisfy $g_0^* = id$ and $g_t^* \omega_t = \omega_0$. Finding this vector field requires exactly the two properties held by the symplectic forms ω_t, closedness and non-degeneracy.

The very first thing we will show is something that will be needed in a few moments. Since ω_t is a symplectic form then we know that it is closed, $d\omega_t = 0$. We want to show that $\frac{d}{dt}\omega_t$ is closed as well. We don't know what form ω_t takes but at least we know it is a two-form so we know we can write it in-coordinates as

$$\omega_t = \sum_{i<j} f_{i,j}(t)\, dx_i \wedge dx_j.$$

So we have

$$\begin{aligned}
d\left(\frac{d}{dt}\omega_t\right) &= d\left(\sum_{i<j} \frac{d}{dt} f_{i,j}(t)\, dx_i \wedge dx_j\right) \\
&= \sum_{i<j} \frac{d}{dt} f_{i,j}(t)\, d\left(dx_i \wedge dx_j\right) \\
&= \frac{d}{dt}\left(\sum_{i<j} f_{i,j}(t)\, d\left(dx_i \wedge dx_j\right)\right) \\
&= \frac{d}{dt}\, d\left(\sum_{i<j} f_{i,j}(t)\, dx_i \wedge dx_j\right) \\
&= \frac{d}{dt}\, d\omega_t \\
&= 0
\end{aligned}$$

Using the Poincaré lemma we know that $\frac{d}{dt}\omega_t$ is exact and so there exists a one-form λ_t such that

$$\frac{d}{dt}\omega_t = d\lambda_t.$$

The next thing we do is take the derivative with respect to time of $g_t^*\omega_t$. This gives us

$$\begin{aligned}
\frac{d}{dt}\, g_t^*\omega_t &= \lim_{h\to 0} \frac{g_{t+h}^*\omega_{t+h} - g_t^*\omega_t}{h} \\
&= \lim_{h\to 0} \frac{g_{t+h}^*\omega_{t+h} - g_t^*\omega_{t+h} + g_t^*\omega_{t+h} - g_t^*\omega_t}{h} \\
&= \lim_{h\to 0} \frac{g_{t+h}^*\omega_{t+h} - g_t^*\omega_{t+h}}{h} + \lim_{h\to 0} \frac{g_t^*\omega_{t+h} - g_t^*\omega_t}{h}
\end{aligned}$$

$$
\begin{aligned}
&= g_t^* \left(\lim_{h \to 0} \frac{g_h^* \omega_{t+h} - \omega_{t+h}}{h} \right) + g_t^* \left(\lim_{h \to 0} \frac{\omega_{t+h} - \omega_t}{h} \right) \\
&= g_t^* \Big(\mathscr{L}_{v_t} \omega_t \Big) + g_t^* \left(\frac{d\omega_t}{dt} \right) \\
&= g_t^* \left(\mathscr{L}_{v_t} \omega_t + \frac{d\omega_t}{dt} \right) \\
&= g_t^* \Big(d\iota_{v_t} \omega_t + d\lambda_t \Big) \\
&= g_t^* \Big(d\big(\iota_{v_t} \omega_t + \lambda_t \big) \Big).
\end{aligned}
$$

There are a couple minor details in the above calculation that should be addressed. First, that

$$
\mathscr{L}_{v_t} \omega_t = \lim_{h \to 0} \frac{g_h^* \omega_{t+h} - \omega_{t+h}}{h}
$$

may look a little different than the definition of the Lie derivative given in Sect. A.7, but here we are not keeping track of the base point. Second, we used Cartan's homotopy formula and the fact that ω_t is closed to get

$$
\mathscr{L}_{v_t} \omega_t = \iota_{v_t} \underbrace{(d\omega_t)}_{=0} + d(\iota_{v_t} \omega_t) = d\iota_v \omega_t .
$$

We want our functions g_t to satisfy the equality $g_t^* \omega_t = \omega_0$. Taking the derivative with respect to time we find that this is equivalent to

$$
\frac{d}{dt}\, g_t^* \omega_t = \frac{d}{dt}\, \omega_0 = 0.
$$

Substituting in the identity we just found, this in turn is equivalent to

$$
g_t^* \Big(d\big(\iota_{v_t} \omega_t + \lambda_t \big) \Big) = 0,
$$

which in turn is true if

$$
\iota_{v_t} \omega_t + \lambda_t = 0.
$$

This is sometimes called Moser's equation. If we can solve this equation for v then we can find the integral curves of v to obtain the mappings g_t that satisfy the required $g_t^* \omega_t = \omega_0$. Note, if this is satisfied the g_0 is the identity function trivially.

So now our goal becomes to solve the equation $\iota_v \omega_t + \lambda_t = 0$. Suppose that M is a $2n$-dimensional manifold with local coordinates $(x_1, \dots, x_{2n})$. Letting $\frac{\partial}{\partial x_k}$ be the basis for TM and dx_k be the basis for T^*M we can write

$$
\begin{aligned}
\lambda_k &= \sum_{i=1}^{2n} \lambda_i(t, x)\, dx_k, \\
v_t &= \sum_{k=1}^{2n} v_k(t, x)\, \frac{\partial}{\partial x_k}, \\
\omega_t &= \sum_{\substack{k,\ell=1 \\ k<\ell}}^{2n} \omega_{k,\ell}(t, x)\, dx_k \wedge dx_\ell .
\end{aligned}
$$

Using this we can write

$$\iota_{v_t}\omega_t = 2\sum_{\ell=1}^{2n}\left(\sum_{k=1}^{2n}\omega_{k,\ell}(t,x)v_k(t,x)\right)dx_\ell$$

and so write $\iota_{v_t}\omega_t + \lambda_t = 0$ as

$$2\sum_{\ell=1}^{2n}\left(\sum_{k=1}^{2n}\omega_{k,\ell}(t,x)v_k(t,x)\right)dx_\ell + \sum_{\ell=1}^{2n}\lambda_\ell(t,x)\,dx_\ell = 0.$$

This actually gives us a system of $2n$ equations

$$\lambda_\ell(t,x) + 2\sum_{k=1}^{2n}\omega_{k,\ell}(t,x)v_k(t,x) = 0,$$

which of course can be written as

$$\begin{aligned}
\lambda_1(t,x) + 2\sum_{k=1}^{2n}\omega_{k,1}(t,x)v_k(t,x) &= 0,\\
\lambda_2(t,x) + 2\sum_{k=1}^{2n}\omega_{k,2}(t,x)v_k(t,x) &= 0,\\
&\vdots\\
\lambda_{2n}(t,x) + 2\sum_{k=1}^{2n}\omega_{k,2n}(t,x)v_k(t,x) &= 0,
\end{aligned}$$

which in turn can be written in matrix form as

$$\begin{bmatrix}\lambda_1\\ \vdots\\ \lambda_{2n}\end{bmatrix} + 2\underbrace{\begin{bmatrix}\omega_{1,1} & \cdots & \omega_{2n,1}\\ \vdots & \ddots & \vdots\\ \omega_{1,2n} & \cdots & \omega_{2n,2n}\end{bmatrix}}_{\Omega}\begin{bmatrix}v_1\\ \vdots\\ v_{2n}\end{bmatrix} = 0.$$

or more succinctly as

$$\boldsymbol{\lambda} + 2\Omega\mathbf{v} = 0.$$

Since the ω_t is a symplectic form for $0 \le t \le 1$ it is non-degenerate, which implies the matrix Ω is non-singular and hence can be inverted. If you have not had a linear algebra course this may be unclear to you, but we will not try explain this detail here and will instead refer you to a linear algebra textbook. But this fact allows us to solve for **v**,

$$\mathbf{v} = -\frac{1}{2}\Omega^{-1}\boldsymbol{\lambda}.$$

Thus we have found a unique vector field v_t as a solution to the Moser equation $\iota_{v_t}\omega_t + \lambda_t = 0$. This vector field v_t can then be integrated to give the integral curves g_t which are in fact the mappings we needed to find, thereby proving the Moser lemma. With the Moser lemma now in hand the proof of the Darboux theorem is almost trivial.

Theorem B.1 (Darboux Theorem) *If (M, ω) is a symplectic manifold of dimension $2n$ then in a neighborhood U of each point $p \in M$ there exist local coordinates $(x_1, \ldots, x_n, y_1, \ldots, y_n)$ such that the symplectic form ω can be written as*

$$\omega = \sum_{i=1}^{n} dx_i \wedge dy_i.$$

Given a symplectic form ω, then at any point $p \in M$ it is possible to find local coordinates $(x'_1, \ldots, x'_n, y'_1, \ldots, y'_n)$ such that

$$\omega\big|_p = \sum_{i=1}^{n} dx'_i \wedge dy'_i\big|_p.$$

This is fairly straight-forward to see. Any coordinate patch will have a symplectic form of the form $\sum_{i=1}^{n} dx'_i \wedge dy'_i$ on it. Just scale the coordinates until equality is obtained at the point p. Of course at this point there is no reason to expect that ω will equal $\sum_{i=1}^{n} dx'_i \wedge dy'_i$ on U not at p.

So now we have two symplectic forms on U, $\omega_0 = \omega$ and $\omega_1 = \sum_{i=1}^{n} dx'_i \wedge dy'_i$. We can use this to define the family $\{\omega_t\}$ of forms

$$\omega_t = \omega_0 + t(\omega_1 - \omega_0).$$

Of course since both ω_0 and ω_t are symplectic forms it would be natural to assume that ω_t is for each t. But this is something that has to be checked. Since both ω_0 and ω_1 are symplectic forms they are closed, so it is easy to see that ω_t is closed too,

$$d\omega_t = \underbrace{d\omega_0}_{=0} + t\big(\underbrace{d\omega_1}_{=0} - \underbrace{d\omega_0}_{=0}\big) = 0.$$

Showing that ω_t is non-degenerate requires a little more work and this is the reason we took such pains to find ω_1 such that $\omega_1(p) = \omega_0(p)$. This means that $\omega_t(p)$ is actually independent of t. In other words, we have that $\omega_t(p) = \omega_0(p) = \omega_1(p)$ for $0 \le t \le 1$. Since we defined $\omega_t = \omega_0 + t(\omega_1 - \omega_0)$ we see that ω_t is a continuous function of ω_0 and ω_1. In other words, ω_t changes continuously from ω_0 to ω_1 as t goes from 0 to 1. Since both ω_0 and ω_1 are non-degenerate in a neighborhood of p then *by continuity* of ω_t, ω_t is non-degenerate in some neighborhood of p for all t between 0 and 1 and is therefore a symplectic form for $0 \le t \le 1$. The phrase "by continuity" may seem strange to you. This basically refers to the properties that continuous functions have, something that is typically studied in an analysis or functional analysis course. Again, we will not try to explain this in any more detail here and will instead refer you to an analysis textbook.

So ω_t is both closed and non-degenerate for each $t \in [0, 1]$ and hence $\{\omega_t\}$ is a family of symplectic forms so we can apply the Moser lemma. According to the Moser lemma in some neighborhood U around the point $p \in M$ there exists a map $g_t : U \to U$ such that $g_t \omega_t = \omega_0$. In particular we have $g_1 \omega_1 = \omega_0$. In other words,

$$\begin{aligned}\omega &= g_1^* \left(\sum_{i=1}^{n} dx'_i \wedge dy'_i \right) \\ &= \sum_{i=1}^{n} d(x'_i \circ g_1) \wedge d(y'_i \circ g_1).\end{aligned}$$

We now choose the new coordinates $x_i = x'_i \circ g_1$ and $y_i = y'_i \circ g_1$ thereby proving the Darboux theorem.

B.5 A Taste of Geometric Mechanics

In this section we give a very brief introduction to geometric mechanics. We will show that it is possible to write the Hamiltonian equations of a simple mechanical system in terms of symplectic forms. We will not go into very much detail, and so we will simply state, and not show, a very important correspondence. The purpose of this section is to see where and how the canonical symplectic two-form on the cotangent bundle is used. You may not follow every detail, but that is okay, for now simply grasping the big picture is enough.

In physics and mechanics the manifolds we are actually working with come from the particular physical problem we are considering. For example, if we are considering a very simple physical system of a point particle under the influence of Earth's gravity the manifold we are concerned with comes from the x, y, and z-coordinates of the particle. So our manifold is $M = \mathbb{R}^3$ with coordinates (x, y, z). If we are considering a system of n particles, far away from the Earth, all under the influence of gravity between each other then we are concerned with the x, y, and z-coordinates of each of these particles, so our manifold is $M = \mathbb{R}^{3n}$ with coordinates $(x_1, y_1, z_1, x_2, y_2, z_2, \ldots, x_n, y_n, z_n)$.

In physics and mechanics the manifold that comes from a physical system is called the **configuration space** of the physical system. Of course, as you probably already know from your physics classes, if you choose the right coordinate system you can simplify your manifold considerably. We will not worry about that here except to say that when we are able to make simplifications in our manifold based on symmetries or other properties of the actual physical system under consideration, the manifold is still called the configuration space, but our coordinates are often called **generalized coordinates**. When we are considering generalized coordinates, or for that matter any coordinates for a configuration space, the notational convention is to use the variable q instead of the variable x. Thus we will say the manifold has coordinate (or coordinate functions) $(q_1, q_2, \ldots, q_n)$. If you see q used as a manifold's coordinate functions instead of x you can be fairly confident that the book or paper is written from the point of view of physics or mechanics, and the techniques used are probably useful for physical or mechanical systems.

Now we turn our attention to the ideas of **flows**. Flows are essentially same things as integral curves, see for example Fig. 2.4, that appear in the context of physics or mechanics problems. Suppose you have a manifold M and it has a flow "on it." You can imagine a flow in a manifold as a fluid that is flowing through the manifold. Each particle of fluid determines a trajectory on M. The flow on the manifold is depicted as the collection of trajectories of particles of fluid on M, see Fig. B.8

As the fluid flows, at each moment of time each particle of fluid has a velocity, and hence a velocity vector. For example, at time t_0 all the particles at all the points in M have a velocity. This collection of velocity vectors at the time t_0 gives us a velocity vector field v_{t_0} on M; that is, v_{t_0} is a section in TM. Similarly, at a later time t_1 the particles flowing in M produce a velocity vector field v_{t_1}, which is of course also a section in TM. See Fig. B.9.

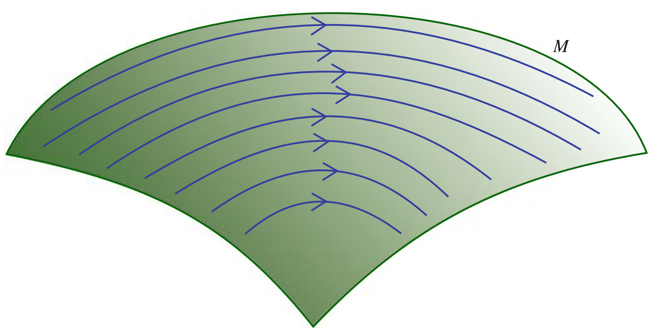


Fig. B.8 The manifold M shown with several flow trajectories

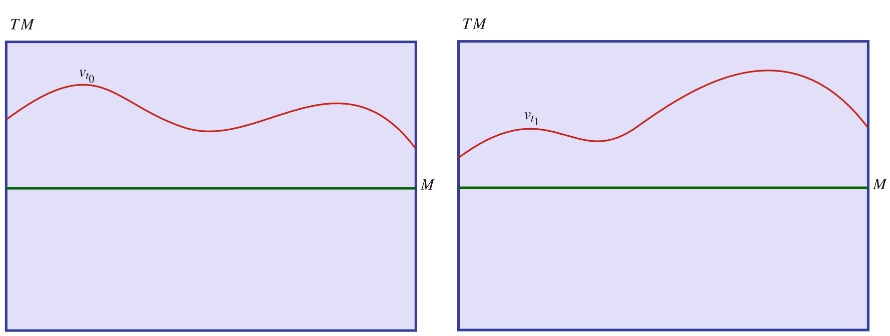


Fig. B.9 The tangent bundle TM shown with section v_{t_0} (left) and section v_{t_1} (right)

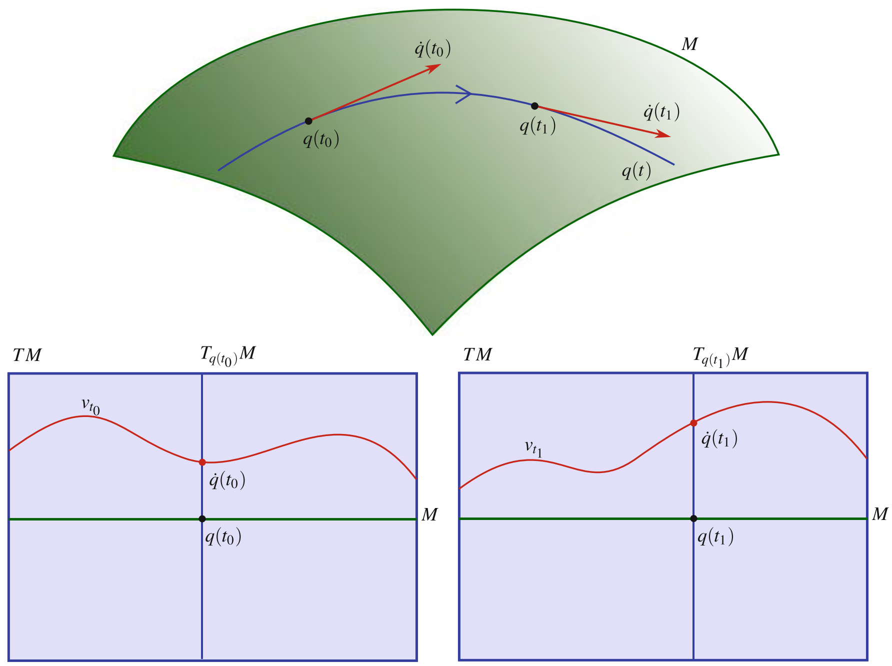


Fig. B.10 Top: The manifold M with a single flow trajectory $q(t) = \big(q_1(t), q_2(t), \ldots, q_n(t)\big)$. At the point $q(t_0)$ the velocity vector $\dot{q}(t_0)$ is shown, and at the point $q(t_1)$ the velocity vector $\dot{q}(t_1)$ is shown. Bottom: The points $q(t_0)$ and $q(t_1)$ along with the vectors $\dot{q}(t_0)$ and $\dot{q}(t_1)$ shown in the vector bundle TM

Let us single out one flow line on M and look at it closely, as in Fig. B.10, top. The trajectory of the flow gives us a curve on M. We will denote the curve by

$$q(t) = \big(q_1(t), q_2(t), \ldots, q_n(t)\big).$$

At time t_0 the velocity vector of the curve is given by

$$\begin{aligned}\dot{q}(t_0) &= \big(\dot{q}_1(t_0), \dot{q}_2(t_0), \ldots, \dot{q}_n(t_0)\big)\\ &= \left(\frac{\partial q_1(t)}{\partial t}\bigg|_{t_0}, \frac{\partial q_2(t)}{\partial t}\bigg|_{t_0}, \ldots, \frac{\partial q_n(t)}{\partial t}\bigg|_{t_0}\right).\end{aligned}$$

This velocity vector $\dot{q}(t_0)$ is at the point $q(t_0)$. We get the velocity vector at a later time, t_1, in an identical fashion. See Fig. B.10.

So, if we were somehow to know the velocity vector field $q(t)$ of some flow q and we wanted to find the actual curves of the flow q we would need to integrate $\dot{q}$; that is,

$$q(t) = \int \dot{q}(t)dt.$$

Thus the flow curve q is the **integral curve** of the velocity vector field $\dot{q}$. We won't even pretend to do this here for a real example, but the general abstract idea should be relatively clear - the velocity vector field of a flow can be integrated to get the flow. So, if we had some nice way to know the velocity vector field of a flow it is relatively straightforward to get the actual flow.

Now we consider the simple example of single particle in $\mathbb{R}^3$, which, for simplicity's sake, we will assume has a mass of one in whatever units we are working with. This is an assumption that mathematicians almost always make, physicists sometimes make, and thankfully engineers never make. We will also basically pretend that this particle has no volume, that it is just a single point that has a mass of unit one. This is another simplifying assumption that is often made, and when this is done the particle is called a point mass. The coordinates of this point mass is given by $q = (q_1, q_2, q_3)$. Again, as part of simplifying notation the coordinates are often just written as q, but it needs to be understood that this q does, in reality, mean (q_1, q_2, q_3).

We will assume that this point mass has some force acting on it, which we will denote F. Often the force that is acting on the particle depends on the location of the particle, and so is written as a function of the coordinates, $F(q)$. If the force is also a function of time then we would write $F(q,t)$. Often forces F can be written in terms of what is called a **potential function**, which is very often denoted by U. This is not a physics class, this is simply a very brief introduction to geometric mechanics, so we will not go into depth discussing potential functions. What this all means to us is simply that we can write the force F in terms of the potential function U as

$$F(q) = -\frac{\partial U(q)}{\partial q} \qquad \text{or} \qquad F(q,t) = -\frac{\partial U(q,t)}{\partial q}.$$

Often we say that the force F comes from the potential function U.

We will assume this is the case within our example, that the force F comes from a potential function U. So what does the equation $F(q,t) = -\frac{\partial U(q,t)}{\partial q}$ actually mean in our example with the single point mass in $\mathbb{R}^3$? Recall that in actuality $q = (q_1, q_2, q_3)$ so the equation $F(q,t) = -\frac{\partial U(q,t)}{\partial q}$ actually means we have the system of equations

$$\begin{aligned}
F_1(q_1,q_2,q_3,t) &= -\frac{\partial U(q_1,q_2,q_3,t)}{\partial q_1},\\
F_2(q_1,q_2,q_3,t) &= -\frac{\partial U(q_1,q_2,q_3,t)}{\partial q_2},\\
F_3(q_1,q_2,q_3,t) &= -\frac{\partial U(q_1,q_2,q_3,t)}{\partial q_3},
\end{aligned}$$

where F_i is the component of the force in the q_ith direction.

By Newton's law we know $F = ma$, where m is the mass and a is the acceleration. Of course, we are assuming $m = 1$ so with that simplifying assumption we can write $F = a$. But we also know that acceleration is just the derivative of velocity, so for our particle we have

$$a = \ddot{q} = \frac{\partial^2 q(t)}{\partial t^2},$$

resulting in

$$F = a \quad \Rightarrow \quad \ddot{q}(t) = -\frac{\partial U(q,t)}{\partial q},$$

which is clearly a second-order differential equation. Actually, like with force above it is actually a system of three second-order differential equations:

$$\begin{aligned}
\ddot{q}_1(t) &= -\frac{\partial U(q_1,q_2,q_3,t)}{\partial q_1},\\
\ddot{q}_2(t) &= -\frac{\partial U(q_1,q_2,q_3,t)}{\partial q_2},\\
\ddot{q}_3(t) &= -\frac{\partial U(q_1,q_2,q_3,t)}{\partial q_3},
\end{aligned}$$

where $\ddot{q}_i$ is the component of acceleration in the q_ith direction.

In physics **momentum** is defined to be $p = m\dot{q} = \dot{q}$, where the last equality is because $m = 1$. Including in our notation the dependence on time we would write $p(t) = \dot{q}(t)$. Like with the potential functions, we will not get sidetracked into describing or explaining what momentum is or why it is important, for us we will simply use it mathematically to turn our second-order ordinary differential equation into a system of first-order ordinary differential equations

$$\begin{aligned}
\dot{q}(t) &= p(t),\\
\dot{p}(t) &= -\frac{\partial U(q,t)}{\partial q}.
\end{aligned}$$

Again, like our F and $\ddot{q}$ above, this is really a set of six ordinary first-order differential equations.

The total energy of the system is the sum of the potential energy and kinetic energy of the particle. The potential energy of the particle located at q at time t is given by the potential function $U(q,t)$ while the kinetic energy of the particle is given by $\frac{1}{2}\big(p(t)\big)^2 \equiv \frac{1}{2}p^2(t)$. Thus the total energy, which is denoted by H_t is given by

$$H_t(p,q,t) = \frac{1}{2}p^2(t) + U(q,t).$$

The total energy function is an example of what is called a **Hamiltonian function**, which is why we denoted it with an H, while the subscript t is included in the notation to help us recall the function's dependence on time. We can take the derivative of H_t with respect to both p and q. Doing this and combining with the above system equalities we get

$$\begin{aligned}
\frac{\partial H_t}{\partial p} &= \frac{\partial}{\partial p}\left(\frac{1}{2}p^2(t) + U(q,t)\right) = p = \dot{q},\\
\frac{\partial H_t}{\partial q} &= \frac{\partial}{\partial q}\left(\frac{1}{2}p^2(t) + U(q,t)\right) = \frac{\partial U(q,t)}{\partial q} = -\dot{p}.
\end{aligned}$$

Thus the above system of first-order differential equations can be written in terms of the Hamiltonian function as

$$\begin{aligned}
\dot{q} &= \frac{\partial H_t}{\partial p},\\
\dot{p} &= -\frac{\partial H_t}{\partial q}.
\end{aligned}$$

This is a system of equations are called **Hamiltonian equations**. At each particular moment in time t the equations give us a velocity vector $(\dot{q},\dot{p})$. The Hamiltonian equations describes the motion of our point mass particle under the influence of the force F that comes from the potential function U. In other words, by solving this system of equations we can get the motion of the particle. It turns out that almost any physical system has a Hamiltonian function that can be used to write Hamiltonian equations for the system.

Actually, instead of just the variables q we now have the variables q and p. By increasing the number of variables we turned a second-order differential equation into a system of first-order differential equations. This makes solving the differential equations easier, but at the cost of having to solve for more variables. So instead of just configuration coordinates q we now have momentum coordinates p. The configuration coordinates and momentum coordinates together give what is called the **phase space manifold** of the system. In our example the phase space manifold is $\mathbb{R}^6$ with coordinates $(q_1,q_2,q_3,p_1,p_2,p_3)$.

Let us sidetrack for a moment and think about what this momentum p actually is. Recall, we defined it as $p = m\dot{q}$. Also recall that we claimed the kinetic energy of the particle was $\frac{1}{2}p^2$. How we arrived at that is not difficult to see, after all,

$$KE = \frac{1}{2}mv^2 = \frac{1}{2}m\dot{q}^2 = \frac{1}{2m}(m\dot{q})^2 = \frac{1}{2m}p^2,$$

and we then assumed unit mass, $m = 1$. But let us consider a different way of writing this,

$$\begin{aligned}
KE &= \frac{1}{2}mv^2 = \frac{1}{2}m\dot{q}^2 = \frac{1}{2}(m\dot{q})\dot{q} = \frac{1}{2}p\dot{q}\\
&= \frac{1}{2}[p_1,p_2,p_3]\begin{bmatrix}\dot{q}_1\\ \dot{q}_2\\ \dot{q}_3\end{bmatrix}.
\end{aligned}$$

So $\frac{p}{2}$ multiplied by $\dot{q}$ gives us a real number representing the kinetic energy of the particle. Clearly, if our manifold M is the configuration space with coordinates q then $\dot{q}$ is a vector at some point $q \in M$, that is, $\dot{q} \in T_qM$. But what eats vectors to give real numbers? One-forms do. Thus our $\frac{1}{2}[p_1,p_2,p_3]$ can be viewed as a one-form or a covector, and hence $p \in T_q^*M$. We admit it, this seems a little bit strange at first glance. In reality the rationale for why we want to view momentum as a one-form is a little more involved and subtle than this, but the above argument will at least give you an idea of why it is true.

Now we have three manifolds, the configuration manifold M with coordinates $q = (q_1, q_2, q_3)$, the tangent bundle TM with coordinates $(q, \dot{q}) = (q_1, q_2, q_3, \dot{q}_1, \dot{q}_2, \dot{q}_3)$, and the cotangent bundle T^*M with coordinates $(q, p) = (q_1, q_2, q_3, p_1, p_2, p_3)$. The cotangent bundle T^*M is also called the phase space manifold of the system. As odd as it may seem, this phase space manifold T^*M is actually the important manifold for studying physical systems.

In order to understand why, we need to make another sidetrack and define **Hamiltonian vector fields**. A Hamiltonian function is a time dependent function defined on the phase space of a physical system. The exact form a Hamiltonian function has depends on the physical system. It is usually the total energy of the system, which for a closed system is the kinetic energy plus the potential energy. The Hamiltonian function for a single particle in a force field was given above. In other words, suppose we had a physical system with the configuration manifold M, then a Hamiltonian function is some real-valued function

$$
\begin{aligned}
H_t : T^*M \times \mathbb{R} &\longrightarrow \mathbb{R} \\
(q, p, t) &\longmapsto H_t(q, p, t).
\end{aligned}
$$

If we hold the time fixed at a particular moment t_0 then we get a real-valued function $H_{t_0} : T^*M \to \mathbb{R}$. For simplicity's sake we will continue to use the notation H_t to denote our Hamiltonian function at a fixed time instead of H_{t_0}.

We can use this Hamiltonian function on T^*M along with the canonical symplectic form ω on T^*M to define a vector on T^*M. Because the definition of this vector field relies on the Hamiltonian function it is called the Hamiltonian vector field and is very often denoted by X_{H_t}. It is defined by the following formula

$$
\iota_{X_{H_t}} \omega = dH_t.
$$

This formula is clearly a little cumbersome, with subscripts down to the third level. Recall how the interior product operator works, if α is a k-form then $\iota_v \alpha$ plugs the vector v into α's first slot and turns it into a $(k-1)$-form. Here the symplectic form ω is a two-form, so when the vector field X_{H_t} is plugged into the first slot we have $\omega(X_{H_t}, \cdot)$, which is a one-form. And since the Hamiltonian function H_t is a real-valued function then the exterior derivative of it is the one-form dH_t. So another way of writing this formula would be as

$$
\omega(X_{H_t}, \cdot) = dH_t(\cdot).
$$

This formula means that for any vector field v at all, we have that $\omega(X_{H_t}, v) = dH_t(v)$.

The canonical symplectic form ω is of course fixed, and once we are given a Hamiltonian function then clearly dH_t is fixed as well. The Hamiltonian vector field is defined to be exactly that vector field that makes the one-form $\omega(X_{H_t}, \cdot)$ exactly the same as the one-form dH_t. We will not discuss why we know such a Hamiltonian vector field exists or is unique here, simply suffice it to say that it does and is.

Since our Hamiltonian vector field X_{H_t} is a vector field on the manifold T^*M then we can write it as

$$
X_{H_t} = \begin{bmatrix} v_1 \\ \vdots \\ v_{2n} \end{bmatrix}
$$

where n is the dimension of manifold M. Noting that we can write $\omega = \sum_{i=1}^n dq_i \wedge dp_i$ we have

$$
\begin{aligned}
\iota_{X_{H_t}} \omega &= \Big(\sum_{i=1}^n dq_i \wedge dp_i \Big) \big(X_{H_t}, \cdot \big) \\
&= \Big(\sum_{i=1}^n \big(dq_i dp_i - dp_i dq_i \big) \Big) \big(X_{H_t}, \cdot \big) \\
&= \sum_{i=1}^n \big(dq_i(X_{H_t}) dp_i - dp_i(X_{H_t}) dq_i \big) \\
&= \sum_{i=1}^n \big(v_i dp_i - v_{n+i} dq_i \big).
\end{aligned}
$$

Taking the exterior derivative of the Hamiltonian function gives us

$$dH_t = \sum_{i=1}^{n} \frac{\partial H_t}{\partial q_i} dq_i + \sum_{i=1}^{n} \frac{\partial H_t}{\partial p_i} dp_i.$$

Putting this together we see that

$$\iota_{X_{H_t}} \omega = dH_t \Rightarrow \quad v_i = \frac{\partial H_t}{\partial p_i} \quad \text{and} \quad v_{n+i} = -\frac{\partial H_t}{\partial q_i}$$

for $i = 1, \ldots, n$. Recalling that this is true for each time t we can explicitly put in the time dependence,

$$v_i(t) = \frac{\partial H_t}{\partial p_i}(t) \quad \text{and} \quad v_{n+i}(t) = -\frac{\partial H_t}{\partial q_i}(t)$$

so we can write the Hamiltonian vector field as

$$X_{H_t}(t) = \begin{bmatrix} v_1(t) \\ \vdots \\ v_{2n}(t) \end{bmatrix}.$$

This allows us to integrate the Hamiltonian vector field to find its integral curve

$$\Big(q_1(t), \ldots q_n(t), p_1(t), \ldots, p_n(t)\Big).$$

In other words, using the notation associated with the integral curve we can write $v_i = \dot{q}_i$ and $v_{n+1} = \dot{p}_i$, which give us

$$\dot{q}_i(t) = \frac{\partial H_t}{\partial p_i}(t) \quad \text{and} \quad \dot{p}_i(t) = -\frac{\partial H_t}{\partial q_i}(t).$$

But these are exactly the Hamiltonian equations we had from before. Thus it turns out that Hamilton's equations governing the time evolution of a physical system simply fall out of the identity $\iota_{X_{H_t}} \omega = dH_t$. That is, the Hamiltonian vector field associated with some Hamiltonian function H_t is exactly the vector field that needs to be integrated to give the time evolution of the system and thus the formula $\iota_{X_{H_t}} \omega = dH_t$ gives the differential equations that need to be solved. So it turns out that the cotangent bundle T^*M, which comes equipped with the canonical symplectic form, is the appropriate mathematical space to study many mechanical and physical systems.

B.6 Summary and References

B.6.1 Summary

De Rham cohomology creates a link between differential forms on a manifold M and the global topology of that manifold. It does this using the de Rham cohomology group, which is defined as

$$H^k_{dR}(M) = \frac{\left\{ \alpha \in \bigwedge^k(M) \;\middle|\; d\alpha = 0 \right\}}{\left\{ d\omega \in \bigwedge^k(M) \;\middle|\; \omega \in \bigwedge^{k-1}(M) \right\}}.$$

It must be noted that the de Rham cohomology group is not just a group in the mathematical sense of the word, it is a vector space. Using the Poincaré lemma as a conceptual framework, the de Rham cohomology group measures the degree of difference between closed and exact forms and hence how different an n-dimensional manifold M is from $\mathbb{R}^n$.

A symplectic manifold is a manifold M that has a closed non-degenerate two-form ω defined everywhere on it. Suppose the coordinate functions of M are given by $(x_1, \ldots, x_n)$ and the coordinate functions of T^*M are given by

$(x_1, \ldots, x_n, \xi_1, \ldots, \xi_i)$. Then $T_p^*M = \text{span}\{dx_1, \ldots, dx_n\}$ and $T_{\alpha_p}^*\big(T^*M\big) = \text{span}\{dx_1, \ldots, dx_n, d\xi_1, \ldots, d\xi_n\}$, but be careful, the dx_i for these two spaces are not the same. Hence a one-form on T^*M, that is, an element of $T^*\big(T^*M\big)$, would have the general form

$$\beta = \sum_{i=1}^{n} \beta_i dx_i + \sum_{i=1}^{n} \widetilde{\beta}_i d\xi_i,$$

where β_i and $\widetilde{\beta}_i$, for $i = 1, \ldots, n$, are functions on T^*M, $\beta_i : T^*M \to \mathbb{R}$ and $\widetilde{\beta}_i : T^*M \to \mathbb{R}$. Consider the coordinate functions x_i and ξ_i on T^*M, $x_i : T^*M \to \mathbb{R}$ and $\xi_i : T^*M \to \mathbb{R}$. Letting $\beta_i = \xi_i$ and $\widetilde{\beta}_i = 0$ gives the Liouville one-form α on T^*M,

$$\boxed{\text{Liouville One-Form:} \qquad \alpha = \textstyle\sum_{i=1}^{n} \xi_i dx_i.}$$

This one-form always exists on the manifold T^*M. We use this to define the canonical symplectic form on T^*M as

$$\boxed{\text{Canonical Symplectic Form:} \qquad \omega \equiv -d\alpha = \textstyle\sum_{i=1}^{n} dx_i \wedge d\xi_i}$$

where α is the Liouville one-form on T^*M.

$$\boxed{\begin{aligned} \text{Liouville One-Form:} \qquad \alpha(v_{\xi_p}) &\equiv \Pi(v_{\xi_p})\big(\pi_*(v_{\xi_p})\big) \\ &= \Big\langle \Pi(v_{\xi_p}), \pi_*(v_{\xi_p}) \Big\rangle. \end{aligned}}$$

Notice that at no point in this definition were the actual manifold coordinates employed. The canonical symplectic form is then simply $\omega = -d\alpha$. See Fig. B.11.

The Darboux theorem states that if (M, ω) is a symplectic manifold of dimension $2n$ then in a neighborhood U of each point $p \in M$ there exist local coordinates $(x_1, \ldots, x_n, y_1, \ldots, y_n)$ such that the symplectic form ω can be written as

$$\omega = \sum_{i=1}^{n} dx_i \wedge dy_i.$$

A major application of symplectic geometry and the Darboux theorem is geometric mechanics, the attempt to write classical mechanics in geometric terms. For example, it is possible to write the Hamiltonian equations of a simple mechanical system in terms of symplectic forms.

$$\begin{tikzcd}
 & v_{\xi_p} \in T\big(T^*M\big) \arrow[dl, "\Pi"'] \arrow[dr, "\pi_*"] & \\
\Pi(v_{\xi_p}) \in T^*M \arrow[r, Rightarrow] \arrow[dr, "\pi"'] & \alpha(v_{\xi_p}) \equiv \Big\langle \Pi(v_{\xi_p}), \pi_*(v_{\xi_p}) \Big\rangle & \pi_*(v_{\xi_p}) \in TM \arrow[l, Rightarrow] \\
 & \pi(\xi_p) \in M &
\end{tikzcd}$$

Fig. B.11 The Liouville one-form α is defined intrinsically in terms of the vector v_{ξ_p}. The projection mapping $\pi : T^*M \to M$, defined by $\pi(\xi_p) = p$, induces the push-forward mapping $\pi_* : T(T^*M) \to TM$. The mapping Π is also a projection mapping $\Pi : T(T^*M) \to T^*M$ defined by $\Pi(v_{\xi_p}) = \xi_p$. The Liouville one-form α eating the vector v_{ξ_p} is defined by the canonical pairing of the one-form $\Pi(v_{\xi_p})$ with the vector $\pi_*(v_{\xi_p})$

B.6.2 References and Further Reading

Two excellent introductions to de Rham cohomology are Tu [46] followed by Bott and Tu [7]. In fact, Tu wrote the first book essentially to cover the necessary material for the second book. An additional source, more for graduate students and working mathematicians, is the book by Madsen and Tornhave [30].

Symplectic manifolds and the Darboux theorem are covered from a more mathematical point of view by da Silva [11], which we relied on in particular for the proof of the Darboux theorem. One can approach the Darboux theorem using various levels of mathematical sophistication, but this approach is the most appropriate for the intended audience of this book. But see also Abraham, Marsden, and Ratiu [1]. Additionally, any book on geometric mechanics will also cover the basics of symplectic manifolds.

There are quite a number of interesting and good introductions to geometric mechanics for those who are interested in a more complete treatment of the material; Singer [39], Holm, Schmah, and Stoica [26], Talman [44], Calin and Chang [9], and Holm [24, 25].

References

1. R. Abraham, J.E. Marsden, and T. Ratiu. *Manifolds, Tensor Analysis, and Applications*. Springer, 2 edition, 1991.
2. Anonymous. Determinants. http://www.amarketplaceofideas.com/wp-content/uploads/2013/12/Determinant-Derivation.pdf, n.d.
3. V.I. Arnold. *Mathematical Methods of Classical Mechanics*. Springer, 2 edition, 2000.
4. David Bachman. *A Geometric Approach to Differential Forms*. Birkhäuser, 2006.
5. Richard Bishop and Richard Crittenden. *Geometry on Manifolds*. Academic Press, 1964.
6. Richard Bishop and Samuel Goldbirg. *Tensor Analysis on Manifolds*. Dover Publications, Inc., 1980.
7. Raoul Bott and Loring Tu. *Differential Forms in Algebraic Topology*. Springer, 1982.
8. William Burke. *Applied Differential Geometry*. Cambridge University Press, 1987.
9. Ovidiu Calin and Der-Chen Chang. *Geometric Mechanics on Riemannian Manifolds*. Birkhäuser, 2005.
10. Lawrence Conlon. *Differentiable Manifolds*. Birkhäuser, 2 edition, 2001.
11. Ana Cannas da Silva. *Lectures on Symplectic Geometry*. Springer, 2008.
12. R.W.R. Darling. *Differential Forms and Connections*. Cambridge University Press, 1994.
13. George DesChamps. Electromagnetics and differential forms. *Proceedings of the IEEE*, 69(6), June 1981.
14. Manfredo do Carmo. *Riemannian Geometry*. Birkhäuser, 1992.
15. Joaquim Domingos. *Geometrical Properties of Vectors and Covectors*. World Scientific, 2006.
16. Tevian Dray. *Differential Forms and the Geometry of General Relativity*. CRC Press, 2015.
17. C.H. Edwards. *Advanced Calculus of Several Variables*. Dover, 1994.
18. Harold Edwards. *Advanced Calculus: A Differential Forms Approach*. Birkhäuser, 1969.
19. Harley Flanders. *Differential Forms with Applications to the Physical Sciences*. Dover, 1963.
20. Daniel Fleisch. *A Student's Guide to Maxwell's Equations*. Cambridge University Press, 2008.
21. Theodore Frankel. *The Geometry of Physics*. Cambridge University Press, 2 edition, 2004.
22. Stephen Friedberg, Arnold Insel, and Lawrence Spence. *Linear Algebra*. Prentice Hall, 3 edition, 1997.
23. Thomas Garrity. *Electricity and Magnetism for Mathematicians*. Cambridge University Press, 2015.
24. Darryl Holm. *Geometric Mechanics Part I: Dynamics and Symmetry*. World Scientific, 2008.
25. Darryl Holm. *Geometric Mechanics Part II: Rotating, Translating, and Rolling*. World Scientific, 2008.
26. Darryl Holm, Tanya Schmah, and Cristina Stoica. *Geometric Mechanics and Symmetry*. Oxford University Press, 2009.
27. John Hubbard and Barbara Hubbard. *Vector Calculus, Linear Algebra, and Differential Forms: A Unified Approach*. Matrix Editions, 4th edition, 2009.
28. Matthias Kawski. Vector field analyzer ii. https://math.la.asu.edu/~kawski/vfa2/, 2001. Accessed: 2018-05-12.
29. David Kay. *Tensor Calculus*. Schaum's Outlines. McGraw Hill, 2011.
30. Ib Madsen and Jørgen Tornehave. *From Calculus to Cohomology*. Cambridge University Press, 1999.
31. Jarrold Marsden and Michael Hoffman. *Elementary Classical Analysis*. W.H. Freeman and Company, 2 edition, 1993.
32. Jerrold Marsden and Tudor Ratiu. *Introduction to Mechanics and Symmetry*. Springer-Verlag, 1994.
33. Daniel Martin. *Manifold Theory*. Harwood Publishing, 2002.
34. Charles Misner, Kip Thorne, and John Wheeler. *Gravitation*. Freeman and Company, 1970.
35. James Munkres. *Analysis on Manifolds*. Addison-Wesley Publishing Company, 1991.
36. Barrett O'Neill. *Elementary Differential Geometry*. Academic Press, 2 edition, 2006.
37. Paul Renteln. *Manifolds, Tensors, and Forms*. Cambridge University Press, 2014.
38. H.M. Schey. *Div, Grad, Curl, and All That*. W.W. Norton and Company, Inc., 1973.
39. Stephanie Frank Singer. *Symmetry in Mechanics*. Birkhäuser, 2 edition, 2004.
40. Michael Spivak. *Calculus on Manifolds*. Westview press, 1998.
41. Michael Spivak. *A Comprehensive Introduction to Differential Geometry, Vol I*. Publish or Perish, Inc., 3 edition, 2005.
42. Charles Stevens. *The Six Core Theories of Modern Physics*. MIT Press, 1995.
43. James Stewart. *Calculus*. Brooks/Cole Cengage Learning, 6 edition, 2009.
44. Richard Talman. *Geometric Mechanics*. John Wiley and Sons, Inc., 2000.
45. J.A. Thorpe. *Elementary Topics in Differential Geometry*. Springer-Verlag, 1979.
46. Loring Tu. *An Introduction to Manifolds*. Springer, 2 edition, 2010.
47. Gerard Walschap. *Multivariable Calculus and Differential Geometry*. De Gruyter, 2015.

J. P. Fortney, *A Visual Introduction to Differential Forms and Calculus on Manifolds*,
https://doi.org/10.1007/978-3-319-96992-3

48. Karl Warnick and Peter Russer. Two, three and four-dimensional electromagnetics using differential forms. *Turk J. Elec Engin*, 14(1), 2006.
49. Karl Warnick, Richard Selfridge, and David Arnold. Teaching electromagnetic field theory using differential forms. *IEEE Transactions on Education*, 40(1), February 1997.

Index

J. P. Fortney, *A Visual Introduction to Differential Forms and Calculus on Manifolds*,
https://doi.org/10.1007/978-3-319-96992-3